INTERVIEW

# Interview with David A. Case: on force fields, biomolecular modeling, and NMR

David A. Case*

*Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, United States*

Nikolai R. Skrynnikov*

*Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*
*& Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia*

* Correspondence to: david.case@rutgers.edu and nikolai@purdue.edu

Prof. David Case is based in the Department of Chemistry and Chemical Biology at Rutgers University. In this interview he speaks about force field development and validation, predictive power of MD models, combining physics-based force fields with empirical scoring functions, MD modeling of disordered proteins and RNA, synergies between MD simulations and NMR measurements, and chemical shift calculations. The video of this interview is available on the YouTube site (URL https://www.youtube.com/watch?v=bumGM1CWA7Y).

**I. How would you compare the capabilities of the physics-based MD methods and methods employing empirical scoring functions (e.g. ROSETTA, pharmacophore-based docking programs)? Are there any potential synergies between these methods?**

In the field of small molecule - protein interactions, there is no sharp dividing line between different approaches, but rather a continuum. Everything is empirical – even things that are called "physics-based" are based upon experiment and empiricism and simplification. But there are some force fields, like the AMBER force field, or CHARMM, or OPLS force field, that are intended to be very general and operate by building up from small molecules to make large molecules. The term "physics based" gets applied to those; it is a jargony term but it does represent the expected wide relevance and range of applications. You should be able to take a protein under lots of different circumstances and use those force fields and learn something about the behavior of the protein. As you go to more empiricism, in a ROSETTA-like setting, the scoring functions come up (and that is another jargony term) that are at the other end of the spectrum. A scoring function might only work for particular types of proteins – it might be a kinase-specific scoring function. Or it might explicitly take into account things that we see (knowledge-based potentials), i.e. propensities that we see in experiment that we do not really understand from a fundamental view. Yet we want to score more favorably those types of conformations that we see a lot.

The implication of the question is exactly right: There are synergies that can be exploited. We are working on that with the ROSETTA team and the AMBER team. We are working together right now on blind-prediction tests, in which we both compare the two approaches and try to combine them. The notion of a consensus score is very common in bioinformatics; it could be either regression to the mean or removing outliers in any particular score, etc. Very often a "jury" or a "consensus" idea has more validity or more predictive power than any method alone. So a combination, a consensus score between physics-based ideas and ROSETTA-based ideas is increasingly likely. Frankly, as the empirical programs move

forward, they become more detailed and more understandable, and I think in a few years it will be much more common to see combinations of these things.

One of the things that is happening in code development is that molecular dynamic simulation codes are increasingly agnostic to what sort of force field they are being applied with. In previous times if you wanted to run the CHARMM force field, you had to run the CHARMM program, and if you wanted to use the AMBER force field, you needed to use the AMBER program and so on. Some of that is still true and ROSETTA has special features that other programs have not implemented. But increasingly those distinctions are going away and people are working very hard to make program environments (and simulation and docking programs and codes) much more general and agnostic to what sorts of force fields are being used. I fully expect to see that there will be more detailed comparisons between scoring functions and physics-based energies and more combinations of the two – to see what works.

There is a purist in part of me that says, "Well, we really do not want to fudge too much." Even though there are empirical parameters in force fields, we think we understand where those came from and understand what is going on. That is less true for a scoring function or knowledge-based function. For example, it may be trying to capture some cooperative effect that is really difficult to capture, but not really explaining it – even though it might numerically get a good score for a good system. In the long run, I think the physics-based ideas are going to win out because they will be the ones that are least problem-specific and the most generalizable from one place to another. But it is going to be a long time before that nirvana comes and I think we will have lots of combined methods for many years in the future.

**II. Let us assume that someone wants to introduce a (relatively minor) correction into a successful force field. In doing this, he faces two massive challenges: (i) strictly speaking, the entire force field needs to be re-parameterized and (ii) the new force field needs to be thoroughly validated and then accepted by the community. How to address this problem?**

The general problem of how do you evolve force fields from what you have now to something better is a tough question. Historically it has been a tougher question than it is now because we did not have the computer power to really test potential new force fields very extensively before they were published. So one might say we had a certain generation of the AMBER force field, but it might take years and many people looking at that before the real strengths and weaknesses of that became apparent. Therefore if you tried to make a change or a tweak to a force field it could be a pretty dangerous thing to do. I think both programming to automate things, and the availability of computer power to do more tests on new force fields have made a tremendous difference in the past few years. We and several other groups now have really automated the process.

