Mike: OKay. So, there's this number, free energy of insertion. And I think it will be elevated at protein-protein interfaces.

Daisy: This is an inference problem: you have some evidence, this ienergy, and there's this property you want to infer. Let's fill out Bayes theorem.

Mike: Okay. So.. I'm inferring this property. It's a property of a residue: "at an interfacial site", or "not at an interfacial site". And the evidence is also a property of a residue. A numerical property of a residue, though, I never learned how to deal with that...

Daisy: Later, once we've got it all set up. Just make up a cutoff now, so it's a binary property, and we'll fix it later.

Mike: Okay. "Elevated" vs "Not elevated." \*draws a diagram\*

Daisy: And "elevated" is evidence of "interfacial".

Mike: Yes. Exactly.

Daisy: Is that the only thing that's evidence of "interfacial"?

Mike: No... inward-facing residues are... no, they can be interfacial. But when they're elevated, it's not evidence for interfacial. So really what's evidence is "outward-facing & elevated".

Daisy: well, this is an energy of lipid interaction. Is it even really right to say that inward-opnesa re elevated? You wouldn't assign them an energy with this energy funciton, you'd be running the energy function on false pretenses.

Mike: Okay. So I'm not really calculating their energy of interaction with their environment. I'm calculating a hypotetical lipid interaction energy; the energy that they *would* have if they *were* interacting with lipid.

But... is that... reducible, somehow? To something... Bayesian?

Daisy: Right, like you're saying, the idea of something being hypothetical shouldn't just be something you say about E, it should be represented in the probabilities?

Mike: Yeah. Exactly.

Daisy: Hmm. ... I'm inclined... not to think so?

Mike: Sure. Lipid interaction energy given inward facing, vs. lipid interaction energy given not inward facing.

Diasy: That's not the "given" symbol in the probability operators. You can define different properties: a lipid interaction energy, these two hypothetical lipid interaction energies... and that's similar to how we talk about hypothetical probabilities, conditional probabilities. And... and certainly, the "probability that lipid interaction energy is five" changes whetehr it's conditional on winward-facing, vs out-ward-facing. So does "hypothetical out-ward-facing interaction energy", in fact: it's likely to be lower, given outward-facign.

Mike: well, inward-facing doesn't interact with lipid, there would be an energy of interactionw ith surroundings, but not with liipid.

Daisy: No, there is an energy, it's zero, not, like, somehow undefined. There are still forces that hold between the two. Think of oppositely charged points, as the distance between them approaches zero.

Mike: Energy of interaction approaches zero. Got it.

...

Mike: But, this is definitely true. "P(Energy of interaction with surroundings = 5|inward-facing)" depends upon an estimate of its interactions with other residues. "P(Energy of interaction with surroundings=5|outward-facing)" depends upon an estimtate of its lipid interaction energy.

Daisy: Yes. That's true.

Mike: Okay. Maybe we can fill out Bayes theorem now.

...

Okay. This is pretty simple. I just need to know the prior probability that a residue is interfacial. And I need to know the true-positive and false-positive rates - the probability that an interfacial residue will have elevated hypothetical interaction energy, and the probability that a non-interfacial residue will have elevated interaction energy.

Daisy: Alright. Let's restrict our attention to the beta-strand region. And... you said that there's other evidence involved... whether it's outward-facign, fo rexample. Let's restrict our attention to only the most obvious cases.

Mike: Alright. In fact, let's not even think about depth-dependence for now. An outward-facing residue right in the middle of the membrane, part of a β strand. I like that especially because there's experimental evidence concerning their lipid interaction energies.  
But... the chance that it's elevated? What's that mean?

Daisy: Well, think of it as a family of statements. E5 means above 5 joules. E10 means above 10 joules.

Mike: OKay. So... I want to think about .

Daisy: You said it was experimentally determined. You could just plot it using a dataset of monomeric proteins.

Mike: If there are monomeric proteins. First, though, I want to see if I can derive it theoretically.

Daisy: Using what theory? Physics, I guess? Physical chemistry?

Mike: No, natural selection! Or... what's conventional, which is just to assume a Boltzmann distribution of energies and then count hte frequencies.

Daisy: So, you want to derive the energies, using the measured frequencies, rather than the measured energies.

