Introductory Biology re-imagined: exploration-driven learning

Introductory cell/molecular biology textbooks and course design largely reflect inherited traditions established at least half a century ago. Neither have adapted to the explosive accumulation of information, modern thinking about learning, or the accessibility of vast computer power in the hands of every student. This leaves an emphasis on gorging on terminology and factual information even as the value of such diminishes due to its availability in literally everyone's pocket. An unfortunate side effect is that students are exposed to biological science as a massive body of facts and vocabulary rather than a way of investigating and understanding the wonders of life. The approach proposed here is intended to identify and deliver understandings of key concepts in biology through interactive *challenges* and *investigations* on the part of students, guided by interactive software that depicts the molecules of life as they are thought about by real biologists--3D entities with easily assessed surface properties that determine the behavior and interactions of the molecules. The overall vision is of coherent electronic modules combining informative text, interactive challenges focused on tasks and concepts rather than words, and meaningful 3D visualizations of molecules.

There are a number of underlying premises to this approach: 1) introductory biology students are capable of understanding rather than just memorizing terms and labels--and would be more engaged and learn more if they were; 2) better learning will take place if major ideas are presented in an interesting, relevant context; 3) students must be offered opportunities to engage with the ideas—by manipulating molecular representations, by solving problems, by presenting their reasoning alongside 'correct answers'; 4) a building sequence where underlying facts are presented when relevant, rather than a "chemistry/all macromolecules/all basic processes/all advanced studies" will result in better understanding and retention; 5) gathering basic ideas into recurring conceptual/exemplar frameworks will make learning easier, more coherent, and more persistent (i.e. using hemoglobin to teach amino acids, protein folding, structure/function, hydrophobicity, pH, mutation/adaptation/evolution; using an ATPase to teach enzymes, reactions, ATP, phosphotransferase reactions).

This approach is novel, and must confront deeply entrenched classical lecture and textbook approaches. For these reasons, the units might be developed and offered as independent modules, though with the long-term goal of creating a coherent, integrated unit that would be appealing as a framework for a full semester. Many of the modules proposed exist as functioning units, generally with automatic grading features and database storage of student work, and have been delivered at a large public university for 5-10 years. Currently the course requires following textbooks out of order, and is assembled from diverse web and published resources (though Alberts et al.'s "Molecular Biology of the Cell" provides both video and textual supplements in some key areas—I still believe this to be one of the greatest collegiate biology texts available, though it is targeted differently than my course).

Topic order and instructional examples

In order to combat the twin student confusions of “Why are we learning this?” and “Where in my schema does this information *fit*?” I deliver my course in a somewhat unconventional order. I do adhere to the usual introduction of core chemical understandings (atoms, outer shells, hydrophobicity, bonds and interactions). While my course is a combined lab/lecture, many real-world examples could be extracted as videos or text (staples ‘floating’ on water due to surface tension; miscibility of water with ethanol, propanol, butanol; bending of a stream of water by a static-y balloon). These are intended to provide an experimental, observation basis for thinking about molecules and their properties *concretely*.

From here, we proceed directly to protein structure and function (examples below). The reasoning is that students are more familiar and interested in how their bodies work, and the abstract concept of ‘information’ is challenging in the absolute, doubly so without examples of what the information codes *for* or a beginning grasp of how molecular machines work. Proceeding directly to proteins also allows us to implement their learning about the properties of differently bonded atoms in ‘building’ the amino acid toolkit. A lab activity that could also be rendered as a video ‘colors’ one side of an index card with graphite, hole punches out disks, and demonstrates their ‘folding’ graphite-side-up in a mineral oil-water mixture. Detailed case-studies of hemoglobin and an ATPase follow, where students can solidify understanding by observing how actual amino acids in an actual protein bring about actual function or disease.

Having established *what* is made, the nature of the instructions (bases with unique patterns of binary (partial + or -) extensions is introduced. Discussion of “A goes with T” is minimized in favor of a constant return to molecular surfaces and the accompanying inevitability of matching of ‘correct’ pairs. The concept of DNA Polymerase as a machine that *recognizes* pairs, rather than creating them, provides a more evolutionarily and mechanistically correct view, and one consistent with the structural foundation we have provided.

