Supplemental Material

CBE—Life Sciences Education

Jeffery *et al.*

Supplemental Material

Table of Contents

Beaker Dose Analogy

"…'The ΔG of this state. The ΔG of that state.' What does that mean? ... What does 2 kilocalories mean? Okay. Now I can easily explain what 1.3 kilocalories... …means in drug discovery. What does 1.3 mean in binding constants? 1.3 kilocalories. So let's go with something that people understand: dose. You take- do you take over the counter prescription or prescription medications from time to time? Ibuprofen? Got a headache? You take Ibuprofen right? So what does 1.3 kilocalories mean? So if I have a headache and I take a pill and that particular pain medication is weak for me- or just weak in general - it's not working. Alright. I need to increase the dose by ten-fold so I go from 1 ibuprofen to 10 pills because the interaction strength between the drug and the protein target is weaker. I want to strengthen it by ten-fold. That's 1.3 kilocalories. The $\Delta\Delta G$ between - association constant, you know, going from 10- or dissociation constant- going from 10 micromolar to 100 micromolar is ten-fold. At room temperature that's 1.3 kilocalories. That all the sudden means something now."

Gertrude investigates protein drug shelf-life

Gertrude is interested in the physical and chemical modification processes undergone by lyophilized (i.e. freeze-dried) protein drugs in order to improve drug formulations and enhance shelf-life (C). These drug formulations include excipients, which are inactive substances that serve as vehicles for delivering drugs or other active ingredients. Her research group considers the extent to which protein drugs unfold and how they aggregate when they are unfolded or partially unfolded (H). The degree of unfolding is determined by hydrogen-deuterium exchange (HDX): lyophilized protein powders are exposed to deuterium vapor and the resulting peptide mass is measured via mass spectrometer (M). This data is then used to create representations (A) of deuterium incorporation like structural maps (e.g. see Fig. S1), indicating what regions of the protein drugs remain protected during unfolding (H). Gertrude's first excerpt in Figure S1a below provides a clear example of the presence and integration of the MACH model components, and the implicit role of theory in her explanation:
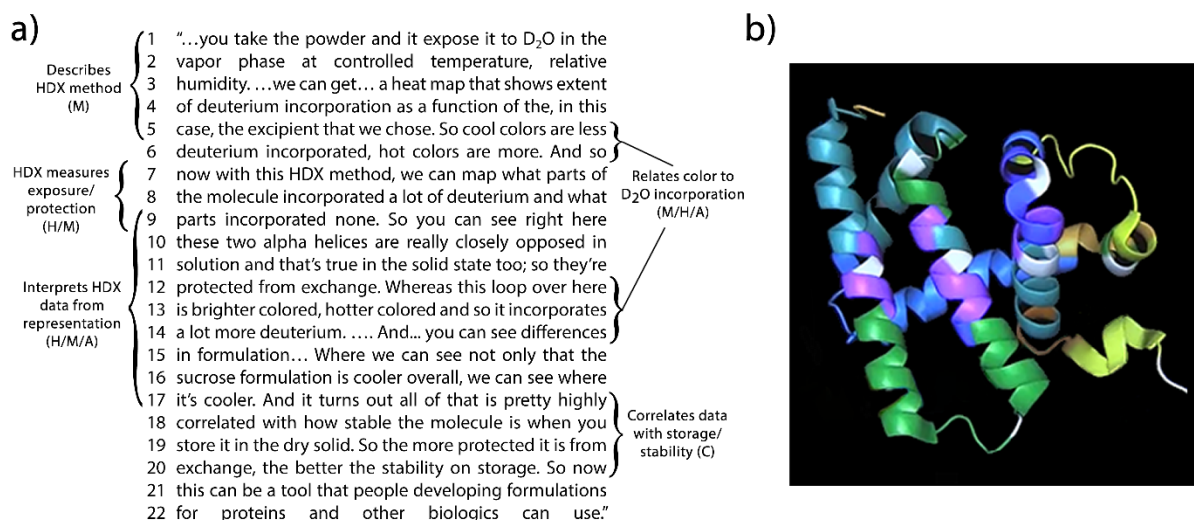

a)

Describes HDX method (M)	1 "…you take the powder and it expose it to D₂O in the 2 vapor phase at controlled temperature, relative 3 humidity. …we can get… a heat map that shows extent 4 of deuterium incorporation as a function of the, in this 5 case, the excipient that we chose. So cool colors are less
HDX measures exposure/ protection (H/M)	6 deuterium incorporated, hot colors are more. And so 7 now with this HDX method, we can map what parts of 8 the molecule incorporated a lot of deuterium and what 9 parts incorporated none. So you can see right here
Interprets HDX data from representation (H/M/A)	10 these two alpha helices are really closely opposed in 11 solution and that's true in the solid state too; so they're 12 protected from exchange. Whereas this loop over here 13 is brighter colored, hotter colored and so it incorporates 14 a lot more deuterium. …. And… you can see differences 15 in formulation… Where we can see not only that the 16 sucrose formulation is cooler overall, we can see where 17 it's cooler. And it turns out all of that is pretty highly 18 correlated with how stable the molecule is when you 19 store it in the dry solid. So the more protected it is from 20 exchange, the better the stability on storage. So now 21 this can be a tool that people developing formulations 22 for proteins and other biologics can use."

