Connecting Common Genetic Polymorphisms to Protein Function: A Modular Project Sequence for Lecture or Lab

Christopher E Berndsen, James Madison University
Byron H Young, James Madison University
Quinlin J McCormick, James Madison University
Raymond A Enke, James Madison University
Connecting Common Genetic Polymorphisms to Protein Function: A Modular Project Sequence for Lecture or Lab

Abstract

Single nucleotide polymorphisms (SNPs) in DNA can result in phenotypes where the biochemical basis may not be clear due to the lack of protein structures. With the growing number of modeling and simulation software available on the internet, students can now participate in determining how small changes in genetic information impact cellular protein structure and function. We have developed a modular series of activities to engage lab or lecture students in examining the basis for common phenotypes. The activities range from basic phenotype testing/observation to DNA sequencing and simulation of protein structure and dynamics. We provide as an example study of the bitterness receptor TAS2R38 and PTC tasting, however these activities are applicable to other SNPs or genomic variants with a direct connection to an observable phenotype. These activities are modular and can be mixed to meet the student capabilities and infrastructure availability. The complete sequence of activities will demonstrate the direct connection between gene structure, protein structure and organism function. © 2016 by The International Union of Biochemistry and Molecular Biology, 44(6):526–536, 2016.

Keywords: genomics proteomics bioinformatics; molecular biology; protein structure function and folding; using simulation and internet resources for teaching

Introduction

As of March 2016, there are nearly 3,700 complete genome sequences across Archaea, Bacteria, and Eukaryotes deposited in EMBL, with new genomes being deposited every year. These genome sequences do not account for the full diversity of a species as there can be genetic variation within the members of species such as ability to taste certain molecules or the differences between a chihuahua and a border collie. In courses, we speak of these differences in terms of changes in gene sequence, but often do not address how alterations to the DNA sequence can change cellular function on many levels including transcriptional regulation or changes to protein structure resulting in functional changes. For students, this direct connection between gene sequence and protein function is not always obvious. For many polymorphisms, the exact cause of the functional difference is not clear, often because of the lack of structural information on the protein. Prediction of protein structure is an area of avid research because of the connection between protein structure and protein function. There are many publicly accessible structure prediction servers and even games increase to increase our knowledge of the breadth of protein folds and push toward a complete understanding of the human proteome.

Here we present a series of lab and lecture activities that can be used to demonstrate the connection between phenotype, genotype, and protein function using human bitter taste perception and the TAS2R38 receptor as an example. The TAS2R38 gene contains three comigrating nonsynonymous SNPs resulting in A49P, V262A, and I296V amino acid changes in the protein receptor causing an undescribed functional change and the observed phenotypes [1]. Persons who are AVI cannot taste the synthetic molecule phenylthiocarbamide (PTC), while individuals who
have the PAV sequence taste PTC intensely [1]. There have been links between this receptor and tasting glucosinolates, which are found in some vegetables [2].

The activities outlined in these inquiry-based activities are interconnected yet modular so that they can be fit to courses of various levels and scope. Because this project can be done with polymorphisms with no identified molecular basis, the data generated can be suitable for inclusion as part of a student research thesis or publication. The goals of these activities were:

- Demonstrate the connections between genotype, phenotype, protein structure, and protein function.
- Have students comprehend different types of scientific data (survey, theoretical, quantitative).
- Expose students to computational and wet lab analysis.
- Have students gain experience with protein structure prediction and modeling.

## General Lab Overview

The subsequent modules are designed to fit into the modern undergraduate laboratory classroom equipped for lecture, wet lab, and computational-based activities (Fig. 1). Wet lab components employ equipment and techniques commonly encountered in molecular and biochemical research. In particular these modules focus on basic skills such as extraction, amplification and genotyping of nucleic acids. Computer-based modules have been designed to emphasize the egalitarian nature of modern bioinformatics analysis. Freely available online tools have made sophisticated computational analysis available to anyone with an Internet connection. These modules emphasize several specific tools for obtaining and annotating sequence information such as the UCSC Genome Browser, the ApE Sequence Editor and the Online Mendelian Inheritance in Man (OMIM) Database (see Supporting Information Table I for a list of all programs). Additionally, students are challenged to explore a variety of available online tools for protein modeling analysis such as Dali, Phyre² COACH, and YASARA. Basic principles explored in each module are highly adaptable and can be modified to fit the preferences, allotted class time and available resources of individual instructors and courses. Detailed protocols for each laboratory module can be found in Supporting Information Document 1 and example activities to teach prediction of structure and function of using online web servers are included as Supporting Information documents.