Controlling the genetic library follows. I teach the *lac* operon as a simple, elegant example (though I am still adjusting to the ‘recent’ suspicion that control may derive almost exclusively through repressor and lactose availability). I then follow with a brief sojourn into eukaryotic development by looking at the *even-skipped* (*eve)* stripe 2 genetic module and its outcomes.

From this point, it seems that there is great flexibility to teach well through a variety of different paths. Energetics (photosynthesis and respiration) is rather modular, though again the emphasis we deliver is on a “follow-the-electrons” view of the citric acid cycle and electron transport chain (it is striking how few textbooks *explicitly* include representations of electrons in the citric acid cycle—troubling given that their extraction is the entire point). This is coupled to ideas of conversion of different forms of potential energy, from the apposition of like charges (established in the discussion of ATPase as example enzyme; see below) to “energy of orbital occupancy” found in sugars and C-H bonds (a video of flammable CH4, and other -CH containing compounds vs. non-flammable H2O and CO2 establishes the ‘energies within) to the potential energy of concentrated protons.

Genetics modules are generally reasonable in several existing textbooks; the one notable addition here is the software generated tasks (see Appendix). In the Mendelian exploration, students are given simulated organisms from which they must deduce dominance and genotype. In the Pedigrees exploration, students are challenged to examine a single pedigree in light of three different inheritance models (autosomal dominant and recessive; X-linked recessive) and conclude about the viability of the model. A notable feature is that each assignment of genotype must be accompanied by a *rationale;* if this is anything other than the necessary-and-sufficient information, the computer attempts to guide the student, but the assignment is not accepted.

My own course then follows a family of timers: EF-Tu in translation, tubulin in chromosome-hunting, and GPCR’s in signaling. The goal here is to show how a common machine element is adapted to different environments and to challenge students to understand how the cell solves problems

To address the cell and cellular organization, signal sequences as ‘cellular zip codes’ and the ‘decoration’ of vesicles with appropriate ‘locks’ matching organelle ‘keys’ (SNAPS and SNARES) completes the molecular-based tour of the cell

Example: relevance through protein function and genetic disease

The classical progression in introductory molecular biology is decidedly strange when viewed in light of current thinking about creating understanding and learning. All the properties of the chemical world are assembled en masse and then placed on a shelf to be taken down at arbitrary future times. pH is a classical example, often finding its first practical application hundreds of pages after being introduced. The building blocks (amino acids, nucleotides) parade by, as does the general concept of polymers, but the roles these perform, and how they play them, are given only label-style status. Perhaps most strangely, most textbooks describe the parts of a protein, describe how one is assembled *but never truly discuss how one functions or provide a deep understanding of its purpose.* Given the breadth of well-understood examples, the argument that 'students at this level are not ready to actually understand things' seems both insulting and misguided.

In an effort to remedy these shortcomings, we have adopted an approach featuring hemoglobin as a concrete example. Hemoglobin has a compact structure featuring a number of alpha helices (allowing presentation of secondary structure in context). Oxygen binding causes notable conformational change, and is facilitated by a number of sidechain interactions as well as pH regulated events (Bohr effect) and small-molecule mediated effects (BPG). Further, hemoglobin study can serve as a springboard for learning about evolution through gene duplication and mutation (fetal hemoglobin and decreased BPG binding), as well as mutation, phenotype, and reinforcement of the importance of hydrophobicity (sickle cell anemia). Linkages to evolution and selection can also be made if the example of sickle cell anemia + malaria is introduced. An example page is here (in class, these are accompanied by automated question/answer software):
https://thinkbio.guru/3D_Directory/!HemoTable.html
[if Java presents difficulties, use the 'HTML5' setting in the 'Display technology' menu at bottom left]

In our curriculum, we then examine hemoglobinopathies, making a point that these are actual mutations with real sufferers. Students are then challenged to generate good hypotheses about the underlying cause of disfunction. Happily, many hemoglobinopathies are quite straightforward (for example, there are mutations in the histidines that contact the heme iron)
https://thinkbio.guru/3D_Directory/Flotsam_and_Jetsam/Hemo_myo/OpathyBPG.html