Relates color to D₂O incorporation (M/H/A)

Correlates data with storage/ stability (C)

b)

Figure S1: Gertrude maps hydrogen deuterium exchange (HDX) data to a protein drug structure to make predictions about drug stability. Gertrude first outlines how a lyophilized protein drug-excipient powder is exposed to deuterium in the HDX method. Unprotected hydrogens in the protein are exchanged with deuterium and the mass of the protein drug is measured via a mass spectrometer. This data is used to make heat maps of the protein like that in panel b, where gradations of color are used to represent the extent of deuterium incorporation (purple represents 0-10% deuterium uptake; dark blue 11-20%; light blue 21-30%; aquamarine 31-40%; dark green 41-50%; light green 51-60%; yellow 61-70%). Gertrude indicates different parts of the heat map in b as she interprets what parts of the protein are protected from exchange. For example, the two parallel alpha helices on the left display cooler colors, which indicates they are not significantly exposed to deuterium in this formulation. On the upper right, the hotter yellow-colored loop indicates significantly more deuterium incorporation in that region. She explains that this data is correlated with stability on storage so they can compare differences in formulations and make judgments about the stability of drug formulations in a shorter time period. The image in panel b was digitally modified to protect the confidentiality of research data.

In this excerpt, Gertrude makes distinct connections between the data collected (M), how it is represented (A), what entities and interactions it describes in the system (H), and what that implies about functionality (C). Using the MAtCH model (Manuscript Figure 1) as a framework, her discussion generally follows an M-H-C pattern against a backdrop of how she interprets one kind of representation (A). During this process, she implicitly uses theoretical knowledge of protein structure and equilibrium. She begins by describing the procedure of exposing the protein drug powder to deuterium which results in data in the form of degree of deuterium incorporation (M, lines 1-5). Then, connecting the M and H components, she explains that the HDX method (M)

allows her to measure the exposure/protection of regions of protein structure (H, lines 7-9). Portions of the protein molecule (entity) have a certain amount of protection and this can be measured by an increase in mass through the replacement (interaction) of hydrogen with deuterium (entities) (H/M). The M and H components are highly integrated in Gertrude's discussion, with protection from HDX exchange seemingly synonymous with degree of unfolding. As the backdrop for this discussion, Gertrude uses a representation (Fig. S1b) which maps HDX mass spectrometry data (M) directly onto a 3D protein structure where color (A, e.g. lines 5-6, 'brighter' and 'hotter') indicates degree of deuterium incorporation (M) and she can thus interpret degree of exposure/unfolding (H, lines 9-17). This demonstrates how Gertrude cycles between the M, H, and A components of the MAtCH model (lines 6-17), using her theoretical knowledge of protein structure and the HDX process to mediate between them. After establishing the connections between these three components, Gertrude states how the representation (A) of where and how much the protein is protected from exchange (H/M) is correlated with the drugs' stability as a dry solid (C, lines 17-20). Thus, she transitions between the A and C components, and cannot only visually (A) compare the relative stability of the entities in different formulations side by side to address her research problem, but she can also do it in weeks rather than year(s)-long shelf studies (C). Although Gertrude does not explicitly use terms like thermodynamics or kinetics in this excerpt, she tacitly employs theoretical knowledge of equilibrium in her discussion of deuterium incorporation.

Gertrude also examines protein drugs from another perspective by looking at their interactions and organization in space and over time (H). The temporal dimension is a significant part of Gertrude's research, from the context of protein drug shelf-life (C), the kinetics of HDX exchange (M), and folding-unfolding and aggregation equilibria (H). The following excerpt in Figure S2a provide another example of how Gertrude integrates the MACH components and theoretical knowledge. Specifically, Gertrude uses her theoretical knowledge to mediate between the H and M components of the MAtCH model, describing what interactions (H) she believes are measured through deuterium exchange (M). Through the use of a narrative (A), she suggests a hypothetical model of a protein drug in solid (lyophilized) form:
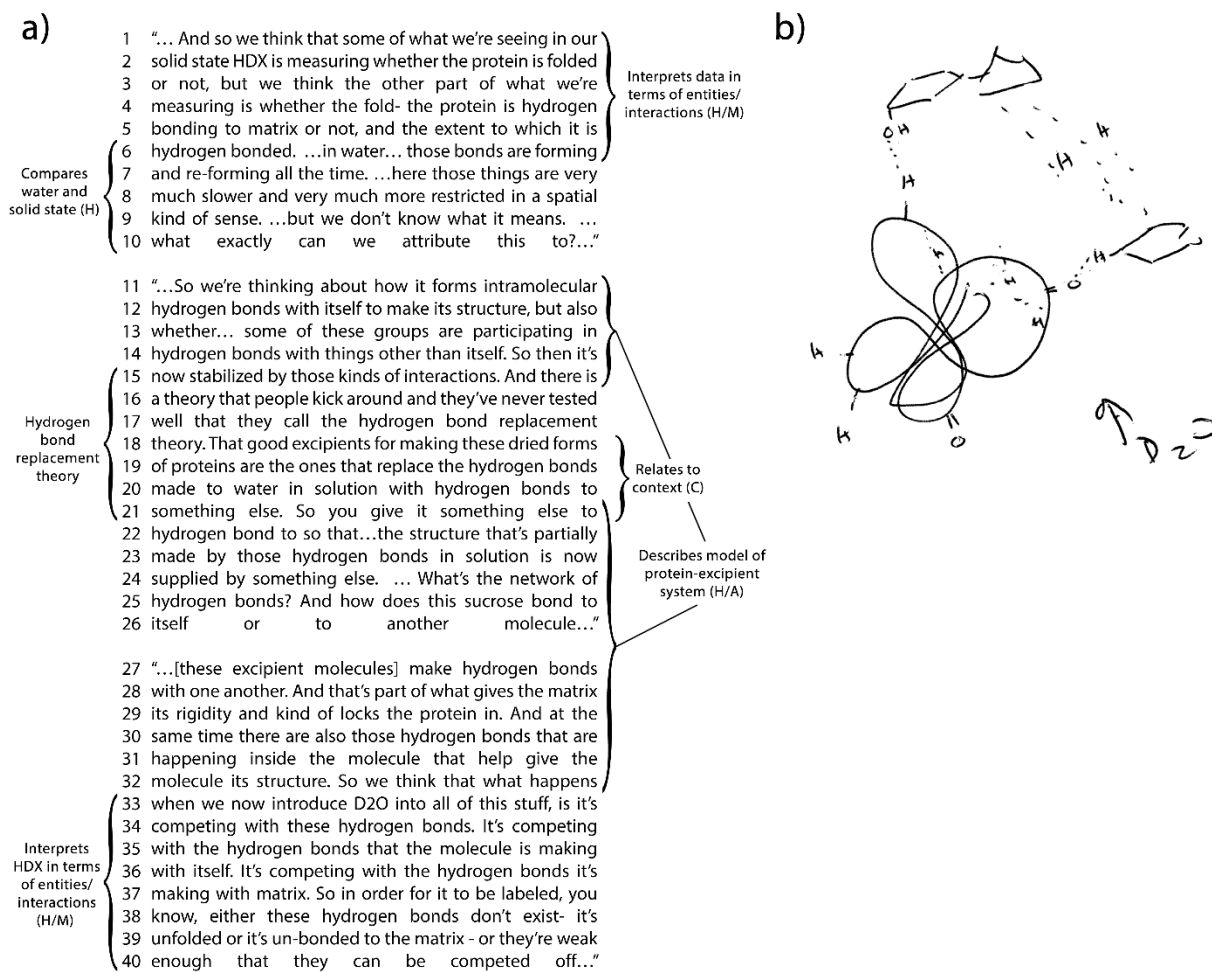
a)

