American Journal of Pharmaceutical Education **2016; 80 (9) Article 147.**

TEACHERS' TOPICS

A Bioinformatics Practicum to Develop Student Understanding of Immunological Rejection of Protein Drugs

Lauren A. O'Donnell, PhD, Wilson S. Meng, PhD, Benjamin J. Andrick, PharmD, Alexa M. Borello, PharmD

Mylan School of Pharmacy, Duquesne University, Pittsburgh, Pennsylvania

Submitted January 18, 2016; accepted August 29, 2016; published November 25, 2016.

Objectives. To design and implement a bioinformatics exercise that applies immunological principles to predicting rejection of protein drugs based upon patient genotype.

Design. Doctor of pharmacy (PharmD) students used the Immune Epitope Database, a freely available bioinformatics tool. Over a 2-week laboratory, students interrogated whether a protein drug would be predicted to induce an immune response based upon patient genotype. Results were presented at the last laboratory session, and students completed reports discussing their findings.

Assessment. Pre-lab quizzes and a final report were graded. Students answered questionnaires assessing perceived learning gains. To determine the impact on student understanding of immunity against protein drugs, the quality of student data analysis and comparisons to class data were graded. Independent measures of student learning demonstrated that students developed a greater understanding of how patient genotype could contribute to treatment failure with protein drugs.

Conclusions. This study indicates that questions related to clinical immunology can be posed using bioinformatics tools.

Keywords: bioinformatics, anti-drug antibodies, biologics, protein drugs, major histocompatibility complex

INTRODUCTION

New protein drugs, including monoclonal antibodies and other recombinant proteins, are being approved for treating a variety of disease states. In 2015 alone, the Food and Drug Administration approved protein drug therapies for treating a variety of conditions, including asthma, plaque psoriasis, high cholesterol, high-risk neuroblastoma, and multiple myeloma as examples.[1] Immunosuppressant protein drugs (eg, etanercept, infliximab) represent 3 of the top 10 revenue-generating drugs in the United States in the past year.[2] As protein drugs become increasingly common in the marketplace and are applied to a wider variety of disease states, there will be an increased need for pharmacists to have an understanding of the factors that may contribute to the success or failure of protein drug therapy. Although many variables may contribute to treatment failure with protein drugs, one known cause is the development of anti-drug antibodies (ADAs). Protein drugs have the potential to be recognized by the immune system, despite quality control efforts to minimize immunogenicity through engineering of amino acid sequences that would be less likely to be immunoreactive. ADAs may then lead to neutralization of the protein drug by the immune system, and thus reduce efficacy or safety. In a cohort of rheumatoid arthritis patients treated with infliximab, an anti-tumor necrosis factor (TNF) monoclonal antibody, Wolbink and colleagues observed that 69% of nonresponding patients had developed ADAs against the drug while only 36% of responding patients had developed ADAs.[3] The presence of ADAs is associated with low plasma concentrations of the drug, which is a negative indicator for therapeutic outcomes.[4] Although it is not yet possible to predict which patients are most likely to mount an immune response against a protein drug, analyzing the potential genetic and environmental factors that contribute to immune recognition and ADA development against this drug class will be important.

Pharmacy students often struggle to understand complex immunological processes and the application of those processes to clinical problems. In order for students

Corresponding Authors: Lauren A. O'Donnell, 441 Mellon Hall, 600 Forbes Ave., Division of Pharmaceutical Sciences, Mylan School of Pharmacy, Duquesne University, Pittsburgh, PA 15282. Tel: 412-396-5133. Fax: 412-396-4660. E-mail: odonnel6@duq.edu. Wilson S. Meng, 413 Mellon Hall, 600 Forbes Ave., Division of Pharmaceutical Sciences, Mylan School of Pharmacy, Duquesne University, Pittsburgh, PA 15282. Tel: 412-396-6366. Fax: 412-396-4660. E-mail: meng@duq.edu

to dissect the role of the immune system in interactions with protein drugs, the students require an understanding of how proteins are processed and "viewed" by the immune system. In particular, the role of the major histocompatibility complexes (MHC) is often difficult for students to master. These molecules, also referred to as human leukocyte antigens (HLA molecules) in humans, are responsible for presenting fragments of foreign proteins for recognition by immune cells. The complexity and diversity of MHC alleles dictates whether individuals will respond to an exogenous protein, such as a protein drug, as foreign. MHC genes are characterized by substantial polymorphism (there are approximately 2000 known alleles), with certain genotypes appearing more often in specific populations or in association with particular disease states. In the case of ADAs, MHC Type II (MHC-II) alleles are especially significant in determining whether antigen-presenting cells (eg, dendritic cells, macrophages, B cells) will present epitopes from the protein drug to helper T cells, subsequently leading to antibody production. Briefly, foreign proteins are engulfed by antigen-presenting cells and broken down into peptide fragments in the acidic environment of the phagolysosome. Upon fusion of the phagolysosome and endosome, the peptide fragments from the foreign protein can be loaded onto the MHC-II molecules. The loaded MHC-II molecules are then transported to the surface of the antigen-presenting cell, where the peptide/MHC-II complex is displayed to naïve helper T cells. If the helper T cell recognizes and binds to the peptide/MHC-II complex, the helper T cell subsequently activates other immune cells, including the activation of B cells to produce antibodies against the foreign protein. Binding of the foreign peptide, or epitope, to a particular MHC-II molecule is specific, with the amino acid sequence of both proteins contributing to the affinity of the interaction. Thus, the diversity of an individual's MHC-II alleles determines the repertoire of antigens to which the helper T cells can respond.