These experiences are extended and 'capstoned' by projects in which students are assigned a genetic disease and tasked with learning about its symptoms, distribution, and analyzing both an actual disease-causing mutation and a hypothetical one chosen on the basis of illuminating some specific feature of protein structure (salt bridge, hydrophobic interaction between monomers, etc.). These are currently being re-done, but an example from Gaucher's disease can be found here:
https://thinkbio.guru/3D_Directory/Flotsam_and_Jetsam/GenDiseases/Gaucher3D.html

A first enzyme: ATP and ATPases

Rather than taking the classical approach of discussing enzymes first (and sometimes *only*) in abstract form concerning activation barriers, we present a concrete example of an ATPase. This offers a number of potential advantages, including looking at ATP's 'energy storage' features in a concrete way, examining in readily understood detail why ATP is stable

despite being an energetic storehouse, and looking at simple ways that a protein can provide a structural framework that encourages ATP breakdown (selection and orientation of a water molecule, 'rewarding' the planar orientation of the phosphate's oxygens during the transition state).
http://www.mrc-mbu.cam.ac.uk/node/449
(I have an annotated version for classroom use)

Further, core events from this example are re-visited during modules on protein regulation via phosphorylation and the 'timers' module that looks at GTPases involved in ribosomal accuracy (EF-Tu), microtubule growth-collapse (tubulin), and signal decay (G-protein coupled receptors)

Advances in our understanding of myriad diseases and drug therapies offer many avenues for developing engaging, interactive, *mechanism-based* activities.

Example: *how* and *why* is DNA the 'information molecule of the cell'

Current treatment of DNA is exemplifies the shortcomings sketched earlier. Students emerge from classrooms only with the 'understanding' that something called 'A' has a relationship with something called 'T' and that this creates the 'code of life' bearing 'information'. While structures appear in textbook margins, key points about their interactions are rarely made in a coherent way. Classical representations fail to capture *properties* and *interactions* that matter; the student is never engaged as an investigator. A shift in focus allows presentation of DNA as a relatively simple chemical whose component structures *innately* bring the property of "copy-ability". Application of this property manifests as copies of specific instructions sent to the cytoplasm for implementation (mRNA) and complete copies generated for cell reproduction. Extension of this 'matching/pairing' toolkit leads to a second set of tools with very different properties--the amino acids, diverse in size, shape, and surfaces compared to the precise, stereotypical 'handshaking' bases. *Now* a review-level examination is appropriate: *how* do the properties of nucleotides vs. amino acids reflect exquisite fitting to their roles: one to allow low-error matching at the chemical level; the other to generate auto-assembling machines of diverse shapes and properties.

An example of how existing 'reinforcement' exercises fail to engage concepts can be found at the often-excellent 'Learn Genetics' site
http://learn.genetics.utah.edu/content/begin/dna/transcribe/
Here, the user is 'challenged' to match the letter A with T and then to use a 'look-up dictionary' to process the resulting string of nucleotide into an alternate string of amino acid letters. Neither of these reflects any aspect of the wonder or *mechanism* of base pairing, translation.

A second deep flaw in current presentations is a failure to consider what aspects of molecular biology represent *concepts* vs. needlessly detailed collections of information. Again looking to DNA for examples, I think the key concept could be termed "replicatability"—the accurate and precise one-to-one (and only one!) matching of base structures is the underlying feature of DNA replication, transcription into RNA, and tRNA selection during translation. A student would be hard pressed to identify this in modern textbooks, which emphasize exhaustive lists of the players, accompanied by complex diagrams and two or three sentences on each participant. Students memorize lists of names corresponding to functions they are not equipped to understand (and whose mechanism is often not even offered to them).

As an alternative, I propose offering students a participatory, detailed tour of how the bases interact (and indeed, fail to interact when Ade meets Cyt and Gua meets Thy). This focuses on the core idea Crick sought to encapsulate in his presentation of the central dogma—

that ‘information’ resides in structural matching of bases, while function resides in machines assembled on the basis of these instructions: proteins. The last key is the flow of this information from storage to implementation via translation.