If I change one parameter, I really need to re-parameterize lots of other things. Say, if I change the size of an oxygen atom, that affects the torsions that are involved with that oxygen atom. I need to re-parameterize against, say, my known database of things that I want to work well – and then I need to test it on lots of real proteins and real systems. But that is not feasible to do in a fairly short turnaround time and in an un-automated way. So there are a number of protocols now that say: if you have even a single new parameter, computer can turn the crank to not only re-parameterize everything else to be consistent with whatever you think you want a force field to do, but also to test that out on lots of new systems (before you even take it further). The requirements that journals or reviewers have now to accept the description of a modified force field are much more rigorous than they used to be. You need to show that the helical

stability is okay, β sheets are okay, turns are okay, folded proteins stay folded, even unfolding temperatures are (roughly) right. These are still relatively small proteins that one is looking at because exploring the phase space of larger proteins is still outside the limit of computational power. But one can have a lot more confidence in a new force field just because computers are so much more powerful and the testing environment has advanced far enough. You can know that you are not doing something really wrong to a protein with a particular change because before you use it in earnest on new problems you will have tested it increasingly on a wide variety of existing examples. Now that may oversimplify the problem – the problem has not completely gone away. But I think that is one of the places where we have made real advances in the past few years in force field development.

**III. Could you give an example of work where the problem of evolving a force field has been approached in a proper and systematic fashion?**

Of course, nothing is truly a sufficiently big sample to be sure that you are not doing something wrong. But let me mention a couple of things. One is the work of Lee-Ping Wang who is now at the University of California at Davis (and who worked at Stanford before this) on what he calls the ForceBalance approach. Any individual step of it is not completely new, but the notion of automating all parts of that procedure is new. Say, when I make even a relatively small change to the parameters, I really go back to ground zero and re-parameterize the entire force field against my database (we do create a real database of things that we want to reproduce). David Cerutti in my group has done a similar thing with what we call the ff14ipq intrinsic polarization potential. Part of that has to do with polarization, which we will come to later, but a big part of that has to do with really good work that Dave did in automating the whole procedure of regenerating the force field. Other people do this as well. I think that to a greater or lesser extent, there are really quite a number of examples of that. Having said that, this is still an expensive process; not every change deserves that sort of process. So we are kind of in the middle of things. The fact that all the parameters are coupled to each other in odd ways, in unpredictable ways is not a completely solved problem, but the community has gone a long way to solving that problem. The main thing is just the increased computer power that is available now to test things out.

**IV. What is the current status of polarizable force fields – and the future outlook?**

It is one of those things that always seems like it is almost ready to happen, but it has been "almost ready to happen" for the last twenty years and has not really come to the fore. There are certainly very good arguments that one of the big approximations made in common protein force fields has to do with charges. In the so-called fixed charge force fields, the atomic charges are independent of their environment and represent some kind of averages. The real charges are not fixed – they actually fluctuate as the environment changes. But you choose some average polarization, e.g. of a carbonyl bond, and keep that fixed in most current force field practices. That is what is done, and you hope that average behavior comes out. And certainly one of the things that are missing from current force fields is the mechanism allowing the electrostatic behavior of side chains or backbone to change as a function of their environment. But that is not the only thing that is wrong with modern force fields, and Jay Ponder is one of the people who really emphasize that. As he tries to make his AMOEBA force field, it does add polarizability but there are lots of other things: static multipolar moments, and careful description of non-bonded interactions, and other things that are at fault with current force fields. It may have been too simplistic to think that if we just add polarizability, we are going to make dramatic advances.