The Immune Epitope Database and Analysis Resource (www.iedb.org) is a freely accessible bioinformatics website supported by the National Institute of Allergy and Infectious Diseases.[5] The IEDB offers a collection of search tools that can be used to ask immunological questions about epitope processing, MHC presentation, and prediction of T and B cell epitopes. The database is curated from published experimental data from immunological assays as well as current, updated data from author submissions.[6] Based on the collected experimental data, users can interrogate whether a given amino acid sequence would be likely to bind to a particular MHC-II molecule. In querying the database, the user can choose to examine their protein sequence within the context of specific MHC-II alleles, thereby providing an opportunity to compare immunogenicity of a protein across MHC genotypes. By comparing a protein sequence of interest to known epitopes within the database, IEDB provides an initial step toward predicting whether a protein sequence has the potential to be presented by an MHC-II molecule and induce an immune response.

The goal of our study was to design a bioinformatics practicum for pharmacy students that would deepen their understanding of immunogenicity and allow them to apply those concepts to a clinically related question. By using the IEDB website, students were asked to predict whether a protein drug would be presented by a specific MHC genotype, to determine whether the epitopes identified within their drug shared sequence homology with known allergens, pathogens, or self-antigens within the database, and to interpret how their findings could impact on therapeutic outcomes. As science curriculum standards move toward problem-based and inquiry-based laboratory design, this exercise provides an opportunity to ask questions about the interactions of an important class of drugs with the immune system, while improving basic understanding of complex immunological processes.[7]

DESIGN

The bioinformatics practicum took place within the Ability-Based Laboratory Experience (ABLE) course for second-year PharmD students. The ABLE is a mandatory laboratory series that incorporates clinical and applied subjects from current course work. Students are enrolled in the ABLE series throughout the first through third years in the program. The bioinformatics practicum was placed within the spring semester of the second year in order to coincide with the infectious disease course, where immunity is frequently discussed, and with the rheumatology course, where many of the protein drugs are introduced. Within the rheumatology course, approaches to engineering protein drugs are discussed, with emphasis on differences between chimeric, humanized, and fully human monoclonal antibodies.[8] Basic immunology is taught within the Human Physiology and Pathology course in the first year, so the practicum occurred roughly 1 year after the concept of MHC molecules and antibody production had been taught in the classroom.

Laboratory sessions for the practicum met for 1 hour in week 1, and for 2 hours in week 2 (total of 3 contact hours with the instructors). Five laboratory sections were included in the study (n=150 students). Students were required to bring a laptop or tablet with Microsoft Word installed and Internet accessibility. In the first week, students were assigned 2 prelaboratory videos that reviewed the immunology background and the question posed by the laboratory exercise. Video 1 included the basics of

MHC function, including a comparison of MHC-I and MHC-II, as well as brief description of the genetic diversity at the MHC alleles (Video 1). Video 2 described the overall clinical issue of anti-drug antibodies in protein drug therapy, as well as an outline describing the tasks the students would perform in the laboratory (Video 2). At the start of the first laboratory session, a quiz was administered to students to test their understanding of the content in the prelaboratory videos. The instructors then explained the expected outcomes and learning objectives of the laboratory session (Table 1) and presented a short tutorial (roughly 10 minutes) to explain the IEDB interface, the data output that would be generated from the first analysis of the database search, and a worksheet that would be used to record and annotate the initial dataset.

Each laboratory section was assigned a MHC-II allele to analyze over the 2-week laboratory. MHC-II alleles were selected by the instructors to include alleles associated with the development of autoimmune disorders such as rheumatoid arthritis (DRB1*0401, DRB1*0101 and DRB1*1001) or alleles without an association with rheumatoid arthritis (HLA-DR1*0103 and HLA-DR1*1502).[9] These MHC alleles were selected in order to include patient populations that would be likely to receive protein drugs, such as the TNF inhibitors, in the class dataset. Within each laboratory section, students paired off independently for the exercise. Students chose 1 of 12 protein drugs that could be queried, which were displayed from a Microsoft Excel file that was projected in the room for the duration of the laboratory. Each pair of students picked a protein drug of their choice, and entered their names in the spreadsheet next to the corresponding drug so that each drug was selected only once within a laboratory section. After choosing a protein drug to analyze, students downloaded a corresponding Microsoft Word document from the university Blackboard site, which included the name of the drug, the amino acid sequence of the drug presented in FASTA format, and a chart for recording predicted epitopes found by the IEDB analysis.