In order to focus student understanding on base matching, I developed a software activity where students are presented with actual 3D structures of the bases (‘**BasePairer**’; see Appendix). While they begin with the classical representations (carbon atoms are gray, oxygens red, etc.) students are first challenged to identify the (+) and (-) ‘feels’ of the relevant atoms. This is a critical step, as these assignments ‘flesh out’ the bases and generate their most critical information-bearing feature: the ‘magnetic’ complementarity of the basepairing surfaces (hydrogen bond donors and acceptors). Once the bases are appropriately ‘labeled’ students are tasked with rotating/flipping the bases to discover/confirm pairings--as well as to check and see how ‘non-pairs’ (A-C, G-T) fit (Appendix I). This places the student in the role of *independent investigator* of claims and reinforces the idea that these are *real* entities whose details have readily observed consequences. Finally, students investigate how simple chemical changes (tautomerization, deamination) change the meanings of the bases by changing their partnering characteristic.

Other examples in use or under development

—CRISPR: how the specificity of base-matching is enabling precision repair of genes

—Molecular medicine: how size and shape of molecules alters bodily function. The ‘Philadelphia chromosome’ and Gleevec feature, but there are now myriad examples.

—Opsin: how mutations allowed the development of new functionality (human green opsin as a duplicated, mutated red opsin).

—Skin color: an extended investigation of human skin color, vitamin D synthesis, Inuit peoples (dark-skinned peoples living with limited Sun exposure), fish oil (source of vitamin D), and the Atlantic conveyor (alters climate allowing grain [low in vitamin D] production far North

—pH, the flu, and you: how the flu virus takes advantage of pH and cellular trafficking to Trojan Horse its way into the body

—Molecular routes to white: lizards at White Sands: a paper chosen for *relative* readability at an introductory level that includes evolutionary concepts, the Mc1r gene (introduced in the skin color module above), protein sorting (**PNAS** 107(5):2113)

Appendix: The software library

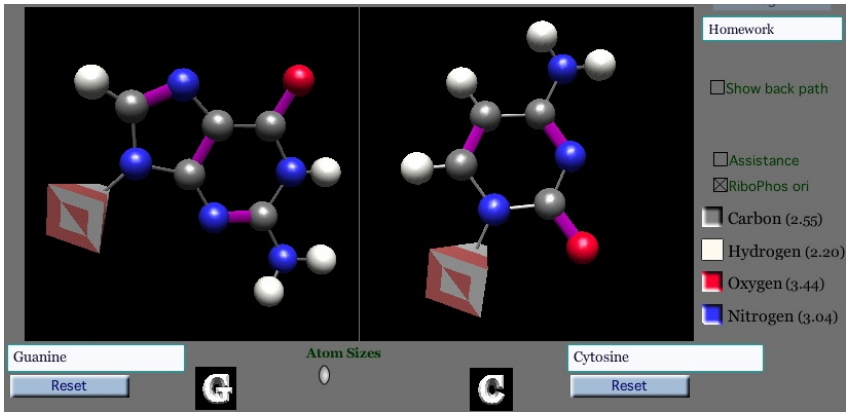
Overview: The suite of software I developed can be found at (use the 'Control Center' from 'Master launcher' linkh; download the one for your OS; no user account is required to investigate in the 'non-Assessor software' tab*):
https://thinkbio.guru/Modules/All_Modules_Table.php
I am the sole creator of this software and have rights to do with it as I would.
*Note that in summers, much of this software is updated and re-vamped; contact
bruce.patterson@princeton.edu
with any questions.

Information Flow

In order to provide students with a 'roadmap' of the processes behind the creation of cellular machines, we compare the processes of the Central Dogma to an heirloom family recipe book. When a family member leaves home, the book in its entirety will be copied and bestowed on the family member (replication, daughter cell). When going to a picnic, a single, disposable copy of a recipe (mRNA) is all that is needed. There is an accompanying interactive exercise where students participate in a mechanical copying and translation using pieces that embody the structural matching elements of bases. Finally, nucleotide instructions are converted to arrows with simple 'folding' rules (continue straight, rotate clockwise or counterclockwise, 45 or 90 degrees, and lengths of 3 different units). When 'translated', simple shapes (house, llama, bicycle) are created, exhibiting the diversity that can be created from a simple, linear code.

