Development of a Web Compute Server for a Bioinformatic Tool, the Story of www.ConfBuster.org

Gabriel Bégin
Laboratory of Pr Patrick Lagüe
October 2018
# Tables of Contents

<table>
<thead>
<tr>
<th>A</th>
<th>Server Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>What is ConfBuster?</td>
</tr>
<tr>
<td>2.</td>
<td>System Architecture</td>
</tr>
<tr>
<td>3.</td>
<td>Web UI Development</td>
</tr>
<tr>
<td>4.</td>
<td>Queue Design</td>
</tr>
<tr>
<td>5.</td>
<td>Compute Process</td>
</tr>
<tr>
<td>6.</td>
<td>New Software Integration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>Implementation for Non-Bioinformatician</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Parameters Selection</td>
</tr>
<tr>
<td>2.</td>
<td>Web UI Simplification</td>
</tr>
<tr>
<td>3.</td>
<td>Data Transport Flexibility</td>
</tr>
<tr>
<td>4.</td>
<td>Results Delivery</td>
</tr>
<tr>
<td>5.</td>
<td>Data Privacy Policy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>Appendix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Full-Stack Architecture</td>
</tr>
<tr>
<td>2.</td>
<td>UML System Design</td>
</tr>
<tr>
<td>3.</td>
<td>Web Front-End</td>
</tr>
<tr>
<td>4.</td>
<td>Web Back-End</td>
</tr>
<tr>
<td>5.</td>
<td>Queue System Versatility</td>
</tr>
<tr>
<td>6.</td>
<td>Multi-Node Computing</td>
</tr>
</tbody>
</table>
First Part

Server Development
**Server Development**

**What is ConfBuster?**

ConfBuster is an Open Source Suite of Tools for Macrocycle Conformational Search and analysis. It searches if a more stable conformation for a given macrocycle is present. Search Results are based on:

- Energy
- RMSD

Bioinformatics knowledge is needed.

**Macrocycle**

- Molecule with a ring of at least 8 atoms
- High interest in drug development

**Conformers Visualization with PyMOL**

**Result Matrix with RMSD-Energy relationship**


ConfBuster is a suite of tools written in Python with the goal of identifying the lowest energy conformations of macrocycles. The suite also includes tools for the analysis and visualization of the conformational search results. Coordinate sets of a single molecule in PDB/SDF/MOL2 format is required as input. A set of conformation coordinates is returned as output, as well as PyMOL script and graphics for results analysis. Identified conformations are sorted from the lowest to the highest energy.

The link to download the suite and this server are provided below. The present server allows you to freely use ConfBuster without software installation.

Please cite:


Download Examples

Sketch a Macrocycle (external links)

Additional Information

- ConfBuster and ConfBuster Web Server are under GNU General Public License v3.0
- ConfBuster uses an heuristic algorithm: the results may vary
- ConfBuster needs 3D coordinates: a file with only 2D coordinates will be rejected
- ConfBuster needs hydrogen atoms: a molecule without hydrogen will be rejected
- The input file, the result files, and the email address will be deleted after the compute session

http://ConfBuster.org

Home Page
Server Development
System Architecture

Unified Modeling Language

White 3D Boxes = Physical Node
Unified Modeling Language
Grey Boxes = (Software) Component
Server Development
System Architecture

Unified Modeling Language

Colored Boxes = (Package) Layer

Gabriel Bégin
Unified Modeling Language

Small White Boxes = (Class) Module
Unified Modeling Language

Dotted Arrows

A uses B
Front-End Technologies

1) Homemade Files
   - Homemade Pages & Libraries

2) Bootstrap CSS
   - Page Layout Management Libraries
   - Language CSS
   - Simple Syntax

3) jQuery JS
   - UI Logic Library
   - Simplifies the usage of Javascript
   - Language : (jQuery &) Javascript
Bootstrap CSS

- CSS Library
- **Used to simplify Web Interface Design**
- Layout based on a 12 column grid
- Rows are added manually
- Allow to configure Responsive UI Resizing
  - Computer
  - Tablet
  - Mobile
- Also provide esthetic buttons/controls
- With current template and examples:
  - **Fast and Easy to Learn (~1h)**

https://getbootstrap.com/
ConfBuster is a suite of tools written in Python with the goal of identifying the lowest energy conformations of macrocycles. The suite can take structures in PDB format as input. A set of conformational search results. Coordinate sets for results analysis. Identified conformations.

The link to download the suite and this server are provided below. The present server allows you to freely use ConfBuster without software installation.

Please cite !