A detailed outline of the IEDB workflow is provided for instructors for weeks 1 and 2 (Appendix 1). In week 1, the students' first task was to examine the protein drug for potential epitopes that could be presented by the assigned MHC-II molecule. Students were instructed to copy the amino acid sequence of their chosen protein drug into the MHC-II Binding Predictions tool on the IEDB site (http://tools.iedb.org/mhcii/). To limit the parameters of the search to the MHC-II allele assigned to their laboratory section, students selected the appropriate allele from the "Select MHC allele" dropdown menu. The instructors informed the students to leave the other search parameters (eg, output of rankings) at the default settings to ensure that the group was viewing uniform datasets. Upon submitting the drug sequence into the search engine, IEDB generates a list of potential epitopes in the drug sequence that could be presented by the MHC-II allele based upon homology to known epitopes for that allele. The predicted epitopes are ranked by a percentile score where lower percentiles equate to good binders to the MHC-II molecule. The score of a given peptide is generated from averaging of the rankings from several prediction algorithms within IEDB. A percentile cutoff of 10% or lower was used in order to focus student analysis on the epitopes with the greatest likelihood of binding to the MHC-II molecule. The students recorded the amino acid sequence and location (eg, residue numbers of the polypeptide spanning the epitope) of the predicted epitopes into the chart in their worksheet. Overlapping epitopes that shared more than 50% of the same amino acids were excluded by choosing a representative epitope from the overlapping group that showed the lowest percentile score. At the conclusion of the laboratory, students recorded the number of predicted epitopes that were found from their search in the laboratory section's spreadsheet in class. The spreadsheet was saved by the instructors and used as the starting point for data collation in the next session.

In the second week of the laboratory, students compared the predicted epitopes identified from the protein drugs in week 1 with experimentally validated T cell epitopes in the IEDB. The goal of this exercise was to identify potential cross-reactive epitopes in the protein drugs that could be found in environmental stimuli or infectious agents. Despite the specificity of the adaptive immune response, cross-reactivity between different antigens is a common occurrence, and the role of cross-reactivity in ADA development is poorly understood. Cross-reactivity occurs when a T cell receptor (or antibody) binds to an antigen that is not the same one used to elicit the immune response. In the laboratory, the instructors explain an

Table 1. Learning Objectives for a Bioinformatics Practicum on Immune Recognition of Protein Drugs

I.	Explain basic immunological concepts that underlie the basis of recognition of foreign proteins
II.	Identify potential epitopes in protein drugs that could be recognized by specific MHC genotypes using online bioinformatics tools
III.	Hypothesize how immune recognition of a protein drug could lead to treatment failures

example of cross-reactivity, where a T cell epitope from Epstein Barr virus was shown to mimic an epitope from myelin basic protein.[10]

Using the predicted epitopes recorded during week 1, students entered each unique peptide into the "Linear Epitope" search on the IEDB home page. The lowest possible stringency was used for the matching algorithm ("BLAST – 70%" from the Linear Epitope dropdown menu, allowing for 70% sequence identity or similarity[11]) in order to capture the greatest number of matches in the database. For each match to the queried drug sequence, IEDB returns the amino acid sequence of the matching epitope, the antigen from which the epitope is derived, the species of origin of the antigen (eg, human, mouse, pathogen), and the references for the experimental binding data. Students recorded the data from each homologous epitope, grouping epitope matches by origin of the antigen, and repeated the search with any remaining predicted epitopes from their protein drug. If no matches were found in the database, students recorded in a separate chart that there were no homologous epitopes discovered.

After each pair of students had analyzed their data, they recorded any matching epitopes in a master spreadsheet that was projected on a screen in the classroom. Thus, each class produced a list of potential crossreactive epitopes for a given MHC-II allele from a variety of protein drugs. The instructors then led a discussion with the class to examine the types of antigens that were identified as potential sources of cross-reactivity, and how prior or coincident exposure to cross-reactive antigens could influence therapeutic outcomes with a given drug. In subsequent laboratory sections, students were shown the data from another laboratory group as a point of comparison, as each laboratory section analyzed the same list of drugs within the context of presentation by a different MHC-II allele.

After the data collation and group discussion in week 2, students completed a report based on their findings. The laboratory report included a series of questions that asked the students to apply their findings to a patient case involving treatment with the drug they had analyzed in laboratory, and to discuss the role of the potential role of immune response in treatment failure. The students had 1 week to complete the report, which was graded by the instructors. The practicum was deemed to be exempt by Duquesne University's Institutional Review Board.