Mike: Yeah. People do that a lot. I want to know the theory behind that. Also, it's more general: the measured energies are only for a single depth.

Daisy: Okay. This is good, actually. It seems like you have a lot of different ways to derive the same information. You have physical modelling, with ProtCAD or Amber or something...

Mike: The hard part there is to get a model of a membrane, and then let the simulation run enough that I've got the energy of a typical interaction, rather than withthe molecules in whatever shapes I happened to program in.

Daisy: Okay. So, that's hard, but it's the most reliable.

Mike: Not necessarily. All of our uncertainty about what the membrane is like becomes uncertainty in our energy prediction.

Daisy: Mike. Don't just dismiss applying the best theory available.

Mike: even when you have a good theory, you can't use a reductionist theory if you don't have the reduction available!

Daisy: Have you tried?

Mike: no.

Daisy: Have youe ver even looked for a phsyical model of a membrane?

Mike: No...

Daisy: Alright. So don't dismiss it. Somebody might have already estimated the lipid interaction energies with a simulation like that. A theory as good as physics is worth the look.  
So, there's physics simulations. There's...

Mike: There's something almost... childish about physics. Like, the functions are so simple. All these simple polynomial functions of distance.

Daisy: That only makes it more likely to be true. And they're well-tested, which is more important than their simplicity.

Mike: I heard Manfred Sippl tried to apply Lennerd-Jones potentials to explain protein structure and it didn't work.

Daisy: That's... very interesting, actually.

Mike: Yeah. Like... here's an example Vik gave me. A dehydrated hydrogen bond. Some protein-protein interfaces apparently have them... I should read more about this, actually, since the serines at β-barrel membrane protein interfaces are probably dehydrated hydrogen bonds. But... okay, think of a serine at a porin interfaces... surrounded by hydrophobic side chains. So, if it participates in a hydrogen bond, it's going to be a much stronger one than would happen in watter, because it's not competing with water, it's in the middle of a membrane... or something? But the real point here is that there's no such thing as an interaction energy for hydrogen bonds. It depends upon their surroundings.

Daisy: There's nothing you just described that wouldn't be captured by a molecular dynamics simulation. The lack of water would be captured in a molecular dynamics simulation. The only thing that wouldn't be aare things that involve the energies of electrons. Or a lot of polarizability, I hear that can't be aptured in molecular dynamics simulations very well yet either. And all of that would be perfectly captured, hypothetically, by a quantum mechanics simulation. Physics is not inaccurate. There are only innaccurate approximations to it. So don't dismiss the possibility that there are accurate approximations for estimating lipid interaction energy in β-barrels. Whatever Sippl was doing might be good to look into, so you odn't try it. But approximating simple, well-tested models is not something to rule out without investigation. It might work, so the value of information is high.  
So, there's that. Physics modelling, which, not that we get into it, is actually a whole range of possible approximations. There's... this Boltzmann thing? What is that, exactly?

Mike: It's something Sippl did...

Daisy: It's something a lot of people do, you said. Talk about the science, not the people.

Mike: Okay. A lot of... well-regarded knowledge-based potentials... you can't just talk about the science. The evidence that I *have* is that people like these potentials.

Daisy: Do you think that scientists like models taht don't work?

Mike: Honestly, Daisy, I'm not sure, and I would really like to become sure.People talk about "science" as if it's one thing... there's an equivocation here. People like Salmon say, "hooray science, the transistor, the internal combusion engine, antiretrovirals". But that's still what we would see if 90% of academic science was complete nonsense, as we *know* tha tmuch of the theory of practicing psychoanalysts was, for a long time.

Daisy: Engineers must have some way of telling good science from bad science, since it's fairly easy to test, for example, the battery life of a battery or the brightness of a screen.

Mike: Yes, but *I* haven't learned to do it yet! So what am I to make of the affection for knowledge-based potentials?

Daisy: What about the excellent correspondence between weakly stable strands and contacts in a partial model?

Mike: Yes... they tested it. They didn't test it perfectly, but they tested it sufficiently. They found... elevated meant non-native. As if htere's... see, I don't know the underlying theory that connects this to energy, actually.

Daisy: Sure you do. High energy, so it's not the thermodynamic minimum. Anfinsen's theory.

