

3-1-2014

# Homology Modeling and Molecular Docking for the Science Curriculum

Owen M. McDougal  
*Boise State University*

Nic Cornia  
*Boise State University*

Chris Mallory  
*Boise State University*

Julia Thom Oxford  
*Boise State University*

Tim Andersen  
*Boise State University*

## Homology Modeling and Molecular Docking for the Science Curriculum

Owen M. McDougal,<sup>†</sup> Nic Cornia,<sup>‡</sup> S.V. Sambasivarao,<sup>£</sup> Andrew Remm,<sup>§</sup> Chris Mallory,<sup>†</sup> Julia Thom Oxford,<sup>§</sup> C. Mark Maupin,<sup>£,\*</sup> Tim Andersen<sup>‡,\*</sup>

<sup>†</sup>Chemistry and Biochemistry, Boise State University, Boise, ID

<sup>‡</sup>Computer Science and Engineering, Boise State University, Boise, ID

<sup>£</sup>Chemical and Biological Engineering, Colorado School of Mines, Golden, CO

<sup>§</sup>Biological Sciences, Boise State University, Boise, ID

\*Corresponding authors: C. Mark Maupin ([cmmaupin@mines.edu](mailto:cmmaupin@mines.edu)) & Tim Andersen  
([tim@cs.boisestate.edu](mailto:tim@cs.boisestate.edu))

### Abstract

DockoMatic 2.0 is a powerful open source software program (downloadable from [sourceforge.net](http://sourceforge.net)) that simplifies the exploration of computational biochemistry. This manuscript describes a practical tutorial for use in the undergraduate curriculum that introduces students to macromolecular structure creation, ligand binding calculations, and visualization of docking results. A student procedure is provided that illustrates use of DockoMatic to create a homology model for the amino propeptide region (223 amino acids with two disulfide bonds) of collagen  $\alpha 1$  (XI), followed by molecular docking of the commercial drug Arixtra<sup>®</sup> to the homology model of the amino propeptide domain of collagen  $\alpha 1$  (XI), and finally, analysis of the results of the docking experiment. The activities and supplemental materials described are intended to educate students in the use of computational tools to create and investigate homology models for other systems of interest and to train students to be proficient with molecular docking and

analyzing results. The tutorial also serves as a foundation for investigators seeking to explore the viability of using computational biochemistry to study their receptor-ligand binding motifs.

**Key words:** Homology modeling, molecular docking, computational biochemistry, DockoMatic

## **Introduction**

Advances in hardware, coupled with ever more affordable computer systems, have allowed computational experimentation that once were limited to research laboratories, to enter into classroom settings. A complement to the wide-spread adoption of molecular modeling and computational biochemistry laboratory exercises has been the development of computer software applications to facilitate homology model creation and molecular docking activities [1]. These types of applications that once required license agreements on the order of tens of thousands of dollars per year to access well documented software are now emerging through open source sites, e.g. sourceforge.net [2]. Databases of template structures may be accessed through web portals like National Center for Biotechnology Information (NCBI) using the efficient search program BLAST, and fast computer processing and cluster computing now allow the creation of homology models for biomacromolecules, from a template structure, in minutes rather than hours, days, or weeks [3,4].

The current work emerged from a week-long computational chemistry short course developed by an interdisciplinary team of science educators with backgrounds in chemistry, biology, computer science, and engineering. Students that participated in the course were introduced to software applications, theory, and molecular docking applications. The course assessment was used to refine the content and computational tools presented to students in order

to optimize the material and activities that were deemed most significant to foster student understanding. The main programs required for homology modeling and molecular docking are integrated into DockoMatic to create a user-friendly computational tool accessible across scientific disciplines.

A practical software tutorial is provided for use within the time constraints of a traditional laboratory course for introductory biochemistry, cell and/or developmental biology, or pharmacology courses. The objective is for students to learn how DockoMatic can be used as a computational tool to better understand complex biochemical processes by creating homology models, and setting up, conducting, and analyzing docking calculations. The computational exercises introduced in this tutorial allow students to visualize complex macromolecular processes using software that is accessible and affordable. The focus of the tutorial is collagen  $\alpha 1$  (XI) (Col  $\alpha 1$  (XI)). Col  $\alpha 1$  (XI) is a fibrous protein significant in thrombosis that contains an amino propeptide domain (NPP) consisting of 223 amino acids and two disulfide bonds. Using the Col  $\alpha 1$  (XI) NPP domain as the test case, the tutorial leads students through two processes: (1) homology model creation of the three dimensional structure for Col  $\alpha 1$  (XI) NPP [5], using the open source DockoMatic software [6-8], and (2) molecular docking calculations for the ligand Arixtra®, an anticoagulant medication, binding to the NPP domain of Col  $\alpha 1$  (XI) [5].