1 "… And so we think that some of what we're seeing in our
2 solid state HDX is measuring whether the protein is folded
3 or not, but we think the other part of what we're
4 measuring is whether the fold- the protein is hydrogen
5 bonding to matrix or not, and the extent to which it is   } Interprets data in terms of entities/ interactions (H/M)
6 hydrogen bonded. …in water… those bonds are forming
7 and re-forming all the time. …here those things are very
8 much slower and very much more restricted in a spatial   } Compares water and solid state (H)
9 kind of sense. …but we don't know what it means. …
10 what   exactly   can   we   attribute   this   to?…"

11 "…So we're thinking about how it forms intramolecular
12 hydrogen bonds with itself to make its structure, but also
13 whether… some of these groups are participating in
14 hydrogen bonds with things other than itself. So then it's
15 now stabilized by those kinds of interactions. And there is
16 a theory that people kick around and they've never tested
17 well that they call the hydrogen bond replacement
18 theory. That good excipients for making these dried forms   } Hydrogen bond replacement theory
19 of proteins are the ones that replace the hydrogen bonds
20 made to water in solution with hydrogen bonds to   } Relates to context (C)
21 something else. So you give it something else to
22 hydrogen bond to so that…the structure that's partially
23 made by those hydrogen bonds in solution is now   } Describes model of protein-excipient system (H/A)
24 supplied by something else. … What's the network of
25 hydrogen bonds? And how does this sucrose bond to
26 itself   or   to   another   molecule…"

27 "…[these excipient molecules] make hydrogen bonds
28 with one another. And that's part of what gives the matrix
29 its rigidity and kind of locks the protein in. And at the
30 same time there are also those hydrogen bonds that are
31 happening inside the molecule that help give the
32 molecule its structure. So we think that what happens
33 when we now introduce D2O into all of this stuff, is it's
34 competing with these hydrogen bonds. It's competing   } Interprets HDX in terms of entities/ interactions (H/M)
35 with the hydrogen bonds that the molecule is making
36 with itself. It's competing with the hydrogen bonds it's
37 making with matrix. So in order for it to be labeled, you
38 know, either these hydrogen bonds don't exist- it's
39 unfolded or it's un-bonded to the matrix - or they're weak
40 enough   that   they   can   be   competed   off…"

b)

Figure S2: Gertrude proposes a possible model of the interactions between protein drugs and excipients in the powdered (solid state) form. She explains that beyond the HDX data indicating if a protein is folded, they believe the data might also indicate if the protein drug is hydrogen bonding to the matrix. Referring to previous water absorption experiments and plots of percent deuterium incorporation v. time, Gertrude explains that while they can interpret the information in terms of protein dynamics (data not presented), they cannot explain at a molecular level why they see a difference in the liquid and solid states. She elaborates on what they think might be happening in the solid state, using an online image of the pentapeptide Leu-enkephalin to explain hydrogen bond donors and acceptors (not pictured). Gertrude then draws a "cheater picture of what's in [her] head" for the liquid (not pictured) and solid states (panel b). The large scribble represents the backbone of a folded protein drug with several hydrogen bond donor and acceptor groups. The ring-like structures on the top and right represent excipient molecules, like sucrose, with donor and acceptor groups. Dotted lines represent possible hydrogen bonding interactions between excipient molecules, between excipient molecules and protein drug, and within the protein drug itself. Gertrude references a 'hydrogen bond replacement theory' in panel a, thus proposing that good excipients protect hydrogen bond donors and acceptors from deuterium exchange, and possibly chemical degradation in general, by participating in hydrogen bonding interactions that would normally be made to water. The arrow labeled 'D2O' in panel b corresponds to Gertrude's final explanation of what they believe happens when deuterium is introduced to the solid state and what it indicates about the system.