Download

Start a Macrocycle Conformational Search

Let our servers run ConfBuster for you

Email

recipient@example.com

- The results (zip file) will be sent to this address
- Attention: The email client has to allow attachment
jQuery

- Javascript Library
- Used to simplify Web Interface Logic
- Different Syntax based on Javascript
- Allow to configure Dynamic Animations
  - Bouton Fade
  - Text Fade
- With current template and examples :
  - Fast and Easy to Learn (~1h)

https://jquery.com/
Job Submission Flow

1) Home.html
- Data Validation
- Data Submission

2) DataReception.php
- Data Validation
- Data Encapsulation

3) Synchronous HTTP POST Transaction
- Transmission by B2B Transaction

4) Confirmation.html
- Transmission Confirmation for the user
Job Submission Flow

3) Data Transaction

- Synchronous HTTP POST
- Web Server to Web Server
- Transmission by B2B Transaction
- B2B : Business-to-Business

- Prevent Socket Management
- Prevent TCP Port Management

- Apache already implements multi-request management

- Allow Multi-Threaded Requests
Job Submission Flow

5) **ComputeJobPreparation.php**
- Open Data Package
- Send Data to the Database Manager
- Send Data to the Mail Manager

6) **DatabaseManager.php**
- Send the Data into the Queue
- SQL Language

7) **MySQL DBMS**
- Database Software
- Store Jobs & Configurations
- Job Queue

8) **MailManager.php**
- Send a Queue Entry Confirmation to the user
Job Submission Flow

9) QueueManager.py
- Queue Scheduler
- Extract job from database
- Send job to a worker thread
- Runs as an independent server

10) DatabaseManager.php
- Extract the Data from the Queue
- SQL Language

11) ComputeManager.php
- Abstract Structure of a Worker
- Contains Processing Instructions
- Asynchronous Job Processing

12) FileSystemManager.py
- Simplifies File System Management

13) MailManager.py
- Send Results to User
Server Development
New Software Integration

Current System

Business Layer

FileSystemManager.py

ComputeManager.py

MailManager.py  QueueManager.py

ConfBuster-Single-Molecule-Minimization.py
ConfBuster-Macrocycle-Linear-Sampling.py
ConfBuster-Analysis.py
Server Development
New Software Integration

Compute Command Customization

Any Commands Supported by the Operating System of the Server.
Server Development
Compute Process Summary

**Database Manager**
(Queue Interaction)

**Queue Manager**
(Scheduler)

**Compute Manager**
(Worker)

**File System Manager**
(WorkBench Access)

**Mail Manager**
(Results Delivery)

---


---

Gabriel Bégin
Second Part

Implementation for Non-Bioinformatician
**Implementation for Non-Bioinformatician**

**Parameters Selection**

### Challenges

- Make ConfBuster accessible for chemist
- Selection of the most relevant parameters

### Solutions

- No option selection for the first version
- Keep most of the parameters by default

---

### ConfBuster tools

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Description</th>
<th>Parameters</th>
</tr>
</thead>
</table>
| ConfBuster-Single-Molecule-Minimization                                   | Performs a simple minimization of the given molecule (recommended before).  | -i input filename [mandatory]  
-o output name prefix [default: replace input file]  
CONFBUSTER-Rotamer-Search Identify rotational isomers of a molecule. |
| ConfBuster-Rotamer-Search                                                 | Identify rotational isomers of a molecule.                                  | -i input filename in mol2 [mandatory]  
-g number of generations [default: 100]  
e the energy cutoff to discriminate conformations in units of kcal/mol [default: 50]  
d output directory name [default: use the prefix of the input filename]  
f format of outputted molecules [xyz or default: mol2] |
| ConfBuster-Macrocycle-Linear-Sampling                                    | Performs a conformational search of a cyclic molecule.                      | -i input filename [mandatory]  
-r rmsd cutoff in Angstrom [default: 0.5]  
-n for each cleaving point, number of rotamer searches performed [default: 5]  
-N for each cleaving point, number of molecules extracted from each rotamer search [default: 5]  
-o output directory name [default: prefix of the input filename] |
| ConfBuster-Analysis                                                      | Perform post-analyses to visualize a clustering based on RMSD values between the conformations. | -i directory name of the search results [mandatory]  
-r rmsd cut off [default: none]  
-n number of conformations to include in the analysis [default: all]  
e midpoint value of the energy color scale [default: 0] |

[https://github.com/patricklague/ConfBuster](https://github.com/patricklague/ConfBuster)
Implementation for Non-Bioinformatician
Web UI Simplification

Straightforward Design

- No menu
- All content on the same page
- Important content at the left
- Form at the right

- Only 3 operations needed
  - 1) Email Input
  - 2) Macrocycle file
  - 3) Form submission