Mike: But.... why are is everything *just stable enough*? How come there aren't superstable regions that I can't recognize *as* elevated? Sure, of course non-native will be elevated relative to the native state. But why is it elevated relative to the *mean energy of the protein?*

Daisy: Interesting question. Maybe make some evolutionary models?

Mike: So quick to recommend new work!

Daisy: Better than laboring over something on mistaken foundations.

Mike: Vik knows the foundations. Maybe I should just write exactly the programs he tells me to. But, then again, Vik wants to educate me, and I want to learn. Maybe in the first year of grad school, just doing what they tell me is teh right thing to do, but I defintiely don't want it to be my habit.

Daisy: Alright. So, there's this question of, what is the empirical support for knowlecge-based potentials? And don't forget, once you know the empirical support, you also know the domain of applicability.

Mike: Yes... thiough, still, Vik knows the domain of applicability. I don't *need* to, for my research. I just want to.

Daisy: Which is fine.

Mike: Also, I don't need to make evolutoinary models, necessarily. The introduciton to the most recent β-barrel structure discusses this.

Daisy: Check it out.

Mike: Of course. Alright. SO physics, boltzmann...

Daisy: Yes... what exactly is this boltzmann idea?

Mike: Basically... there's this relationship between energy and frequency that holds in many systems. Gases, for example. And it also holds between the frequencies of inserted vs. not-inserted transmembrane helices, and the energy of the inserted vs. not inserted state, according to a paper that I would like to find and read thoroughly. The relationship is, that if N1 and N2 are the number of particles in state 1 and state 2, and E1 - E2 is the difference in Gibbs free energies between these two states, then .

Daisy: Hmm... I've beeen thinking, does specifying ther atio between any pair of probabilities specify the probability distribution? It seems like with normal numbers, no, because you can scale the whole thing by a constant, but maybe that's the only change that's possible. So thee's only a one-dimensional space of possibilities. Adding the additional constraint that the sum ofthe probabilities is one removes a degree of freedom, and then you're left with a unique solution.

Mike: Later? Please?

Daisy: Okay. Tell me more about Boltzmann.

Mike: Well, basically, Sippl... I mean, scientists... I mean, basically, there are a *lot* of circumstances where, if you assume that this relationship holds, you get what looks like a good energy function. The one example I know of is TmSIP, and its success in finding non-native broken contacts.

Daisy: That was a... how should I say it?

Mike: Let's call it a knowledge-based potential. As far as I know, they all assume this Boltzmann relation.

Daisy: Or a Boltzmann-derived potential.

Mike: Sure. A boltzmann-derived potential. I have a paper on the theory ehind them, but I haven't read it. I really don't know why they use this particular relation.

Daisy: Okay, so that's what we know about the Boltzmann potentials. So. There's this Boltzmann assumption, there's real physics, there's... wasn't there something else?

Mike: Yes... make a natural selection model. I've given it a little thought, and here's what I thought of. Suppose that a bacterium uses energy for two things: producing membrane protein B, or growing. And there's a whole population of these bacteria, with different versions of membrane protein B. And, there's various things that can happen to B. It can misfold, or, it can pop out of the membrane and be lost. Whenever this happens, the bacterium needs to replace B, which costs energy it could otherwise be using for reproduction. The energy of lipid interaction somehow determines whether it pops out. And, maybe it also helpds determine whether it folds? That seems like a much more complicated question. But you would end up with different mutations, and in general different sequences, ahving different frequencies. And you can run this simulation, or maybe there's a good analytic approach.

Daisy: That's awesome. And then you would get the frequency distribution implied by the assumption that the lipid interacting surface is optimized for insertion.

Mike: Yes. Exactly.

Daisy: Okay. So you have four approaches to the enrgy, for htis one slice through the membrane. Experimentally derived energies from mutation studies.

Mike: The more the better.

Daisy: Yes, the more the better. Physical modeling. The Boltzmann distribution. And evolutionary models.

Mike: All this for that one probability, .

Daisy: That's the wonderful thing about Bayes' theorem. It gets your priorities straight. John Platt thought strong inference got your prioriteis straight. And he's right, but it can't generate a hypothesis for what computational method to use, whattheory to use. Platt only knew what to do when the hypotheses were obvious.

Mike: So, I'll get to it. I'll read those papers, and I'll see what probabilities of elevated energy I get from each estimated energy.