In the excerpt provided above, Gertrude cycles through the M, H, and A components of the MAtCH model. We see from Gertrude's initial remarks (lines 1-5) that she interprets the HDX data (M) in light of her theoretical knowledge of the HDX process and theoretical knowledge about hydrogen bonding (interactions), and water, protein, and excipient structures (entities with properties) (H). After establishing the connection between what entities (H) are being measured (M), Gertrude uses her theoretical knowledge to suggest a hypothetical model (in narrative form) of what may occur in the protein-excipient system (H/A, lines 11-15, 21-29). Gertrude integrates

theoretical knowledge of a "hydrogen bond replacement theory" which has been suggested in her field (lines 15-21) with other theoretical knowledge to construct her hypothetical model (A, lines 21-32). She provides a drawing to assist her explanation (A, Fig. S2b). We can also see that Gertrude connects her hypothetical model of the protein-excipient system (H/A) to her research goal of predicting good excipients (C, lines 18-21). As with the first excerpt, it is possible to see from this discussion and representation how Gertrude uses her theoretical knowledge to closely tie her research methods (M) to a hypothetical model of the physical process (H/A). Both implicitly throughout, and at times explicitly (lines 1-5, 33-40), we can see that Gertrude discusses HDX (M) in terms of what it can measure about the scale of unfolding (H), as well as what it implies about the interactions between different entities in the system and the relative strengths of those interactions at the molecular level (H).

In other parts of her interview, Gertrude provides additional examples of how she transitions between the MACH components using theoretical knowledge, often against the backdrop of a representation (A). For example, Gertrude's research group also investigates protein aggregation because proteins that become partially unfolded after lyophilization have a tendency to form aggregates (H) when they are reconstituted, which can cause immune responses in patients (C) (see for example Ratanji, Derrick, Dearman, & Kimber, 2013). Through the use of episodic exposure to deuterium (M), Gertrude can measure what protein regions appear to participate in exchange or are buried during aggregation (H/M). The kinetics and equilibria underlying the episodic incorporation of deuterium into the partially unfolded proteins are particularly important as the relative amount of deuterium that is incorporated over time (M) reflects how fast residues become buried in the aggregated form, as well as where residues are buried (that is, the aggregation interface) (H). As before, theoretical knowledge plays a critical role in this process by allowing Gertrude to mediate between the representation (A) of the measureable world of HDX data (M) and the molecular world of interacting entities (H). However, in this instance, Gertrude uses a special type of line graph called a butterfly plot where information about aggregation (H/M) is not mapped directly onto a protein structure (A) as with the Figure S1, but some structural information in the form of residue number (H) is still provided and combined with percent deuterium incorporation (M). This is sufficient for Gertrude to interpret what the representation (A) implies about the protein system (H). Thus, Gertrude's research enables her to more quickly make inferences regarding which peptide drug formulations will have longer shelf-lives through the application of HDX methods. We can see in the following case how William's efforts similarly aim to improve predictions, but address an entirely different research problem.

William simulates protein dynamics to improve drug metabolism prediction

William's work focuses on incorporating protein dynamics into computational models (M/A) in order to improve predictions about where drug candidates are metabolized and by what enzymes, so as to aid the development of more metabolically stable drugs (C/H). Unlike the other experts interviewed here, William's goal is the development of a predictive method to model possible drug and protein movements and interactions (M/A), which is validated and trained using experimental site metabolism data (M). The end product of his research – a process incorporating a variety of techniques like molecular dynamics simulations, molecular docking, and statistical techniques (M/A) – can then be used to produce data of its own (M). By considering protein dynamics (which he defines as the trajectories of atoms and residues in a protein (H)), he can produce an ensemble of protein structures to represent the multitude of possible conformations and average them to suggest the most likely preferred conformation (M/A). This conformation can then be used in the simulated docking of drug candidates to make predictions (M/A). Because of the goal and computer model-based nature of his research project, the H, M, and A components are extremely integrated in William's discussion and his understanding of thermodynamics similarly appears to intertwine or align with his simulations (A). The MAtCH model allows us to make sense of the complexity by focusing on the connections. In the following excerpt in Figure S3a, we can see how William connects the components, as well as how his understanding of thermodynamics aligns with his simulations (A):

a)

Describes entities / properties (H)

1 "…here you have your catalytic center. And let's
2 imagine you have- and we saw this- …a glutamate,
3 which is more flexible. And if you have, for example,
4 a ligand which- let's see- an aromatic ring here- and
5 we have something in between and, let's say, you
6 have a positively charged amine here. Then, because
7 it's so unspecific, there's a huge amount of different
8 structures which can bind. So what can happen is
9 you might have the same aromatic ring for another
10 compound, but it has a much larger chain to the
11 charged amine. So in order to stabilize, what you, for
12 example, see is that this glutamate is changing its
13 side chain and now stabilizes with this negatively
14 charged- this amine for other compound as this here
15 is much longer than this part. But if you don't include
16 this protein dynamics, you would not be able to
17 predict this compound in the same way, or in the
18 same close proximity to the catalytic center than if
19 you used just this conformation of your glutamate
20 residue. …but then you would just show… protein
21 structures with the ligand and overlay them with the
22 static structure to show them what kind of dynamic is
23 involved and how this is really critical for making
24 better   predictions   for   drug   metabolism."