## **Homology Modeling**

In the first activity, students access the Timely Integrated Modeler (TIM) utility through the DockoMatic graphical user interface (GUI) to create the Col  $\alpha 1$  (XI) NPP homology model [8]. DockoMatic draws upon subservient software programs to facilitate user ability to load the

template protein sequence, perform a BLAST search, identify a template protein, align amino acids, generate a protein structure, and evaluate the resultant homology model for accuracy compared to the template structure. This homology model creation and validation exercise in combination with the molecular docking portion of the tutorial may all be accomplished within a two hour laboratory period with a group of students that have never used the software.

DockoMatic draws upon BLAST, MODELLER, and PyMol to identify a suitable template structure, create a homology model, and validate the model structure [4,9,10]. A set of instructor notes is provided (see supplemental materials) to lead instructors through the proper installation of the DockoMatic software and ensure subservient software programs are accessible. The homology model laboratory exercise provides an overview of the following topics: (1) publically available protein structures, (2) flow chart for homology model creation, (3) step-by-step instructions to create a homology model for Col  $\alpha 1$  (XI) NPP using DockoMatic, and (4) a procedure to validate the resultant structure by Ramachandran plot analysis, sequence alignment, and root mean square deviation from template structure (see supplemental materials). Figure 1, a flow chart included in the student procedure, outlines the steps in the homology model creation process.

### **FIGURE 1**

The learning objective for the homology modeling activity is for the student to learn how to search for and identify an appropriate template protein using a knowledge of similar biological activity for the protein and sequence comparison. The homology models created will then be evaluated for accuracy based on sequence and structure comparisons. They will achieve this objective by using DockoMatic to perform a BLAST search on a target protein, selecting an appropriate template protein structure based on established criteria (E value, % query coverage,

and sequence alignment), generating a 3D homology model, and refining and optimizing the resultant protein structure.

## **Molecular Docking**

The second activity takes students through the process of using their recently created homology model to perform molecular docking calculations. An overview of the protocol for molecular docking segues into the DockoMatic GUI. DockoMatic 2.0 uses either AutoDock 4.2 or AutoDock Vina to perform molecular docking calculations; the user may select the docking engine based on preference [8,11,12]. AutoDock 4.2 is the default docking engine for DockoMatic. A ligand structure file for the anticoagulant drug Arixtra® has been provided along with a grid parameter file (gpf), created in AutoDockTools (ADT), for the molecular docking experiment students are to perform. Students are led through the following processes: (1) accessing the required ligand (Arixtra®), target (Col  $\alpha$ 1 (XI) NPP), and grid parameter files (gpf) for the molecular docking experiment, (2) putting the files in the correct entry field in the DockoMatic GUI, (3) adjusting the parameters of the experiment to allow sufficient trials in the time available, (4) following the progress of the calculations, and (5) accessing the results of the docking trials. The number of trials is set to ten to allow the experiment to complete in under a minute, thus providing time to analyze the results. An overview is provided to students so they understand how to evaluate the resultant log files and interpret the statistics associated with docking studies. Attention is given to show and explain the “Cluster Histogram” produced as output from the docking studies. Identifying the most energetically favorable ligand binding pose from the log files leads to visualization of the result (see Figure 2). DockoMatic draws

upon PyMol for structure analysis, and students are led through the process of using PyMol [10] to first visualize Arixtra®, then Col  $\alpha$ 1 (XI) NPP, and finally the ligand bound protein complex.

## FIGURE 2

The learning objective for the molecular docking activity is for students to learn how to access ligand and receptor structure files, generate a grid parameter file that defines the ligand binding domain on the receptor, run docking jobs, review results, and visualize the lowest energy binding poses for a ligand-receptor complex using PyMol. Students are provided a visual depiction of ligand binding and they learn to use the computational tools required to assess the intermolecular attraction leading to more favorable ligand binding conformations.

## Evaluation

To evaluate the degree to which the experiment can be conducted within a two hour laboratory session, the materials were piloted at the 93<sup>rd</sup> annual meeting of the American Association for the Advancement of Science Pacific Division, Boise, ID, 2012. There were eight participants, five undergraduate, two postdoctoral researchers and one faculty member. Three of the participants had a background in computer science, three in chemistry, and two in biological sciences. After the tutorial, participants were asked about their experience (see Table 1). The strengths of the tutorial given by the participants were the detailed background material, moderate pace of student activities, clarity of instruction, and the degree of confidence that they could utilize the DockoMatic software to create a receptor homology model and perform molecular docking calculations. Weaknesses of the tutorial were listed as the desire to see more coverage of the creation of grid parameter files, usage of PyMol to manipulate and visualize the ligand to

receptor interactions, and assignment of an independent project for participants to pursue as a test of their capabilities to use the software effectively.