### Analyzing Genetic Variability and the Subsequent Effect on Phenotype

We have developed several hybrid wet lab and computer-based modules for students to extract DNA and analyze variability within their own genomes to be used to determine the molecular basis of a common phenotype variation, bitter taste perception. Class data obtained from this first set of modules can be used in subsequent modules using molecular modeling to characterize the protein basis for polymorphic phenotypes. Figure 1 lists a workflow of the possible paths students can follow during these modules.

#### Hybrid Molecular Biology Module #1: DNA Extraction, Amplification and Analysis

Prior to this lab activity, students have received lecture information on topics generally related to this lab including DNA structure/function, composition and variation of genomes and replication/amplification of DNA. Additionally, students have engaged in prior lab activities analyzing DNA using PCR amplification and agarose gel electrophoresis as well as introductory activities exploring the UCSC Genome Browser [3] and the ApE sequence editing software (Wayne Davis, personal communication [4]). A pre-lab lecture reviews concepts of genomic variation, the molecular and biochemical basis of taste perception, as well as ethical considerations of human genetic and genomic analysis. During this module students will extract their own genomic DNA and use it as template for a PCR reaction and subsequent genotyping assay for a common SNP in the TAS2R38 gene associated with bitter taste perception. Computational components of this module will focus on DNA sequence analysis of the TAS2R38 gene.

In the first wet lab component, students extract their own DNA from hair and buccal samples using an efficient
and economical alkaline lysis technique [5]. This protocol illustrates fundamental differences in nucleic acid extraction versus purification emphasizing that crude DNA extracts are sufficient for PCR analysis. During a long incubation step in the DNA extraction protocol, students prepare agarose gels for analysis of restriction fragment length polymorphism (RFLP) PCR products in a subsequent lab. Students will also begin computer-based DNA sequence analysis of the \textit{TAS2R38} gene during this incubation period.

In the first computational lab component, human gene sequence and genetic variation data are obtained from the UCSC Genome Browser and annotated using the ApE sequence editor and information obtained from the OMIM database [6]. During this computational module, students use the genome browser to navigate to the \textit{TAS2R38} gene in the most recent human genome assembly and configure the visualization tracks to view available genetic variation data for this region (Fig. 2A). Students then navigate to and download the \textit{TAS2R38} genomic DNA sequence from the genome browser and import it into the ApE sequence editing software where it will be subsequently annotated with genetic variation data tracks obtained from the browser. Further information pertaining to genetic variants resulting in altered protein structure and function are obtained from the OMIM database and annotated in their sequences. Finally, students annotate their sequences with instructor-provided primer information that they will use for downstream analysis of their own \textit{TAS2R38} alleles (Fig. 2B).

Upon completion of the DNA extraction protocol and computational analysis, students prepare PCR reactions to specifically amplify and genotype their own \textit{TAS2R38} alleles from hair and buccal DNA preparations. Students set up PCR reaction using forward primer provided by the instructor containing a single mismatch site that will introduce an allele-specific restriction site to be utilized for downstream RFLP analysis. PCR reactions are loaded onto a preprogrammed thermocycler. Products are collected and stored by the instructor for analysis in the following lab meeting.

Activities in module 1 are reinforced using individual and group assessments. Individual students are required to recreate the computational activities in this module outside of class while elaborating on their findings by answering several assessment questions. These questions are designed to assess the understanding of computational information gathered during the module and how this information integrates into wet lab portions of the module (Supporting Information Document #1). For example, students are challenged to explain how a mismatched forward PCR primer can be used for downstream genotyping analysis. Additionally, students in groups of three met outside of class to prepare a mini proposal outlining the multiweek genotype/phenotype experiment. The proposal serves as an effective
assessment of big picture understanding of the comprehensive experiment as well as an opportunity for the instructor to provide feedback for students to incorporate into downstream assessments associated with this lab activity.