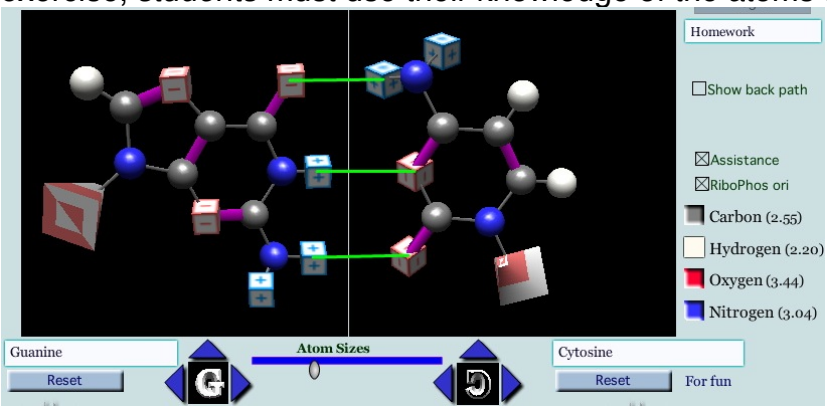
QuantumMine and PatternMaster: These are tools that challenge students to employ and recognize the Scientific Method. It's not clear these would find wide acceptance in an Intro Bio course; convincing students that thinking and problem solving are part of biology has proven to be a perennial challenge. PatternMaster is notable in that, unlike virtually all other challenges, students seek to identify underlying *rules* of a system, rather than working in a well-understood system to enumerate *facts*.

Basepairer: Exploring *how* and *why* the components of the Genetic Code 'know' one another. The introduction to the bases consists of two components. First, students are presented with the classical structural representations:



It is useful t learn this view, but it hides critical challenges for students. Chemically, the nitrogens (blue) and oxygens (red) are virtually identical in behavior in bases, but receive symbols that suggest they are very different. Even worse, a hydrogen (white) attached to a carbon (grey) is neutral in charge, whereas one attached to nitrogen or oxygen is positive. So 'pairing' these

molecules is not at all visually intuitive. To solve this, I have created a different display. In the exercise, *students* must use their knowledge of the atoms to re-label the bases as shown:



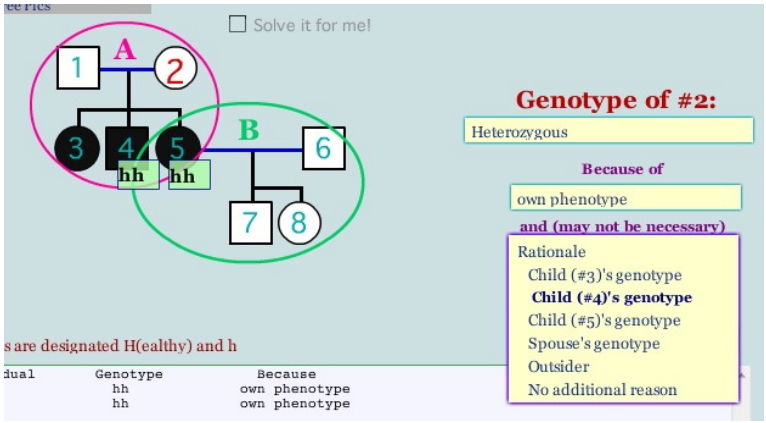
Here, the left base has been left in its original orientation, but informative (+) and (-) cubes have been substituted in. They are colored per the *other* coloring convention that is commonly accepted--blue for positive charge, red for negative. Visual inspection is enough to see that after the manipulation of the base on the right, the two 'match up'--at each position, a blue (+) faces a red (-).

StructViewer: A 3D representational tool that I am adapting to include the charge coloring conventions used above along with a yellow that indicates neutral/hydrophobic. Such an approach is vital because it is not our mythical *representation* (atom coloring that is conventionally used) but how each molecule 'feels' that illuminates its properties. Once a 3D engine is modified to allow this form of presentation, any molecule can trivially be added to the library.

Gameter: Punnett squares are a *means*, not an *end.* The Gameter program allows students to 'construct' arrangements of up to four different genes and position them on two different chromosomes. It generates an animation of recombination and 'explains' how different gamete genotypes arise from the initial disposition of alleles.

Mendelstar/MendelPede: This program is not truly unique at the level used in introductory biology, but it's usage is too rare. Instead of having students solve endless word problems or Punnett squares, it presents organisms of readily identifiable genotype but unknown phenotype and requires them to deduce through experimentation the underlying genotype (including the dominance relationship). x_plorer ('Cross explorer') is a support tool that helps students walk through the abstractions generally used to solve problems like those in Mendelstar. It should be viewed as beta software.