EVALUATION AND ASSESSMENT

A pre-laboratory quiz, the practicum exercises, and a final laboratory report were graded for all students in the practicum (n=150). The quiz was comprised of brief short-answer questions that assessed whether students had completed the pre-laboratory video tutorial and gained an understanding of MHC function and cross-reactivity. Each laboratory section was given a different quiz to protect the integrity of assessment at the beginning of the week 1 session, and there were no significant differences between the laboratory sections ($p>0.05$ by one-way ANOVA). Regardless of the laboratory section, student performance on the prelaboratory quiz varied (Mean (SD) = 77.7± 24). We speculate that the differences in performance on the prelaboratory quiz may have been because it was a relatively low-stakes assessment (ie, worth less than 1% of the final grade for the spring ABLE sequence). Another possibility is that the immunological processes described and the application to antibody development against protein drugs may have been difficult to learn thoroughly from watching a short video. Despite the reasons for the wide range of scores on the prelaboratory quiz, 63 students (42% of the class) received a grade of >90% on the quiz, suggesting that many of the students had learned the underlying immunological concepts before entering the laboratory. While we could not track which students had watched the video in order to correlate that with quiz performance, future iterations of the laboratory will increase the weight of the prelaboratory quiz grade to ensure that students complete the prelaboratory preparation.

The mean grade on the final laboratory project was 87.1%±10.2% (n=150). Students performed strongly on sections of the project that required explanations of immunological processes, such as the role of MHC-II molecules in inducing an immune response against foreign proteins (related to Learning Objective I; Mean = 88.9%) and how proteins with regions of amino acid similarity could lead to the production of cross-reactive antibodies (related to Learning Objective III; Mean = 84.9%, Table 2). Students also were asked to consider a scenario where a patient was initially responsive to the therapy, but 4 months later had become unresponsive to the protein drug despite increased dosage (Table 2; Patient Case sections). Scores related to the patient case were lower in comparison to questions regarding immunological prinicples. For example, the final section of the project asked the students to suggest alternative explanations for why the example patient became unresponsive to therapy. On average, students received the lowest scores on this section (Mean = 76.1%; Table 2.)

In addition to direct measures of student learning, we also measured perceived learning gains as reported by the students. Pre- and post-laboratory survey instruments were distributed to the students 1 week prior to the first laboratory meeting and 1 week after the final project was submitted. Completion of the surveys was voluntary and anonymous. The response rate was 96% (144/150 students completed the survey) for the pre-laboratory survey and 89% (133/150 students completed the survey) for the

Table 2. Student Performance on the Final Project for a Bioinformatics Practicum on Immune Recognition of Protein Drugs

Laboratory Report Section	Average Scores (SD)	Average Percent	Related Learning Objective
Identification of potential drug epitopes and candidate crossreactive proteins by patient MHC allele using iedb.org	9.3/10 (1.5)	93.2	II
Descriptions of the relationship between antigens and epitopes in the induction of an immune response	3.6/4 (0.7)	89.7	I
Explanation of differences in MHC-I and -II in recognition of foreign proteins	3.6/4 (0.5)	88.9	I
Discussion of immune crossreactivity with clinically-relevant examples	3.4/4 (0.7)	84.9	III
Patient case: role of the immune system in treatment failure with a protein drug	3.2/4 (1.0)	81.1	III
Patient case: speculation on other causes of treatment failure with protein drugs	3.0/4 (0.9)	76.1	III

Average score out of the total points allotted for each section are shown. Sections are connected to the learning objectives described in Table 1

post-laboratory survey. Eight statements regarding the student's perception of their understanding of different topics raised in the practicum were assessed using a 5-point Likert scale (1=strongly disagree with the statement; 5=strongly agree with the statement). Table 3 shows the comparison of pre- and post-laboratory survey data. In the surveys, the term "biological disease modifying anti-rheumatic drug (DMARD)" was used in place of "protein drug" because the students had been introduced to biological DMARDs in a pain management course prior to the start of the laboratory. Students showed increased confidence in their understanding of each subject presented ($p<0.001$ by Wilcoxon signed-rank test for all questions). The greatest perceived learning gains were reported in understanding the roles of MHC molecules (+1.9) and differences between MHC-I and MHC-II (+2.3), and in comfort levels for describing to a patient how the immune response might interfere with protein drug therapy (+1.9). The highest score on both the pre- and post-laboratory surveys was for the students' understanding of cross-reactivity. The lowest perceived improvement (+1.5) on the surveys was on understanding how pharmacogenomics tools could be applied to predict differences in immune responses in patients.