**Hybrid Molecular Biology Module #2: Genotype/Phenotype Determination**

After an introduction to the molecular basis of bitter taste perception explored in the first module, students will elaborate on these activities by completing genotyping assays of their own TAS2R38 alleles, conducting bitter taste phenotype experiments, and using sequence alignments for molecular phylogeny analysis. Computational portions of this lab module utilize additional tools available in the UCSC Genome Browser and the ApE sequence editing software.

Similar to module #1, the second module is a combination of computer-based activities designed to run in tandem with wet lab activities. Prelab lecture material reviews concepts in PCR and RFLP analysis. Students retrieve their PCR products from the previous class meeting and set up RFLP digest designed to differentially digest TAS2R38 alleles [7]. During restriction digest incubations, students conduct bitter taste phenotype experiments. In the first phenotype experiment students use PTC tasting paper to determine their bitter tasting ability. Students then taste various vegetables in the brassica family containing glucosinolates, which interact
with the TAS2R38 receptor and are associated with bitter tasting (Fig. 4A). Students anonymously score the vegetables rating their bitter taste on a scale of 1–10. Bitter taste phenotype data for the class is collectively compiled using a public Google Sheet preformatted by the instructor. These data are subsequently integrated with class genotype data following gel electrophoresis of RFLP PCRs (Fig. 3A).

Following bitter taste perception assessment and restriction digest incubation, students analyze RFLP PCR fragments using agarose gel electrophoresis analysis. Banding patterns visualized on 2% agarose gels indicate each student’s genotype for a SNP in the TAS2R38 gene associated with bitter taste perception (Fig. 3B). TAS2R38 genotype data for the class is collectively compiled on the preformatted public Google Sheet and integrated with bitter taste phenotype data (Fig. 3A).

During the 30-min gel electrophoresis analysis experiment, students conduct computational sequence alignment analysis of TAS2R38 alleles using the alignment tool in the ApE software. Students first align and compare human taster and non-taster alleles of the TAS2R38 gene. Students then obtain TAS2R38 sequence information for several non-human primates from the UCSC Genome Browser for alignment analysis with the human taster and non-taster alleles. Phylogenetic analysis of TAS2R38 alleles enables students to construct hypotheses about the evolution of bitter taste perception in primates (Fig. 4B).

Following module 2, we took this opportunity to discuss issues with the phenotype survey that could decrease the accuracy or reliability of the results such as subjective ratings of vegetable taste and sample size. Individual assessments
are also used to challenge students to recreate and interpret sequence alignment analysis outside of class. For example, students are asked to use their phylogeny results to determine the likelihood of the human nontaster allele arising before or after the human lineage split from other primate species. These assessments challenge students to use higher order cognitive domains requiring application and analysis of molecular and computational data.

Modeling the Protein Basis for Polymorphism Phenotype

When connecting genotype to phenotype, there is often a gap in the student’s understanding stemming from the missing molecular basis for the phenotype. To fill in this gap, we have developed several computer-based modules for students to discover the molecular basis by modeling the protein structures and then using the structures to propose the differences in function. Rather than explicit instructions, students are given free rein to choose from a variety of possible databases and programs for molecular modeling activities. A workflow of the possible paths students can follow is shown in Fig. 1 and a list of web-based or freeware programs is provided as in Supporting Information Table I along with a brief explanation of the use. Every effort was made to use web-based or free software to make this activity accessible to multiple audiences who do not have local access to a high performance computer and to keep pace with new trends in educational computing such as Chromebooks and iPads.

We perform these activities with students who only have limited experience with molecular biology and biochemistry and upper level students who have taken previous courses in molecular biology and biochemistry. We also present example protein modeling activities that were used in an introductory Biochemistry course of 80 students demonstrating the use of these modules in a variety of settings and student skill levels. Before beginning the computational modules, we recommend a brief introduction or review of the levels of protein structure, amino acid structure and basic properties, and general overview of protein/enzyme function. If performed with the molecular biology modules, this is a good time to reiterate the connection between DNA sequence and protein structure which dictates protein function.