- Downloadable Examples
- Molecule sketchers

- Privacy policy
Usage of Transport Objects

- Increase flexibility for further changes to the software stack.
- Prevent configuring every property on each function.
- Allows to rapidly modify the system according to users needs.
Implementation for Non-Bioinformatician
Results Delivery

Delivery Method

- Result Sent by Email
  - Universal & Accessible

- Zip Package
  - Conformations Identified
  - PyMOL Visualization Script
  - RMSD/Energy Matrix
Implementation for Non-Bioinformatician
Data Privacy Policy

Public Privacy Policy

- No Data are kept on the servers after the compute session
  - Compute data are deleted
  - Result data are deleted

Additional Information

- ConfBuster and ConfBuster Web Server are under GNU General Public License v3.0
- ConfBuster uses an heuristic algorithm: the results may vary
- ConfBuster needs 3D coordinates: a file with only 2D coordinates will be rejected
- ConfBuster needs hydrogen atoms: a molecule without hydrogen will be rejected
- The input file, the result files, and the email address will be deleted after the compute session
Third Part

Appendix
Appendix
Full-Stack Architecture

Multi-Layer
Architectures

Gang of Four
- Erich Gamma
  - Richard Helm
  - Ralph Johnson
  - John Vlissides

OOP Design Patterns
- Abstract Factory
- Singleton
- DAO/VO (*Composite)
- MVC (*Observer)

* Ensued from …

Web Stack Models

Examples of Web Stacks

<table>
<thead>
<tr>
<th>GNU/Linux</th>
<th>Windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNU/Linux</td>
<td></td>
</tr>
<tr>
<td>Apache</td>
<td>Windows</td>
</tr>
<tr>
<td>MySQL</td>
<td>IIS</td>
</tr>
<tr>
<td>PHP</td>
<td>SQL Server</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GNU/Linux</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GNU/Linux</td>
<td></td>
</tr>
<tr>
<td>Apache Tomcat</td>
<td>Windows</td>
</tr>
<tr>
<td>MySQL</td>
<td>IIS</td>
</tr>
<tr>
<td>JavaEE</td>
<td>SQL Server</td>
</tr>
<tr>
<td></td>
<td>ASP.Net</td>
</tr>
</tbody>
</table>

(Web) Full-Stack
Architecture

Examples of Web Stacks:

1. GNU/Linux
   - Apache
   - MySQL
   - PHP

2. GNU/Linux
   - Apache Tomcat
   - MySQL
   - JavaEE

https://docs.microsoft.com/en-ca/iis/get-started/introduction-to-iis/iis-web-server-overview

Appendix
UML System Design

Unified Modeling Language

- Standard for Software Modeling
- Uses Specific Symbols
- Categorized in Diagrams

This diagram is based on:
- Application Diagram
- Component Diagram
- Package Diagram
- Class Diagram


Flexible Front-End

- Implementation with popular libraries:
  - jQuery
  - Bootstrap CSS

- Documented HTML / CSS / JS

- Various models of web pages
  - Home
  - Confirmation
  - 404 / 403

https://github.com/patricklague/ConfBusterWebServer
Appendix
Web Back-End

Customizable Back-End

- OOP PHP Implementation
- Documented PHP
- User Guide for Server Installation
  - Step-by-Step
  - Available on GitHub
- Possibilities for B2B Only Implementations
  - Scripted Submissions
  - Batched Transactions

https://github.com/patricklague/ConfBusterWebServer
Appendix
Queue System Versatility

DBMS Based Queue Access

- DBMS : DataBase Management System

- Usage of SQL Language
  - Simplify Data Interaction

- Usage MySQL Connectors
  - PHP PDO
  - Python MySQLdb

- MySQL DBMS already implements asynchronous query management
  - Simplifies Queue Traffic Management
Appendix
Queue System Versatility

Web UI  Desktop GUI  Mobile Appl.

PHP / Python / Java / C++ / …

Queue Connectors

Compute System

Job Queue
Appendix
Multi-Node Computing

Current System

Job Input System

Compute System

Job Queue

Job Input System
- Shared Node
- Dedicated Node

Compute/Queue System
- Shared Node
- Dedicated Node

Business Layer

ComputeJobPreparation.php

MailManager.php

Database Layer

DatabaseManager.php

MySQL MariaDB Database Server
Appendix
Multi-Node Computing

Job Input System

Compute Node #1

Compute Node #2

Compute Node #3

Job Queue

Gabriel Bégin
Acknowledgments

**Director**
Pr Patrick Lagüe

**Colleagues**
Xavier Barbeau
Antony Vincent

**Stéphane Larose**
- Web Server Hosting
- HTTPS Certificate
- Institutional DNS
  confbuster.ibis.ulaval.ca

**Gabriel Bégin**
Gabriel.Begin.4@ULaval.ca

Try it now!
www.ConfBuster.org