DISCUSSION

The bioinformatics practicum was designed to address 2 perceived gaps in the pharmacy curriculum. First,

Table 3. Pharmacy Students' Perceived Learning Gains After Completing a Bioinformatics Practicum

Survey Questions	Pre-laboratory Score[a,b]	Post-laboratory Score	Difference
I understand the phenomenon of cross-reactivity.	2.7	4.5	+1.8
I can explain the role of MHC molecules in the immune response.	2.1	4.0	+1.9
I understand the different roles that MHC-I and MHC-II play in immunity.	1.7	4.0	+2.3
Imagine that there is a patient who needs counseling for one of the biological DMARDs. I would feel comfortable counseling the patient on the how their immune system might reject this drug.	1.8	3.7	+1.9
I understand how pharmacogenomics would apply to predicting treatment outcomes in patients on biological DMARDs.	1.9	3.6	+1.7
I understand how pharmacogenomics tools can be applied for studying differences in immune responses in the human population.	2.3	3.8	+1.5
I am comfortable explaining how anti-drug antibodies are formed.	2.3	3.9	+1.6
I can explain the role that anti-drug antibodies play in treatment failures with biological drugs.	2.3	4.0	+1.7

MHC = major histocompatibility complex; DMARD = disease modifying anti-rheumatic drug

[a]Responses to survey questions were based on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree

[b]Data are reported from 5 laboratory sections

we have found that pharmacy students often fail to retain their understanding of immunology from the first year. In Duquesne's PharmD curriculum, immunology is taught in 6 contact hours within a broader Human Physiology and Pathology course. Because the immunology material is embedded within a much larger course, the students may not recognize its importance or application to other subjects, and thus not prioritize retention of the information. By providing an inquiry-based laboratory that incorporates basic immunology, the bioinformatics practicum provides students with a mechanism to apply their understanding of immunology to a more clinically related question. Second, protein drugs are becoming more widely used, particularly for chronic disease states where prolonged administration is necessary. Thus, it is important for student pharmacists to consider how these drugs may interact with the immune system. Although protein drugs are engineered to minimize immunogenicity, the diversity of MHC molecules and the potential for cross-reactivity present challenges for predicting whether a drug will be recognized by an individual's immune system.[11,12] Thus, the practicum was designed to highlight these challenges so that students could appreciate the complexity of using a foreign protein as a therapeutic.

Students's final report scores suggest that they completed the data acquisition and analysis from IEDB with relative ease (93.2%±15%, Table 2), and were able to address corresponding questions about the protein structure (eg, humanized monoclonal, chimeric proteins) and the mechanism of action of their protein drug of choice. The scores also suggest that many of the students improved their understanding of immunological processes (eg. MHC presentation, cross-reactive epitopes) after completing the pre-laboratory quiz and laboratory exercises. However, students struggled more with applying their results to a hypothetical patient case involving the protein drug assayed in the laboratory. Most students were able to offer a reasonable explanation of how the immune response could have come to recognize the protein drug based upon the make-up of their MHC-II haplotype, but we unable to explain whether the timeframe of the treatment could reasonably be expected to involve immune rejection. In other words, some students were confused as to whether 4 months was a reasonable period of time for an antibody response to be mounted against a foreign protein. While we did not delve specifically into the timing of an antibody response, we had expected that the students would be able to extrapolate a sense of the timing from their infectious disease course, where the development of immune responses is often discussed in relation to the start of an infection and to vaccination. Students struggled most when asked to hypothesize about possible explanations for why a patient would become unresponsive to therapy over time. We suspect that students were weaker in this section because it was the most speculative part of the assignment, requiring the students to consider not only the immune processes that were the focus of the laboratory but also specifics of the disease state for which their drug of choice would be prescribed. As an example, treatment failure with cetuximab, a monoclonal antibody used in colorectal and head and neck cancers, or with daclizumab, a monoclonal antibody used to prevent rejection of organ transplants, may involve factors that are specific to the disease state as opposed to an issue of immunogenicity of the drug. In future iterations of the course, there may be a need to orient students to the broader context of therapeutics, with development of anti-drug antibodies as just one possible cause of secondary treatment failures. Regardless, most students (92.7%) received some partial credit on this section because their responses suggested that they were able to begin reasoning through some of the additional factors that can contribute to lack of efficacy with protein drugs.

In the pre- and post-laboratory surveys, students expressed the greatest confidence in their understanding of cross-reactivity (Table 3). Based on the quality of the answers in the prelaboratory quiz, we did not anticipate a relatively high degree of confidence on the subject in the prelaboratory survey. However, cross-reactivity is discussed in other courses, such as Infectious Disease I and II, which also take place in the second year. Thus, students would have had repeated exposure to the topic of cross-reactivity in the same semester, which may explain their comfort with the topic. The lowest response in the post-laboratory survey addressed how pharmacogenomics applied to prediction of treatment outcomes. We speculate that students were less comfortable with this statement as we had explained the laboratory in terms of a bioinformatics exercise, and had not emphasized potential applications to pharmacogenomics in the laboratory sessions. Nevertheless, students reported a positive trend in their understanding of the subjects presented in the survey.

Students were more successful at explaining multi-step immune processes than at explaining the possible role of cross-reactivity in the development of anti-drug antibodies. We suspect that the application to cross-reactivity was more challenging because it involves multiple immune recognition steps within an undefined timeframe. Moreover, it is not known how much of a role cross-reactivity plays in treatment failure with protein drugs. The outstanding questions regarding cross-reactivity were emphasized by the instructors in order to encourage the students to view the exercise as an experiment. Therefore, the students' uncertainty may be appropriate because it is a very experimental question rather than

a clearly documented process that can be cited and memorized. Nevertheless, students reported confidence in their understanding of cross-reactivity itself, even if they were less certain about how the process might unfold with protein drug therapy.