Data Analysis Software

In the course of modeling the protein structure and function, students will generate a wide variety of data from numerical outputs of secondary structure propensities to protein structure ensembles and simulation files. To analyze all the data, students will need to acquire familiarity with at least a structure viewing program and a plotting program. For viewing protein structures, we recommend local installation of a program that can display PDB files as an interactive graphic and permits measurement of

FIG 5

Student generated TAS2R38 modeling results. Example of a student generated model of the TAS2R38 receptor with sites of the PTC tasting polymorphism shown in dark blue (a). Example of a student generated alignment of a tasting model (Blue) and a Non-taster model (green) with PTC bound to taster shown in light blue and nontaster in yellow (b). Example of a student generated alignment of a tasting model (Red) and a nontaster model (Blue) with PTC bound to taster shown in Yellow and nontaster in Green (c). Example student generated figure showing a peptide of Gustducin (shown in yellow/orange) bound to a tasting (left) and a non-tasting (right) receptor (d). [Color figure can be viewed at wileyonlinelibrary.com]
**TABLE 1**

**Student feedback on protein modeling modules**

<table>
<thead>
<tr>
<th>What did you learn in doing this project?</th>
<th>What was the most challenging aspect?</th>
<th>Has anything you learned from the project changed how your approach research or think about science?</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;I gained a solid understanding of many basic computational techniques over the course of the project from secondary structure prediction to homology modeling and docking simulations. It helped me realize the strengths and weaknesses of computational studies and their complementarity for wet lab results.&quot;</td>
<td>&quot;Developing a question that could be answered given the data that we were capable of generating and relating the two together was challenging.&quot;</td>
<td>&quot;The scientific process is more complex than I realized before doing this project. It is relatively common for scientists to obtain data that can change the direction of their research because the research process involves seeking information that is not yet known.&quot;</td>
</tr>
<tr>
<td>&quot;I learned how computer simulations such as structural modeling and ligand docking can lay the foundation for follow up experiments by providing preliminary data on how proteins interact with possible ligands...I also learned the theory behind these programs and how to properly use the data to support or disprove a hypothesis.&quot;</td>
<td>&quot;Analyzing the large amounts of data these programs produced was also challenging as we needed to figure out how the data was calculated, if it was significant, and how it was relevant to the overall goals of the project.&quot;</td>
<td>&quot;The project definitely furthered my understanding of enzymology and different ways to think about mechanisms. It also provided me with the knowledge of a few more computational tools that I feel comfortable in employing if I feel it’s necessary.&quot;</td>
</tr>
<tr>
<td>&quot;I learned some basic uses of YASARA and gained a deeper appreciation of how difficult it can be to answer a seemingly simple question using molecular modeling. I also learned how to develop future experiments based upon molecular modeling data.&quot;</td>
<td>&quot;The vast majority of my time was consumed in ensuring I was using the correct file types, formatting input files, and troubleshooting the programs and software I was using.&quot;</td>
<td>&quot;The modeling project changed my outlook on the usefulness and prevalence of in silico studies and gave me an invaluable basic skill set that I can draw from in my future studies. Most other students that I’ve interacted with have had little to no experience with computational methods since they are not widely employed unless one does research in the field...Now, I look for areas of my research that can be supplemented with in silico studies to help strengthen my data or guide future experiments.&quot;</td>
</tr>
</tbody>
</table>

Distances, angles, and so forth. There are many available (Supporting Information Table SI) but we recommend PyMOL [8], YASARAView [9], or SwissPDBViewer [10] which are freely available, have cross-platform compatibility, and have reasonable user manuals. Many of these programs are extendable to include additional features such as molecular dynamics or structural analysis for more advanced users usually for a fee. If you do not wish to install a program, there are cloud-based PDB viewers such as Jolecule, which works well on Chromebooks and iPhones [11], and browser-based viewers such as Cn3D [12]. For numerical data, we recommend Google Sheets or Plot.ly as free, simple plotting programs which can produce publication quality figures and are easy for students to collaborate.
during data analysis. Moreover, if students are working in groups, it is helpful to have a cloud-based drive provided by the institution or via a commercial provider such as Dropbox or Google Drive for students to store the data.