Describes changing organization (H)

Dynamics affect organization (H/M)

Importance of dynamics to prediction (C)

b)

Figure S3: William provides an example of the binding of two different drug compounds. To illustrate the significance of including protein dynamics in simulations, William asks to imagine a binding site, pictured in panel b, which has a specific flexible glutamate residue some distance away from the catalytic center. He explains that the glutamate residue can change its conformation to stabilize different drug compounds (aromatic rings with hydrocarbon chains of different lengths ending in amine groups). William argues that if protein dynamics – like the changing conformation of the glutamate residue – are not included, it is not possible to predict how the ligand interacts with the catalytic center. Thus, in papers, William shows how the inclusion of protein dynamics in simulations leads to better drug prediction by overlaying images of predicted protein-with-ligand structures over the static structure of the protein.

In this particular excerpt, William's discussion generally follows an H-M/A-C pattern against the backdrop of a representation (A, Fig. S3b). William begins by describing the significant structural

components of the binding site, their properties, and two hypothetical drug compounds (H, lines 1-11). He then explains how those residues might change their spatial organization to accommodate different drug compounds (H, lines 11-15; see Fig. S3b) and thus alter a compound's distance in relation to the catalytic center (H/M, lines 15-20). He argues that because of this, including dynamics in simulations (M/A) is critical to improving the predictive capabilities of current methods and thus aiding the drug discovery process (C, lines 15-17, 22-24). William's tacit use of theoretical knowledge allows him to productively mediate between the measurable world (M) and what it implies about the molecular world of (simulated) protein structures and their interactions (H/A). We also begin to see in the above excerpt that William relates residue flexibility to protein dynamics, but what is not yet apparent is his unique way of assigning meaning to theoretical thermodynamic concepts. The following excerpt in Figure S4a provides an example of how William maps meaning onto mathematical models and symbols (A), as well as how he applies his theoretical knowledge of thermodynamics, particularly of enthalpy, entropy, and free energy, to the context of developing protein dynamics simulations (M/A):
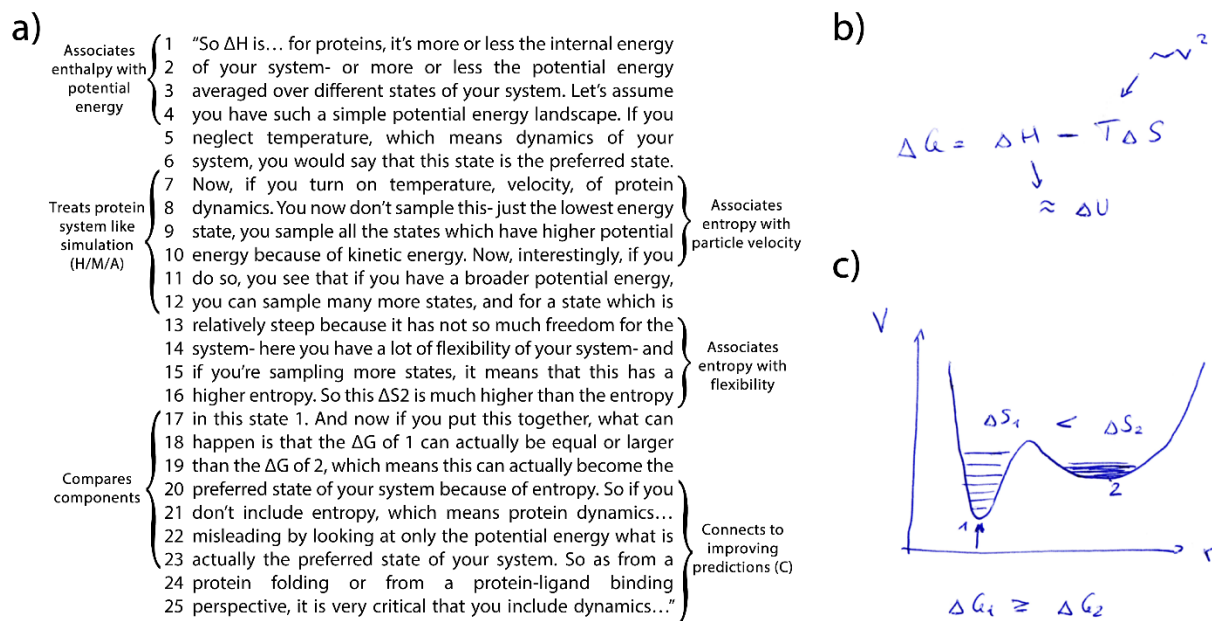

a)

Associates enthalpy with potential energy

1 "So ΔH is… for proteins, it's more or less the internal energy
2 of your system- or more or less the potential energy
3 averaged over different states of your system. Let's assume
4 you have such a simple potential energy landscape. If you
5 neglect temperature, which means dynamics of your
6 system, you would say that this state is the preferred state.