The T Cell Epitope Prediction Tool in IEDB functions as a Web-based server in ranking potential MHC ligands (La Jolla Institute for Allergy and Immunology, La Jolla, CA). It operates by calculating the median percentile ranks of peptides generated from a collection of prominent MHC predictive methods. This mode is referred to as the "Consensus" method. However, the user has the option of choosing a specific algorithm for epitope prediction.[13] These include NETMHCIIpan, NN-align, SMM-align, Combinatorial library, and Sturniolo. The most efficient option would be to let the IEDB server select the most appropriate method; it could either select Consensus or apply one of the algorithms, depending on the MHC allele being queried. In the laboratory exercise, we used the default option of allowing the server to select the best method. Added to this flexibility is that the MHC prediction function is interfaced with experimental T cell assay data mined from the literature. We propose that having a working knowledge of IEDB would equip students with a broad view of adaptive immunology, a learning outcome unlikely to be achieved by designing the exercise around a specific MHC prediction algorithm.

In addition to the MHC binding prediction and T-cell epitope search, students may benefit from engaging other IEDB features. Skill-based activities designed around the suite of B-cell epitope prediction tools would reinforce a 3-dimensional perspective of antigen and antibody interactions. Using known examples, the tools can be used to illustrate the difference between linear and discontinuous epitopes. Hands-on bioinformatics exercises using different protein sequences would provide opportunities to further explain antigenic drift and point mutations in pathogenic proteins. An example of assessment would be asking students to explain the challenges in developing effective HIV vaccines. NetChop is another feature that can be integrated into laboratory courses; this tool is used to predict proteasomal cleavage of MHC class I ligands. This level of granularity is often beyond the scope of immunology courses taught in pharmacy programs. However, understanding the mechanisms involved in processing of MHC-I ligands is crucial in evaluating the pharmacological rationale and clinical performance of cancer vaccines and immune checkpoint inhibitors. Population Coverage is another tool that can used to build assignments; analysis of individuals' responses to a given set of epitopes can be used to highlight the rationale for personalized medicine. Thus, IEDB is an underutilized educational platform on which learning could be enhanced throughout the pharmacy curriculum.

The epitope data in IEDB is curated from the published literature on infectious disease, allergy, transplant rejection, and autoimmunity.[5] While the database is the most comprehensive of its kind, there are limitations to the data that should be considered when using IEDB for an inquiry-based activity. While searching for potentially cross-reactive epitopes, some queries did not receive any hits in the database. Students took an absence of matches to mean that the epitope found in their drug would not cause an immune response because there was nothing similar to it in the database. However, the epitope search can only compare the input sequence to epitopes with previously documented experimental support. Thus, it is possible that the input sequence has homology to epitopes that are not identified yet or that it truly does not share homology with another sequence. While this distinction can be confusing for the students initially, we found it helpful to address it while we had the class discussion to review the data.

Of the articles that cite IEDB as a resource,[5,14] we did not find any that used IEDB in an educational setting. A recent conference presentation reported using IEDB in an undergraduate microbiology course to search for epitopes within methicillin-resistant *Staphylococcus aureus*, which resulted in improved learning outcomes.[15] Despite the limited examples of IEDB as an educational tool, we present findings that IEDB can be readily integrated into a classroom or laboratory setting. IEDB is freely available and only requires a computer and Internet connection to use; thus, other schools of pharmacy could easily access these search tools and data for use in their laboratory teaching. In addition to opportunities for PharmD students, the bioinformatics practicum could be adapted for use as an interdisciplinary project with students from other health care professions. Pharmacy students could be paired with nursing or medical students to investigate whether a series of related protein drugs would be predicted to contain T and B cell epitopes based on a common patient haplotype. Using this practicum as a starting point, the students could then propose a hypothetical clinical trial to test their predictions for immune recognition of the protein drugs as a final assessment. The practicum could also be altered for prepharmacy students or undergraduate students in microbiology or genetics courses, where the emphasis could be shifted from patient outcomes to immune processing.

SUMMARY

We report on a bioinformatics exercise that asks student pharmacists to consider the therapeutic implications of protein drugs and common environmental exposure to antigens. The mechanics of the exercise provided

opportunities for students, working in pairs, to recall fundamental principles of pharmaceutical biotechnology and immunology. Pharmacogenomics is the synthesis of the concepts. The expanding use of protein drugs highlights the need for this and similar bioinformatics problem-solving modules integrating concepts in the pharmacy curriculum.