**Computational Module #1: Protein Sequence Analysis and Prediction.** Some features of protein structure can be predicted based on the primary sequence. These features include ligand/substrate binding sites, cellular localization sequences, and secondary structure. After translating the DNA sequence to a protein sequence students should identify the effect on protein sequence such as amino acid substitution or truncation of the protein. A variety of databases cataloging information on SNP changes such as UniProt [13] and DMDM [14] to protein sequence and that provide an overview of protein structure/function are available (Supporting Information Table SI). After identifying the change in the amino acid sequence, students can then identify sequence motifs and localization features using the multitude of servers accessible from ExPASy [15]. An example of this activity is included as in Supporting Information Document 1. If the polymorphism occurs near to a predicted functional region this suggests a basis for the defect. This information can simplify the tertiary structure modeling, assessment and analysis as it may only be necessary to model a single domain, rather than the entire protein. However, these programs do not consider the 3D arrangement of amino acids, so we recommend to not rely solely on the results from a single server for prediction of the molecular basis for the phenotype. Where time and infrastructure permit, we highly recommend modeling tertiary structure and exploring functional properties of the protein. After completing this module, students are expected to understand the connection between DNA sequence and protein sequence, primary and secondary protein structure, and the limitations of interpreting mutation data based solely on secondary structure prediction.

**Computational Module #2: Structural Modeling.** For proteins with an unknown structure, a homology model is useful to see how the amino acids may orient in 3D space. Homology models use known structures of similar sequence to generate a proposed tertiary structure for a protein of unknown structure. Many existing programs predict protein tertiary structure (Supporting Information Table SI) and the accuracy of the result depends on many factors including similarity to other structures, number of known template structures, and the post-modeling refinement of structure done by each program. For most homology modeling servers and programs, a protein sequence in FASTA format is sufficient to begin modeling. In addition to the structural model, some programs like I-TASSER [16–18] and Phyre² [19] will return information on function or potential ligand binding regions, which can aid students in their description of functional defects. An example activity for homology modeling is included in Supporting Information Document 1. We encourage students to align and compare models from more than one homology modeling program to get a sense of how well they can trust their model. If all the homology models align well, then the structure is plausible, if the structures are distinct, then students will need to use additional methods to support their structure such as molecular dynamics, or break the sequence into individual domains/motifs. Students quickly find that by the traditional methods of structural evaluation such as MolProbity [20, 21], homology models can show poor statistics even if the modeling program says the structure is high quality. Manual inspection of the structures often reveals bad bond angles and poor packing of amino acid side chains, which contributes to the poor statistics. Energy minimization or brief molecular dynamics simulations can often fix these problems which will help with later analysis.

We find this is a good time to discuss why differences appear between the types of programs and discuss the challenges in predicting protein structure and the protein folding problem [22, 23]. For some polymorphisms, the resulting structures will be distinct and will immediately suggest the basis for the phenotypic differences. However, for many non-lethal polymorphisms the structural differences may be subtle and the students may need to perform additional steps to identify the basis for the phenotype. After this module, students should have an understanding of how a homology model is constructed and the importance of different type of chemical interactions to a stable protein. Also, students should be able to describe the features of a well-folded protein and compare and contrast a sequence motif (such as those identified in module 1) versus a protein domain.

**Computational Module #3: Substrate/Ligand Binding.** Changes in protein/ligand interactions provide an obvious way for functional differences to translate into altered phenotype. Identifying ligand binding pockets can be accomplished through several methods. The simplest way is to use the homology model templates and/or the Dali server [24] to identify similar proteins with ligands or substrates bound. These existing structures can be aligned to the students’ model in PyMOL or YASARA to identify which amino acids form the binding pocket. More advanced methods include using the COACH server, which effectively performs the same task described above and tends to return a greater number of possible functions [25]. An example using both the Dali server and the COACH server is included in Supporting Information Document 1. Once the binding pocket is identified, the ligand can be fit manually in the pocket using a viewing program like YASARA, which allows independent movement of an individual molecule within an ensemble of molecules.