## TABLE 1

### Discussion

This work was motivated by software advancements that have allowed homology modeling and molecular docking experimentation to be carried out within the timeframe of a standard undergraduate science laboratory course. This software tutorial provides resources for an instructor to demonstrate the use of DockoMatic and to achieve a foundational degree of confidence for computational investigations by students. The student lab procedure includes an expanded background of computational structure biology, describing not only DockoMatic, but a broader view of the utility of BLAST, MODELLER, and PyMol. Participant feedback from the conference workshop was used to strengthen the presentation of content and guide development of the student tutorial. Based on the review of student-perceived weaknesses, a ‘grid parameter file creation segment’ was added to the student procedure. Student feedback requesting more exposure to PyMol and assignment of independent projects was deemed outside the scope of this project because PyMol tutorials are readily available on the internet and the assignment of independent projects was deemed more appropriate based on the instructor intended usage of DockoMatic. The tutorial is configured to be accomplished within a two hour laboratory session or alternatively, additional time may be spent on the analysis of output docked structures to fill a three hour laboratory session. Instructor notes detail how to download DockoMatic from sourceforge.net and install the software. DockoMatic requires a Linux-based operating system or



emulator, and runs best on PCs with a minimum of 4 gigabytes of ram and a multi-core processor. DockoMatic can run and manage jobs on a standalone PC, but is also designed to run and manage jobs on a Beowulf cluster. The student procedure consists of a step-by-step process describing how students may use DockoMatic for homology modeling, molecular docking, and results analysis. The student procedure and an instructor notes readme file for DockoMatic, Modeller, and PyMol installation are included in the supplemental materials. The student procedure can also be used as a guide for students to pursue independent investigations outside the scope of this tutorial.

### **Acknowledgements**

The project described was supported by the INBRE Program, NIH Grant Nos. P20 RR016454 (National Center for Research Resources) and P20 GM103408 (National Institute of General Medical Sciences), and the Office of Sponsored Programs at Boise State University.

### **References**

- [1] R. B. Jacob, T. Andersen, O. M. McDougal (2012) Accessible high throughput virtual screening molecular docking software for students and educators, *PLoS Comp. Biol.* 8(5): e1002499. doi:10.1371/journal.pcbi.1002499.
- [2] Sourceforge: <http://sourceforge.net>
- [3] R. Edgar, M. Domrachev, A. E. Lash (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207-210.
- [4] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T. L. Madden, (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5-9.

- [5] O. M. McDougal, C. Mallory, L. R. Warner, J. T. Oxford (2012) Predicted structure and binding motifs of collagen  $\alpha 1(XI)$ , *JBio.* 1, 43-48.
- [6] C. Bullock, R. Jacob, O. McDougal, G. Hampikian, T. Andersen (2010) DockoMatic—automated ligand creation and docking, *BMC Res. Notes* 3, 289-297.
- [7] R. B. Jacob, C. W. Bullock, T. Andersen, O. M. McDougal (2011) DockoMatic – automated peptide analog creation for high throughput virtual screening, *J. Comp. Chem.* 32(13), 2936-2941.
- [8] C. Bullock, N. Cornia, R. B. Jacob, A. Remm, T. Peavey, K. Weekes, C. Mallory, J. T. Oxford, O. M. McDougal, T. Andersen (2013) DockoMatic 2.0: A customizable application for high throughput inverse virtual screening and homology modeling, *J. Chem. Inf. Model.* in press.
- [9] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali (2000) Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.* 29, 291-325.
- [10] The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
- [11] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comp. Chem.* 19, 1639-1662.
- [12] O. Trott, A. J. Olson (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *J. Comp. Chem.* 31, 455-461.