REFERENCES

1. US Food and Drug Administration. 2015 biological approvals. http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/BiologicalApprovalsbyYear/ucm434959.htm. Accessed January 18, 2016.
2. Medscape. 100 best-selling, most prescribed branded drugs through March. http://www.medscape.com/viewarticle/844317#vp_2. Accessed January 18, 2016.
3. Wolbink GJ, Vis M, Lems W, et al. Development of antiinfliximab antibodies and relationship to clinical response in patients with rheumatoid arthritis. *Arthritis Rheum.* 2006;54(3):711-715.
4. Kui R, Gal B, Gaal M, Kiss M, Kemeny L, Gyulai R. Presence of antidrug antibodies correlates inversely with the plasma tumor necrosis factor (TNF)-alpha level and the efficacy of TNF-inhibitor therapy in psoriasis. *J Dermatol.* 2016;43(9):1018-1023.
5. Vita R, Overton JA, Greenbaum JA, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* 2015;43(Database issue): D405-412.
6. Vita R, Peters B, Sette A. The curation guidelines of the immune epitope database and analysis resource. *Cytometry A.* 2008;73 (11):1066-1070.
7. American Association for the Advancement of Science. Vision and change in undergraduate biology education: a call to action. http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf. Accessed August 12, 2016.
8. Harding FA, Stickler MM, Razo J, DuBridge RB. The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. *MAbs.* 2010;2(3):256-265.
9. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet.* 2012;44(3):291-296.
10. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell.* 1995;80(5):695-705.
11. Andrick BJ, Schwab AI, Cauley B, O'Donnell LA, Meng WS. Predicting hemagglutinin MHC-II Ligand analogues in anti-TNFalpha biologics: implications for immunogenicity of pharmaceutical proteins. *PLoS One.* 2015;10(8):e0135451.
12. Paul S, Kolla RV, Sidney J, et al. Evaluating the immunogenicity of protein drugs by applying in vitro MHC binding data and the immune epitope database and analysis resource. *Clin Dev Immunol.* 2013;2013:467852.
13. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform.* 2015;53:405-414.
14. Vita R, Zarebski L, Greenbaum JA, et al. The immune epitope database 2.0. *Nucleic Acids Res.* 2010;38(Database issue):D854-D862.
15. Makhluf H. Mining the immune epitope database (IEDB): a novel approach to learn about infectious diseases. 19th Annual American Society for Microbiology Conference for Undergraduate Educators 2012; San Mateo, Calif.

Appendix 1. Tutorial for Instructors

Supplemental material for "*A bioinformatics practicum to develop student understanding of immunological rejection of protein drugs*"

These instructions describe how to utilize the Immune Epitope Database as part of a laboratory exercise for analysis of MHC-II ligands in protein drugs. Video tutorials on the background and other aspects of the IEDB interface are available here:

http://help.iedb.org/entries/140865-How-To-Videos

For the Bioinformatics Practicum, students selected a protein drug from a list of sequence provided by the instructors. Each lab section was also assigned a specific MHC-II allele (*e.g.* HLA-DRB1*0401). Students then performed two tasks within IEDB using the protein sequence of the drug:

1. Identify potential MHC-II ligands for their assigned allele in the drug sequence (Week 1).
2. Determine whether the MHC-II ligands from Week 1 showed cross-reactivity with known T cell epitopes (Week 2).

Below, you will find instructions on how to navigate the IEDB interface for each week's assignment.

INSTRUCTIONS (Week 1): ***Identifying potential MHC II epitopes in the protein drug***

1.) The students choose a protein drug from a list provided by the instructors. After the students have selected a drug, they are instructed to download a corresponding word document from Blackboard that includes the amino acid sequence of the drug. Students are then directed to the website for IEDB.

2.) Enter the IEDB website (http://www.iedb.org)

i. On the IEDB homepage, go to the "Epitope Analysis Resource" box. Within the box, click on the "MHC II binding" link.

1. You will be taken the "MHC-II Binding Predictions" page. Paste the amino acid sequence of the protein drug into the top window.
    a. Enter the protein sequence in FASTA format (*e.g.* add a ">" followed by the name of the name of the drug, hit "Enter," and then paste the protein sequence on the next line).
    b. For the purpose of this exercise, we provided the students with the sequence of the heavy chain only (when the protein drug was an antibody).
2. Below the space to enter the protein sequence, there will be multiple dropdown menus that will define the search parameters.
    a. Under the dropdown menu labeled as "Allele," select an appropriate MHC II allele
        i. We assigned each lab section a unique allele, so that each section was examining multiple drugs with one allele.
    b. We used the preset selections for other choices on this page, so the other dropdown menus were not changed.
    c. Once the appropriate MHC II allele is selected, click on the "Submit" button on the bottom, right corner.
    d. The program will run each contiguous 15 amino acid fragment from the drug. Each fragment will be scored by how likely it is to bind to the chosen MHC-II allele.
3. You will be taken to a page called "MHC-II Binding Prediction Results."
    a. We use the "percentile rank" (last column) to sort the data. A low percentile rank is indicative of protein sequence that has a high likelihood of binding to the MHC-II molecule. A high percentile rank means that the sequence is less likely to bind.
    b. Results are listed in order by ascending percentile rank. Thus, the peptides at the top of the list (and with a low percentile rank) are more likely to be good binders to the MHC II allele of choice.
    c. As a cut-off, we only examine sequences with a percentile rank that is ***less than 5%.***