A better method is to computationally dock the ligand using programs such as Autodock VINA [26] or SwissDock.
[27, 28] (Supporting Information Table SI). In general, these programs simply need the protein structure in PDB format, the structure of the ligand in PDB or SDF format and an idea of the binding pocket location (one residue is often enough). If the ligand is large (>100 atoms) such as DNA or a protein, the HADDOCK web server is recommended [29, 30]. These programs divide the structure into a 3D grid and place the ligand in each box of the grid. The interaction energy between the protein and ligand is then calculated for each placement to determine the best docking site. The result is a PDB file(s) with the protein and ligand bound and a measure of the interaction energy(ies) for students to manually inspect and compare. Two good controls for quality of the docking are for students to have the ligand dock to a known noninteracting protein and a known interacting protein. These two experiments show the lower limit of accuracy with the dock to the noninteracting protein and how well the docking program can reproduce a known structure.

Whether docking is performed manually or computer aided, we take this opportunity to review the types of non-covalent interactions that occur in biological systems (i.e., hydrogen bonding, van der Waals, etc). If the target protein is an enzyme, students may predict how changes in the interaction with the substrate may alter the chemical or kinetics mechanism. After completing this module, students should have an understanding of the role of ligand binding pockets or active sites in protein function. They should be able to explain how the surface composition of these regions contributes to substrate/ligand binding.

Computational Module #4: Molecular Dynamics. A common second step after generating a homology model or docking a ligand is to perform molecular dynamics to allow the protein or protein-ligand complex to reach a stable conformation, which for more advanced applications permits better prediction of the pH, value, interaction energies, and so forth. Traditionally molecular dynamics simulations with proteins has been out of the reach of most institutions, however a number of web servers now allow limited simulation times (~5 ps to 10 ns) to be performed free of charge (Supporting Information Table SI). These servers attempt to energy minimize the protein structure to increase the accuracy of the final structure. Some servers introduce solvent to further improve the structural accuracy. We have included an example activity using MolProbity and energy minimization servers in Supporting Information Document 1. We highly recommend reading the documentation for the programs before delving too deep into simulation exercises; however the default settings for most of these servers are fine for most biomolecules. Local installation of a simulation program is also possible and many programs can perform simulations at a rate of 2–10 ns day\(^{-1}\) on consumer level computers. Some features of the protein mechanism such as large conformational changes may be out of the capabilities of a simple dynamics simulation and require some guidance or steering to explore these possibilities [31]. Regrettably, steered molecular dynamics are not available through free or web-based services to our knowledge; however there are several programs that allow instructors and students to perform this type of simulation (Supporting Information Table SI).

From any type of simulation, this is a good time to discuss with students which chemical properties are being taken into account in the simulation (i.e., bond lengths, atomic interactions, bond angles) while also discussing the assumptions (i.e., solvent/ion concentration and type, oligomeric state, no unusual chemical properties). From molecular dynamics simulations, students can compare the pH values of catalytic residues, flexibility of key structural features, and overall structural stability. After this module students should understand the role of dynamics in protein function.

Computational Module #5: Flexibility Analysis. Dynamics and flexibility are underappreciated effects of genetic mutation on protein functionality. Homology models generated by students are single snapshots of conformations the protein could adopt, however, these are likely not the only conformations adopted by the protein. Determination of protein flexibility can come from molecular dynamics simulations like those described above which often report RMSF values or B factors, or from analysis of segments of the sequence/structure and predicting mobility based on historical trends for the amino acids (Supporting Information Table SI). This latter method does not require a structure and the results are often consistent with more involved methods of analysis. However, this method does not perform well due to the lack of information on the spatial organization of the interface contained in the protein sequence. We have included an example activity using the CABS-Flex server in Supporting Information Document 1. Examples of Student Data from modeling of TAS2R38.