Treats protein system like simulation (H/M/A)

7 Now, if you turn on temperature, velocity, of protein
8 dynamics. You now don't sample this- just the lowest energy
9 state, you sample all the states which have higher potential
10 energy because of kinetic energy. Now, interestingly, if you
11 do so, you see that if you have a broader potential energy,
12 you can sample many more states, and for a state which is

Associates entropy with particle velocity

13 relatively steep because it has not so much freedom for the
14 system- here you have a lot of flexibility of your system- and
15 if you're sampling more states, it means that this has a
16 higher entropy. So this ΔS2 is much higher than the entropy

Associates entropy with flexibility

Compares components

17 in this state 1. And now if you put this together, what can
18 happen is that the ΔG of 1 can actually be equal or larger
19 than the ΔG of 2, which means this can actually become the
20 preferred state of your system because of entropy. So if you
21 don't include entropy, which means protein dynamics…
22 misleading by looking at only the potential energy what is
23 actually the preferred state of your system. So as from a
24 protein folding or from a protein-ligand binding
25 perspective, it is very critical that you include dynamics…"

Connects to improving predictions (C)

b)

$$\Delta G = \Delta H - T \Delta S$$

$\sim v^2$

$\approx \Delta U$

c)

V

$\Delta S_1 < \Delta S_2$

$\Delta G_1 \gtrsim \Delta G_2$

Figure S4 William explains the influence of entropy on free energy in a two-state protein system. William begins by substituting variables in the Gibbs' free energy equation in panel b with other variables that have physical meaning: $\sim v^2$ to indicate particle movement for the $T\Delta S$ term and ΔU for averaged potential energy or interactions (strengths) for the ΔH term. He proposes looking at a simple energy landscape of a two-state protein system, shown in panel c, and explains that the preferred state of the system can change if you ignore the dynamics of the protein by 'turning off' temperature; that is, you ignore the $T\Delta S$ term of the formula in panel b. If temperature is considered, there are many more states that can be sampled. These states are represented as lines within the wells of the graph in c. William explains that the widths of the wells in panel b are a function of greater particle velocity or kinetic energy due to temperature: the greater the breadth of the well, the more possible states can exist, which reflects greater flexibility or freedom of movement. So-called "steep" states have limited flexibility and thus limited states to sample. In the second half of the excerpt, William considers entropy and compares the different ΔG values of the two protein states to illustrate how disregarding entropy in simulations can be misleading and thus the inclusion of dynamics in predictive methods is critical.

Although William integrates the H, M, and A components extensively, using the MAtCH model allows us to make sense of this excerpt by semi-isolating the components. It is first important to note that throughout this excerpt William structures his explanation around two representations (A, see Fig. S4b & c) in addition to talking about a protein system (H) like it is one of his

simulations (M/A, e.g. lines 7-12). He uses his theoretical knowledge of thermodynamics to seamlessly map a description of the states of a protein system (H) to his representation/ simulation (M/A). In doing so, William assigns meaning to mathematical models by mapping entities and their variable states (H) onto particular symbols in the formula and graph (A, e.g. lines 1-3, 16-20 ; also see Fig. S4b & c). William's understandings of enthalpy, entropy, and free energy appear to align with his simulations (M/A) and are mapped to the entities, interactions, and states of a protein-drug system (H). He makes the concept of entropy tangible as "How much an object is moving. How dynamic it is…" (i.e. structural flexibility; lines 13-16) and connects it to temperature and the velocity of particles (H, lines 7-10). He describes enthalpy as internal or potential energy in this excerpt, but also associates it with the sum of interactions and interaction strength (H) in other parts of the interview. In the above excerpt, states both entropy and enthalpy must be considered in order to determine the actual preferred state of the system (lines 17-23). That is, free energy involves "compensation" between interaction strength and protein flexibility. In his simulations (A), temperature can be "turn[ed] on" to allow protein dynamics (entropy) and the resulting different states have different kinds of interactions (enthalpy) (H). William explains that if protein dynamics are ignored, "…you don't have entropy, you're not calculating ΔG's… and ΔG finally determines what… states you observe in nature 'cause we're not living at 0 Kelvin", resulting in incorrect predictions for ligand binding (M/A). Without a method that approximates reality well (M/A), William cannot make reliable predictions about drug candidates (C, lines 20-25).

William also discusses the difficulties students in his research group seem to have interpreting data from simulations (M) and how he must guide them to relate the trajectories of simulated atoms (H/A) to what they might reveal about the simulated system (A):

> "…if you get the statistical analysis out, they stop looking at the- at trajectory- at the atoms moving itself completely. So I have constantly students who will say, 'Oh! I have run the MD simulation. Here's the free energy,' or 'Here's the free energy profile.' And then… so you have your simulation and then they'll look at the ΔG over time and say, 'Oh yeah it goes first up.' And then they see a jump. And then it's equilibrium and then I'll say, 'Wow. What is this jump to you?' And they'll say, 'I don't know.' But didn't you look at the trajectory? Didn't you look really qualitatively at what is happening in the movie? And the structure. And they'll, 'No, I didn't. I just did the analysis.' … Once students have the feeling that they have an analysis to it and they get the values out, they're happy with this. They don't look back- and I don't know if it's hard for them to look at the trajectory and identify what is going on there, or if they're just happy that they have a good quantity coming out? But it is surprising. So I always say, 'Look at the structure. Really look at the- Look at the raw data. Look at the raw data to explain what is going on in the system.' …. It's important and it's valuable information but I want to understand what is the basis of getting this data. And people forget analyzing this part of- so it's in principle like, you're doing an experiment, you do the analysis and you're not really interested in what could happen in the experiment because certain… things [can] be wrong. But if you don't look at the raw data you don't see it."