Pediducer: The unique aspect of this tool is the focus on the RATIONALE for each genotype declaration. Here, the user is shown having assigned genotypes to individuals 4 and 5 and is proposing the (correct) deduction that individual #2 is a heterozygote. The critical component here is the 'Because of' area at bottom right. The student is required to identify up to three *necessary* and *sufficient* elements to the deduction. Here, the correct two are in the process of being identified--having a dominant phenotype oneself but parenting a child of recessive phenotype (= homozygous recessive genotype).



Another key aspect of the exercise is that the user is required to examine each pedigree independently in terms of three different inheritance models. Here, the user investigates the 'Autosomal recessive' model. Each model must be deemed 'plausible' (based on having successfully assigned complete information to every member of the pedigree) or 'ruled out' (based on *identifying* an 'impossible family' where the parents could not have generated the child). This aspect of simultaneously evaluating multiple hypotheses in often neglected in introductory biology, where there is a tendency to say "Here is how they proved the correct hypothesis"

VocabuWary:

Small stuff:
aaDancer is a simple look at animated amino acids and DNA bases, emphasizing the flexibility of the one and the rigidity of presentation of the other

HangTwenty is a game based on hangman that helps users learn the 1-letter amino acid code

Appendix II: eBook features

Vocabuwary: Restricting our usage to an essential core vocabulary is simply not enough, students desperately need help in mastering the language. There are two problems; drilling isn't much fun, and runs the risk of matching a *single* (perhaps meaningless to the student) string of words to accompany each new words. VocabuWary confronts each of these problems. By turning the exercise into a speed-based, scored game, it gives students an incentive to compete against themselves, others, or an instructor-set standard. By employing a (potentially) huge variety of alternate definitions and a number of distinct formats (Is the word used correctly in this sentence? Which definition is correct? Which word matches this definition? Which of these two similar words is the correct one here?) the best strategy is to *understand* the meanings of the words, something the program simultaneously teaches and trains

XWords is an alternative for VocabuWary that provides a weaker learning format but is available to students who are organically challenged with time-based tasks (or who just hate VocabuWary)

Dynamic fill in the blanks: This is a module used in the quizzing program I developed, but would be a powerful addition to an eText. Briefly, it presents users with a paragraph summarizing or introducing a new concept. As students come to a blank, the number of letters is indicated, and the student must type in the correct word. Behind the scenes, it actually accepts any number of sanctioned synonyms. Over time, the a scrambled version of the correct word is revealed and slowly unscrambled ('game-ifying' the experience). Placing a similar activity in-line and having *students* participate in the summarizing of introduced material should offer a more engaging and perhaps effective method of review. FYI, this was inspired from a physical textbook from a course my wife took; the text is
"Rapid interpretation of EKG's" 6th ed. by Dale Drubin

Embedded **video**: My own course is taught with an accompanying lab. While many of the demonstrations (see above) could be done with simple materials in student's homes, it is also true that key points (surface tension, elemental flame colors…) could be somewhat dramatically demonstrated with readily made videos. In an eText context, these would add only to production costs; nothing to per-item student purchases.

Embedded **structure viewer**: Generally, the key aspects of biological molecules are their three dimensional shapes and their surface properties (most notably charges, H-bond donors and acceptors, and neutral areas). While these deeply illuminate complex issues such as filament formation in sickle cell anemia, they can be very helpful in allowing students to think concretely and predictively about simple molecules such as water, amino acids, etc. Especially given Dr. Robert Hanson's work in converting JMOL to HTML-only basis (see previous examples from Hemoglobin and Genetic Disease), it would be relatively trivial to make virtually *every* biological molecule in an introductory eText 'real'.

While deeper issues such as redox potential would be more complicated, representations could be found—while it is a 2D solution, the growing use of showing electron pairs in bonds as closer to one atom than the other is an example of clever clarifying representations:
O-:—H

Embedded **solvelt**: In my view, far too much assessment is based solely on words, and far too much presentation is wholly passive. Solid examples do exist. By way of example, taking this tutorial
http://bcs.whfreeman.com/thelifewire/content/chp08/0802001.html
and making it less passive by requiring the student to PREDICT outcomes based on the two models rather than simply informing them would make for a much more ‘student-as-scientist’ interaction.