3.) Sort the data in the "MHC-II Binding Prediction Results" page to eliminate overlapping peptides.
- a. Each search typically yields over 100 hits. By eliminating hits with a percentile rank >5%, the number of peptides will be reduced. However, even after eliminating peptides >5%, there often were too many peptides left to carry forward into the next analysis.
    - i. In order to focus the analysis, overlapping peptide fragments will be discarded. This step is not required, but we found that it streamlined the lab session.
- b. Start at the top of the list (with the peptide with the "best," or lowest, percentile rank.)
- c. For the 1st hit, look at the "start" number for the peptide.
    - i. For the sake of an example, let's say that the first hit starts at amino acid 75.
- d. Add/subtract 10 from the start number to give a range of start numbers.
    - i. In our example, the overlapping range for the first peptide is 65-85.
- e. Scan the list for peptides that start between 65-85.
- f. Eliminate fragments that start in the overlapping range from the next analysis.
- g. After eliminating the overlapping peptides for the first hit, move to the next peptide on the list and repeat the process.
    - i. Repeat until all of the peptides <5% have been covered.

4.) Students record the non-overlapping peptides (amino acid sequences, start/end numbers, and percentile rank score) in a Word document chart. The students are instructed to bring the chart to the next lab session. A worksheet with questions regarding the structure of the protein drug (*e.g.* chimeric antibody vs. humanized antibody), the molecular target of the drug (*e.g.* TNFα), disease states for which the drug is prescribed, and the effect that the drug would have on pathogenesis is assigned. Students are instructed to hand in their chart and worksheet at the next lab session.

INSTRUCTIONS (Week 2): ***Prediction of cross-reactivity***

1.) In Week 2, students compare the drug peptides that they recorded in Week 1 (*e.g.* those peptides that are predicted to bind to the assigned MHC II allele) to known T cell epitopes (*e.g.* peptides that have been shown in the literature to be recognized by T cells). The goal is to consider whether the drug peptides that are potentially presented by the MHC II molecule possess similarity to known T cell epitopes. If so, then there is potential for cross-reactivity between the drug and the known epitope.

2.) Students are directed to the IEDB homepage.
- a. In the "Epitope Search" window, select "Linear Epitope."
- b. Paste one peptide sequence from the Week 1 analysis into the box below "Linear Epitope."
    - i. In the dropdown menu below "Linear Epitope," select "BLAST – 70%."
        1. The BLAST percentage determines how similar the amino acid sequences must be in order to be considered a match.

2. "BLAST-70%" means that the peptide from the drug will be considered a match to any epitope in the database that has 70% amino acid identity.
3. We chose "BLAST-70%" because cross-reactive epitopes have been identified with 70% similarity.

ii. In the "Assay" window, select "Positive Assays Only" and "T cell assays."
1. "Positive Assays Only" is selected so that negative data is omitted from the comparison. In essence, matches will be considered if a peptide was shown to elicit a T cell response in a published experiment. If a peptide did not elicit a T cell response, it is ignored in this analysis.
2. For a protein drug to elicit an immune response, T helper cells would ultimately have to bind to a fragment of the drug when it is presented by an MHC II molecule. Here, we focused on "T cell assays" to ask whether any other peptides with sequence similarity to the protein drug had been shown to elicit a T cell response. If so, then the peptide fragment from the drug has the potential to be crossreactive to the peptide found in the database.

iii. Hit the "Search" button at the bottom of the frame.

3.) Students record the results of the database search.
a. If a matching epitope is found in the database, the students record the epitope (*e.g.* the sequence of the matching peptide), the antigen (the protein from which the T cell epitope was derived), and the organism from which the antigen is found (typically human, mouse, or a pathogen).
i. It is possible that there will be no matches in the database. This means that there is no experimental data currently published that is similar to the input sequence.
b. Students categorize the match based upon the host: bacterial, viral, autoimmune (human or mouse proteins depending upon the model that was tested in citation), or allergen (such as grass or dust mites).
c. Students repeat the search with any other drug peptides that were recorded during Week 1.
d. Students complete a worksheet regarding how matches could result in a cross-reactive immune response and influence treatment outcome (see text for details.)

4.) For further information on the epitopes in the database, one can select the "Assays" tab or the "References" tab on the results page. The "Assays" tab lists the experimental data that was used to demonstrate that the entry was a T cell epitope (*e.g.* 3H-thymidine proliferation assay). The "References" tab provides a link to the paper that included the epitope data.
a. We did not ask students to look into these tabs. However, the tabs show the availability of the experimental evidence that was used to log the epitope into the database.