For William, connecting the H, M, and A components is obvious. He tacitly uses his theoretical knowledge of thermodynamics to mediate between the measurable world by interpreting the data (M) in terms of what it implies about the (simulated) protein system (H/A). He explains how a

change in free energy on a graph (A) reflects underlying changes in structural movement and/or the formation of new interactions (H) in the simulation (A). It also indicates he must look at the simulated protein system (A) in order to interpret the possible structural cause (H/A) of the data (M). According to William, while producing a numerical or graphical output is doable for students in his lab, interpreting and making connections between the data (M) and the underlying (simulated) physical causes (H/A) is not as obvious. Thus, a combination of experimental and simulated data enables William to improve current methods used to predict the metabolism of drug candidates.

Table 1: Examples of ways in which selected protein folding and dynamics educational materials could be modified using the MAtCH model as a guiding framework. Both manuscript and relevant supplementary material were considered. The most relevant components or connections that the example questions address are indicated in parentheses, although it should be noted that most questions require the application of some amount of theoretical knowledge and may elicit other components.

MAtCH Evaluation	Possible Modifications
Exploring protein structure and dynamics through a project-oriented biochemistry laboratory module *Lipchock, Ginther, Douglas, Bird & Loria (2017)*	
- Provides a social context although module does not ask questions in terms of it (C) - Provides protocols and describes theory of how they work, but does not discuss limitations or alternative methods (M) - Representations are produced and analyzed, but little time is spent discussing purposes, affordances, or limitations (A) - Students are given a hypothesis, rather than producing their own (C) - Practice-oriented, no discussion of thermodynamics and only moderate discussion of kinetics (t)	- What other appropriate methods exist for studying protein structure and dynamics? (M) - Discuss similarities and differences between the acid loops and P-loops of the five protein tyrosine phosphatase sequences you aligned. (H) - What information can be communicated through the ribbon structure of PTP1B you develop in Experiment 1? What are its limitations? (A) - Design the forward and reverse primer sequences for the site-directed mutagenesis of PTP1B. (H/M) - Explain how melting temperature is calculated (H/M/A). - Explain why commercial vectors often contain *lac repressor* sequences. What other kinds of repressor/operator systems are used and in what research contexts? (M, C) - Often the DNA produced through transformation and amplification of a PCR product is sent for sequencing to confirm synthesis of the desired mutation. Why is this necessary? What issues are associated with PCR? (M) - How does one decide on the ratio of bisacrylamide and acrylamide for a polymerization reaction? (M) - What are the purposes of each of the four buffers used in purification of PTP1B? (M) - How does purification of a soluble, well-folded protein differ from purification of natively insoluble or unfolded proteins? Briefly explain the theory behind at least two different methods. (H/M) - Describe the process of creating a Bradford calibration curve with BSA. Explain your choice of wavelength, standard concentrations, and any decisions you made while creating your graph. (H/M/A) - What is the purpose of each of the samples loaded into your gel for SDS-PAGE analysis in Experiment 8? Is there any reason for their order? (M/A) - What information about PTP1B can be obtained from your stained gel? What cannot? (A/H) - Discuss error inherent in kinetic analysis of PTP1B. How is this error summarized in the representation of your average reaction rates? (M, A) - Discuss the fit of your data to the Michaelis-Menten equation. (M/A) - You produced several representations over the course of this project. Explain what each of these representations tells you about the protein you are studying. (A/H) - How does the data you obtained over the course of this project extend characterization of PTP1B catalysis? (M/C/H) - Compare and contrast the methods used in this project with other methods for studying protein structure and dynamics. What are their limitations? What can and what can they not tell you about a protein? (H/M)

	• If you were to conduct further studies on PTP1B (or a similar PTP), what would you do? Explain why you chose those research methods/goals. (C/H/M) • Using literature, identify another enzyme for which protein motions have been shown to be important for function. What is currently known about this enzyme and what research problems or goals currently exist? (H/C) • To what other research could you apply the methods you used in this project? (C, M) • If your aim was to understand more about how PTP1B interacts with its substrate, what would you study (e.g. properties)? What current theories or models would you consider? (H)
Demonstration of AutoDock as an educational tool for drug discovery ***Helgren & Hagen (2017)***	
• Provides opportunity to explore 3D structures of CDK2 and CDK2 inhibitors (A) • Use docking software and apply fragment growth to hit molecules (M/A) • Situates methods in context of drug discovery practices and a specific target molecule, CDK2, but little specific background regarding the latter (M, C) • Introduces a variety of methods used over the course of the drug discovery process (M) • Discusses how models for use in AutoDock are developed/modified (A/M) • Thorough description of how to use software like AutoDock and AutoGrid to modify the protein model (M/A) • Software produces a variety of representations carrying information about the receptor (A) • Limited discussion of affordances and limitations of models/representations (A)	• What interactions/distances are significant to your reasonable docking poses? (H/M/A) • What kinds of modifications can be applied compounds to affect their binding affinity? Explain. (H/M) • Make a recommendation for a compound based on the docking poses you produced. (A/C) • What additional experiments are appropriate after identifying a viable compound(s)? (M/H/C) • What other research problems or contexts employ computational methods as part of their methodologies, and at what point(s) are they used? (C, M) • Explain how dissociation constant and inhibitory concentration resulting in 50% activity reduction (IC_{50}) are related. (H/M) • Explain what information about entities and interactions can be obtained from the methods you used. (H/M) • Discuss any similarities or differences across the possible inhibitors and their interactions with the CDK2 protein. (H) • You modified the CDK2 receptor prior to docking. Discuss these modifications in terms of how well the AutoDock model represents the cellular or *in vivo* environment. (A/H) • How are docking scores calculated? What concepts and/or mathematics underlie score calculation? (M/A) • Explain how variability in ligand and receptor conformations during docking can affect your predictions. Are there implications for your research problem/goal? (M/A/C) • Docking runs can predict highly-scored but physically impossible docking poses, and duplicate docking runs can produce different results. What factors lead to this and how can you 'trust' your results? (M) • Explain the implications of measuring a binding affinity that is overly high or overly low. (M/H/C) • How do *in vivo*, *in vitro*, and *in silico* drug discovery efforts differ? (M/A/C)
Understanding structure: A computer-based macromolecular biochemistry lab activity ***McLaughlin (2017)***	
• Introduces origin of X-ray crystallographic images,	• What is the biological significance of the incorrect residues in the mutated model? How might those mutations affect the structure? (C/H)

