Title
Simulation of Protein Separations

Organization
RIT School of Chemistry & Material Science

Primary Contact
Paul A. Craig

Contact Email
paul.craig@rit.edu

Contact Phone
585.475.6145 (office) 585.305.5557 (cell)

Background Info
Students in the life sciences must learn to use multiple instruments to analyze protein structure and function. Some of these tools are quite advanced (i.e., expensive), which means that students from many programs do not have access either because the instruments simply are not available on their campuses, or because undergraduates are not permitted to use the instruments. To help these students learn these methods and approaches, we have developed JBioFramework (JBF), a suite of analytical simulations of protein structure and function, including 1D electrophoresis, 2D electrophoresis, Tandem Mass Spectrometry, and chemical drawing software to display protein fragments. It has been under development since 1997 using Java and has been used in multiple undergraduate course settings.

Project Description
There have been many contributors to JBF over the years, including students majoring in chemistry, biochemistry, bioinformatics and computer science. The Java code base appears to be stable and functional and can be found at https://sourceforge.net/projects/jbf/. It has more recently been moved to Github (https://github.com/RITJBF/JBioFramework), but the Sourceforge site also includes a functioning executable version of JBF. The 1D electrophoresis (1DE) simulation has more recently been updated and released in a Javascript container and it can be found here: https://people.rit.edu/pac8612/1DE/. The source code for the updated 1DE simulation will be provided by the sponsor. The project for the 2023-2024 academic year is (a) to create a web application of one of the separation simulations according to current standards for software engineering and (b) to create a shell to host the simulation you choose and enable future teams to deploy additional simulations.

Project Scope
Existing applications in JBF will need to be re-engineered from a Java application to a web application in the language of your choice. The team will need to review the existing code from the various repositories and build a functional GUI for deployment in undergraduate biochemistry courses, perhaps extending to the K-12 environment as well. Specific issues to consider include the following:

1.•F†R Æöö² æB fVVÂ öb ÆÂ Æ–6 F–öç2 &R V—FR F FVBÂ 6ò æPw design would need to be created that reflects current sensibilities.
2.•6†VÆÂâ A web interface will need to be developed that can act as a shell for the application that this team pursues, with clear instructions and hooks that will enable future teams to include additional applications under other tabs.
3.” Æ–6 F–öç2 f÷" ¤$`
a.“ DRâ The 1-dimensional electrophoresis app is the most recently developed code and it includes the ability to import data in a standard bioinformatics format called FASTA. Currently the bands that are seen on this gel contain minimal information related to band position in the simulation. The desired 1DE application will make any band on any of these simulations rich in information and connected to multiple resources on the Internet. There will also be an option to select a protein and send it to the Tandem Mass Spectrometer (item 3c) for analysis. It is highly desirable that future users (not coders) will be able customize the information links to resources of their choosing.
b.“$DRâ The 2-dimensional electrophoresis app currently accepts data in a variety of formats (.faa, .genbank, .pdb, .mmCIF). It has tools for calculating molecular weight and isoelectric point, leading to a nice separation of mixtures of large numbers of proteins (up to 3,000) in a 2D matrix based on the size and charge on the proteins. Following the separation, clicking on a protein spot brings up a small window with links for BLAST Search, NCBI Search and Uniprot Search. These buttons need to be customizable so that users can select the databases they want to explore for their proteins. There will also be an option to select a protein and send it to the Tandem Mass Spectrometer (item 3c) for analysis.
c.•DÕ2â The Tandem Mass Spectrometry application is stable and quite useful. The interface needs to be updated to reflect current usage. The TMS will accept inputs (proteins or peptides) from the 1DE, 2DE and RPLC simulations, which it will then process and analyze.
d.•&Pversed Phase Liquid Chromatography (RPLC). This application has not yet been developed, so the algorithmic foundation of the separation will need to be created based on current literature [1–4], which the sponsor will provide and interpret. Proteins are normally pre-digested into smaller fragments (called peptides) which are then separated based on their charge, size and polarity. The results of the separation will be displayed in a chromatogram as shown in the image here, where the peptides appear as peaks on the chromatogram as a function of time. Again, the peaks will be information rich, linked to resources on the Internet, with an additional link that allows the user would to click on a peptide and send it to the Tandem MS for analysis.

e.”Ö 'f–â 6°etch. This chemical drawing program will be implemented to receive peptide fragments of the Tandem MS and display their structures. The current version of JBF contains a very dated version of Marvin Sketch from ChemAxon https://chemaxon.com/products/marvin). A web implementation of Marvin Sketch can be found on the RCSB Protein Data Bank Chemical Sketch Tool page.