Collectively, students were asked to use the modeling modules described above to explain the molecular basis for PTC taste perception. Several examples of student-generated modeling of the TAS2R38 protein receptor are shown in Fig. 5. After comparison of the structures, students found the taster and nontaster proteins to be largely the same and the three mutations were spread across the structure. For example, the P49 mutation is located on the other side of the protein from the other mutations (Fig. 5A). This suggests the functional differences are caused by subtle structural changes or by alterations in function that extend beyond the basic structure. Each group diverged at this point taking separate paths to identify these functional differences. One group verified their structure versus the JPet prediction of secondary structure, which supported the homology model showing no gross structural changes. The group then proceeded to energy minimize the
homology models in YASARA to improve the model quality and then docked PTC into the ligand pocket of the receptor. Ultimately, the students suggested that the entry tunnel to the binding pocket and the PTC binding pocket itself were affected (Fig. 5B). A second group also docked PTC to the binding pocket finding a decreased affinity in the nontaster structure of TAS2R38 and the observed differences in one of the intracellular loops, which they proposed affected downstream signaling. They proposed that these two distinctions combined led to the defects in PTC tasting (Fig. 5C). A third group decided to produce additional homology models including doing structural refinement to a membrane bound receptor model. This group ran molecular dynamics simulations of 2 to 4 ns in YASARA Structure finding that having the protein in the membrane improved the quality of the structures by MolProbity analysis. Docking of PTC to the ligand binding pocket revealed similar interaction energies but with slightly tighter interactions in the taster model of TAS2R38. These students observed significant perturbation in the refined structure caused by the A262/V262, which is near the putative G-protein binding site. Another group then generated a homology model of Gustducin, the G-protein thought to interact with TAS2R38, and docked the interacting sequence of Gustducin to TAS2R38 (Fig. 5D). Ultimately, this group suggested that the Gustducin peptide would not fit in this pocket of TAS2R38 leading to the observed phenotype. These examples of student groups’ modeling results are in agreement with two recent computational studies of TAS2R38 structure and ligand binding, however the proposed defects in Gustducin binding were a novel proposal [32, 33]. These activities also demonstrate the open-ended nature of the molecular modeling modules.

Student Feedback to Computational Modules

In a post-course questionnaire students were asked to reflect on what they learned from performing molecular modeling, what challenged them and whether their perceptions of how scientific data are collected changed (Table I). A clear theme from the student feedback was the difficulty in determining how to use the modeling tools and how to use the data that were generated by these tools. Minimal direction for the individual modeling tools was provided to students in an effort to push the groups to pursue their own interests, thus this criticism was anticipated. In a larger lecture format, more instruction is provided such as that given in Supporting Information Document 1. Each instructor should determine the level of direction appropriate for the audience and the instructor’s comfort with these activities.

Potential for Investigation of Other Loci

Here we use the example of the polymorphic TAS2R38 gene modulating taste perception of the bitter compound PTC. However, these modules can be applied to a broad range of polymorphic loci associated with phenotypic change offering a high degree of versatility to instructors. Projects can easily be tailored to meet the specific needs of each class based on factors such as learning objectives, student interest, and available resources. Naturally, the sensitive implications of human genotyping call for the use of discretion in the selection process. In particular, variants associated with clinical conditions and other phenotypes with sensitive implications should be avoided. In an effort to demonstrate the broad range of loci with potential for investigation, a variety of literature and tools have been utilized to curate five suitable loci taking into account the aforementioned considerations (Supporting Information Figs. S1–S5; [34–38]). These example loci include SNPs as well as variable number tandem repeats (VNTRs), and represent intronic, exonic, and intergenic regions. Brief descriptions of these loci are found in Supporting Information Document 1. Summarization of each alternative locus including live UCSC Genome Browser links and SNPedias or Ensembl genome variant links are included in Supporting Information Table SII.

Summary

Here we have outlined a series of inquiry-based wet bench and computational class activities designed to engage undergraduate students in molecular modeling to determine the basis for a phenotype arising from genetic variation. Our study uses the well described taste receptor TAS2R38 and the ability of bitter taste perception to demonstrate these modules; however, these activities are applicable to other SNPs or genomic variants with a direct connection to a known phenotype. Further, these lab activities are modular and can be mixed to meet the specific capabilities of students and resources available to instructors.

References


536

Connecting Common Genetic Polymorphisms to Protein Function