but provides little social or biological context (C) • Draws connections between electron density maps (A) and amino acid residues (H), but no discussion of how electron density data is measured (M) • Produce images of corrected amino acid residues (A), but do not analyze changes in terms of social or biological implications (C) • Limited discussion of the limitations of X-ray crystallography (M) and the theory (t) behind it • Limited discussion of the accuracy of X-ray crystallography and computer protein models (M/A) • Provides practice using PyMol and *Coot* (M/A), but limited discussion about what these models (A) can describe about interactions and functions (H)	• What are the limitations of the methods used in this activity? What can and what can they not tell you about a protein? (H/M) • What does it mean for a residue to lack electron density? How can such residues be differentiated from mutated residues? (H/M/A) • What are other reasons protein structure refinement software is used? (M/A/C) • What is the purpose of crystallizing proteins to develop protein models? / How can protein models developed from crystallization be used? (M/A/C) • Are there any other strategies to aid crystallization? (M) • How accurately do PyMol and X-ray crystallography models represent the *in vivo* or native state of the protein they represent? (H/A) • What is the purpose of the crystallization solutions used in preparing your crystal tray? (M) • With references, describe two contexts where protein models are used to address research goals. (A/C) • Identify a current area of research which employs similar methods and describe what entities and interactions it investigates. (C/H/M)

Purpose	Interview Question
Phase 1: Exhaust description of research; freeform explanation	**1. Explain your research as you would to a colleague, somebody who is in a related or similar field. Feel free to sketch or show any representations during your explanation. (Let them answer/draw/etc. freely.)** a. Why did you choose to study that (topic of interest)? b. Is there a particular way you want to apply this research? {For clarification: That is, why is this research important, such as to organisms or to society?} **2. What is the role of the living environment** (i.e. the in vivo) in your research?
Phase 2: Probe description of research methods, data, and how data is processed	**1. Can you explain in detail how you study this? (For clarification: …in terms of your data collection, your methods, etc.? How do you actually *do* the science that you do?) (Answer freely.)** a. What kind of data do you collect? i.e. Where does your data come from? **(data source)** Do you use data from other sources (e.g. PDB files) to supplement your own data? If so, where from and how? b. Do you take thermodynamic or kinetic measurements? c. What experimental methods do you use to collect data? **(data collection)** **2. (So) What kind of information do each of those techniques give you? What kind of information do you get from those sources?** **3. Do you use any sort of modeling in studying your protein? (What do you do? How do you use them?)** a. (If applicable) At what stages do you use those models (source, collection, analysis)? b. (If applicable) Can you draw or show the model(s) and describe how you use them? Can you explain how the information for your models/simulation(s) develops from your data OR how your simulation is used as data to explain the phenomenon you study? c. (Limitations) What is this model useful for and what is it not useful for? (What can it do or not do?) **4. How do you analyze your data? (data analysis)** a. When you analyze your data, how does that data help you develop an explanation? How do you piece together the data that you collect and the theoretical aspects of your work? b. How do you represent that data? If you're writing up a paper and in the results section, what sort of data would you present (to communicate your findings)? Do you use (indicate previous drawings) or…? Can you draw an example? i. For the representations that you use (to think about what you do or in publications), what sorts of limitations do they have? Do they communicate too much, too little…?
Phase 3: Probe for additional representations	**1. When you think about your research or when you're trying to explain it, what do you visualize? What do you picture in your mind or draw? Can you draw it for me? (Answer freely).** a. (Clarification) Do you use this/that as a tool for thinking about it during experimentation? Or as a representation for publication? **2. Let's see, you mentioned... (summarize to confirm that you understood their drawings). Apart from those examples, do you use any other visuals in your explanations?**
Phase 4: Research explanation to an upper-level undergraduate student	**1. Could you explain your research like you would to an upper-level undergraduate student (specifically to student in a 300-400-level course)?** **2. Could you tell me a bit more about how you would explain protein folding in general to a student?** a. Would you use entropy to explain (protein folding/dynamics)? If so, how? (Feel free to draw.) b. Would you use enthalpy to explain? If so, how? (Feel free to draw.) c. Would you use free energy to explain? If so, how? (Feel free to draw.) d. You mentioned the concept _____. Can you draw and explain how you would explain that concept in the context of protein folding? e. (If necessary) How would you describe the methods used to get your data in the classroom? Feel free to draw any pictures you would use. **3. That covers everything I wanted to ask. Is there anything else you would like to tell me?**