References

1’ . G. Boswell, J. R. Schellenberg, P. W. Carr, J. D. Cohen, A. D. Hegeman Easy and accurate high-performance liquid chromatography retention prediction with different gradients, flow rates, and instruments by back-calculation of gradient and flow rate profiles. (2011) J. Chromatogr. A. 1218, 6742–6749.
2’ Bâ Abate-Pella, D. M. Freund, Y. Ma, Y. Simón-Manso, J. Hollender, C. D. Broeckling, et al. Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. (2015) Journal of Chromatography A. 1412, 43–51.

3’ ²â 6†–æöF Â Òâ 7Vv–Ö÷FòÂ ââ Yachie, N. Sugiyama, T. Masuda, M. Robert, et al. Prediction of liquid chromatographic retention times of peptides generated by protease digestion of the Escherichia coli proteome using artificial neural networks. (2006) J. Proteome Res. 5, 3312–3317.
4’ òâ V. Krokhin, R. Craig, V. Spicer, W. Ens, K. G. Standing, R. C. Beavis, et al. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. (2004) Mol. Cell Proteomics. 3, 908–919.

Project Challenges

Domain Knowledge. Full implementation of all the requested features will require some knowledge chemistry and biology, and of instrumental methods that are used for signal detection, analysis, and display, as well as understanding how a scientist will want to use the tools. It will also require learning about proteins and how protein sequence strings are manipulated in bioinformatics settings. The sponsor has supported multiple SE teams in the past and has provided training on an ongoing, as needed basis.
Building a Simulation based on Mathematical Model. These separations are all based on mathematical models and algorithms. The team will need to work with the sponsor (and the references provided) to implement the mathematical model in a realistic instrumental interface.
Web Coding. The code base is currently in Java on both Sourceforge and Github. The Github repository will need to be refactored (not sure I’m using that term correctly) and populated with the web application resources as they are developed. As a non-programmer, I cannot assess the amount of work involved here.

Sponsor-Provided Resources

The original Java source code is available on Sourceforge (https://sourceforge.net/projects/jbf/). Compiled versions of the code can also be downloaded from Sourceforge.
The updated source code for the 1DE Java application will be provided by the sponsor.

Constraints and Assumptions

1.•@eam members have taken high school courses in chemistry and biology.
2.”¤$b —2 w&—GFVâ –â ¦ va, and most of the code has not been modified since 2016. The next generation of the JBF project will need to be written in a different language to create a functional web application.
3.•F†W&R &R ×VÇF— ÆR F '2 –â F†R ¦ va version of JBF. The web application should also contain multiple tabs (or a different interface with similar functionality) and the proper documentation to enable future coders to introduce additional simulations.
4.•F†R ¤$b vV" Æ–6 F–öâ v–ÆÂ æVVB Fò &V B æB –çFW' &WB `iles in a variety of formats (.faa, .genbank, .pdb, .pdbx, .mmCIF) and to easily pass data between and among the simulations.

Sponsor Deliverables

•”gVæ7F–öæ Â vV" Æ–6 F–öâ æB 6†VÆÀ
•”gVÆÇ’ Fö7VÖVçFVB 6öFR –â â W ×FòÖF FR v—F‡V" &W ÷6—F÷'•
•”–ç7G'V7F–öç2 f÷" 6öF–ær FV ×2 v†ò v–ÆÂ ¦ö–â F†R &ö¦V7B –â F†R gWGW&P

Proprietary Info
None. the project is open source.

Sponsor Alternate Time

Sponsor Availability Checked
true

Project Agreements Checked
true

Assignment of Rights
open_source

Attachments
SeniorProjectProposal_Paul_Craig_JBF.pdf