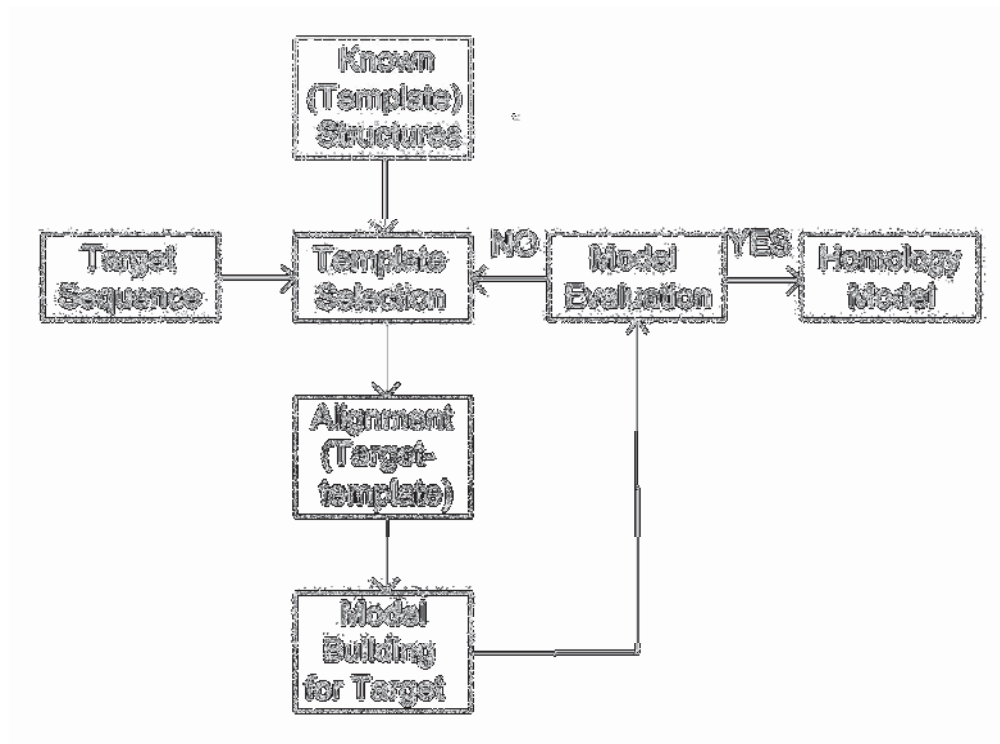
## Figure Captions

**Figure 1.** Flow chart of steps involved in generating a homology model for a protein receptor.

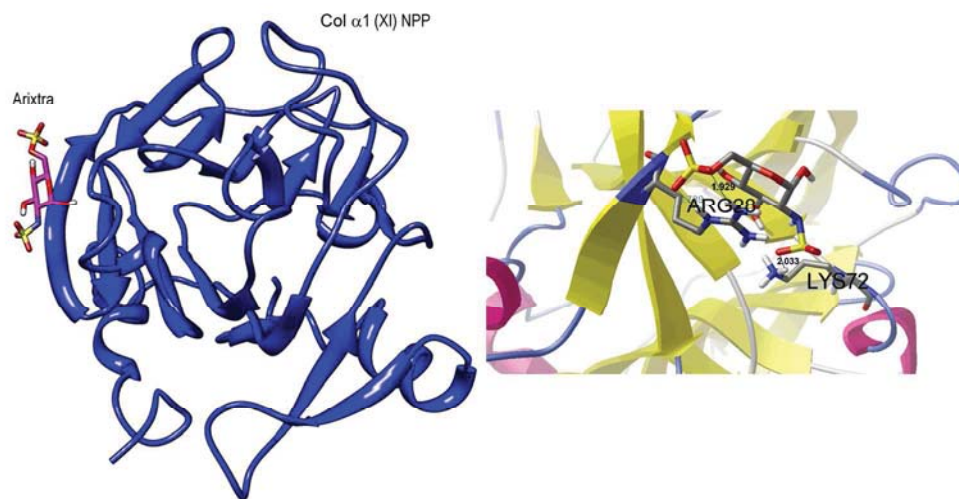
**Figure 2.** Structure of the most energetically favored binding pose for the Arixtra® ligand binding to the Col  $\alpha 1$  (XI) NPP homology model; full view of molecules (left), and expansion of amino acid residues interacting with Arixtra® (right).

### **Table Legend**

**Table 1.** Summary of participant response to a two-hour tutorial session.



88x65mm (300 x 300 DPI)



177x89mm (300 x 300 DPI)

Survey Question	Participant Response from 1 (poor) to 5 (excellent) (N=8)				
	1	2	3	4	5
Well informed of objectives	-	-	-	2	6
Expectations were met	-	-	-	6	2
Objectives were clear	-	-	-	2	6
Activities stimulated learning	-	-	1	4	3
Sufficient practice & feedback	-	-	3	3	2
Appropriate level of difficulty	-	-	2	3	3
Appropriate pace of activities	-	-	3	4	1
Instructor was well prepared	-	-	-	1	7
Instructor was helpful	-	-	-	2	6
Objectives achieved by student	-	-	1	2	5
Student able to use knowledge	-	-	-	3	5
Positive impression of tutorial	-	-	1	-	7
Overall tutorial rating	-	-	-	5	3