The other thing that comes in – and it is an interesting question, there is kind of an Occam's razor thing here – if you do a sample calculation or set of calculations with polarizability included, do you really expect it to be better? Are you sure that you could not have gotten the same improvement in physical fidelity by just making better parameters in a fixed charge force field model? My personal work has been more of the latter: are we sure that we are at the ending point (or near the ending point) of making fixed charged force fields as good as they can be? It is tough to do polarizable force fields. There has been a lot of progress recently. The time penalty that you pay for simulating with polarizability is less than it used to be, and the number of people working in this area is greater than it used to be. So we may finally get to the stage where they break through and become much more common. But I think it is going to be another 5 or 10 years at least before that happens. I do not think we have reached the end of the ability of fixed charge force fields to represent faithfully a lot of things that we want to know about what proteins do. I also kind of agree with Jay Ponder's hypothesis that I mentioned before that we really need to add a lot of things in addition to polarizability to truly have the next generation of force fields. And that is going to be tough, but I am looking forward to reading about it and maybe contributing to it myself.

**V. What are the areas where traditional force fields can be improved?**

The opportunities for improvements in current force fields are not so much in terms of which part of the force field is wrong and needs to be fixed, but which sorts of interactions and balances of interactions do we need to fine-tune? As an example, in the past few years it has become quite obvious that unstructured proteins are not described very well with current force fields (some people have known this for a longer time, but the field as a whole has become much more cognizant of this fact only recently). So the so-called intrinsically disordered proteins, which are unstructured even in their native conditions where most proteins would fold, or proteins that become unstructured either by denaturant or high temperature, are not described very well.

Most people (including me) think that the culprit is the balance between protein-protein interactions and protein-water interactions. When you break a piece of secondary structure to partially unfold a protein, you are actually swapping hydrogen bonds. You are exchanging a protein-protein hydrogen bond for a protein-water hydrogen bond. That equilibrium is probably easily changeable by small alterations in non-bonded interactions. So relative strength of hydrogen bonds, protein-protein vs. protein-water, is probably one of the things that one needs to work on. The experience in my group has been that if you start to make models that look better for intrinsically disordered proteins, then suddenly, if you try to go back and use those for folded proteins, the folded proteins start to unfold. You tend to overshoot in estimating how happy protein segments are in interaction with water (this is what is underestimated in the standard force fields). So one hopes that there is a magic balance, a fine balance that can more correctly describe unfolding and folding. And it is not just total protein unfolding vs. folding, because even proteins that you think of as folded undergo local unfolding, e.g. loops move around. The same physics is involved in fairly local events of breaking and making hydrogen bonds as is involved in the global aspect of protein folding. Getting that balance to work out is a challenge – maybe we cannot do that with current force fields and functional models, but I am not convinced of that yet. I think there is still room for a lot of clever ideas and testing in the space of current protein force fields.

Nucleic acids have their own set of problems because electrostatics and ion atmospheres are much more important there. I think, generally speaking, our ability to understand nucleic acid behavior has fallen behind proteins. Then there are carbohydrates and lipids and all kinds of other biochemicals that are

increasingly a part of the simulation armamentarium that each have unique problems. In my mind, the unique problems are not so much that you need different parameters for a lipid than you do for a protein, but you need to think differently about things, i.e. about tests, that you want a potential force field to satisfy. Those are different for membranes than they are for folded proteins, and so that has to come into play as well.

**VI. You talked about the challenge of constructing a force field which would embrace both folded and unfolded proteins. Is this dependent on our ability to capture a small free energy difference between the folded and unfolded species?**

That is a grand challenge. In the long run we hope that we can understand that better than we do now. Personally, I have not felt that trying to look at the general problem of protein stability (i.e. a free energy difference between folded and unfolded configurations in some given physical environment) is a very productive way to go about this. Because it is such a hard thing. Describing the phase space of the unfolded state and the configurational entropy and the solvent entropy and all the other things, as well as just energetic consequences, is very hard. My predilection is more or less to look at what I call "brute-force" simulations. Can you just run a simulation under a given environment (without being clever theoretically the way that many people in the protein-folding field are really clever in understanding what goes on) and say, "I just want the computer to tell me the answer." Well, that is still way out there in terms of computational feasibility for the disordered state of proteins.

As a community, we are almost at the stage where we can reliably and reproducibly model oligopeptides in both folded and unfolded configurations and really sample the configurational space of the unfolded state in a way that, if not fully converged, seems to be functionally converged. One does that with replica-exchange calculations and very large amounts of computer time. Similarly, with nucleic acids in the past couple of years people have been able to convincingly show that they can do reproducible simulations of tetranucleotides, which are unstructured and pretty floppy things. This is only four nucleotides long and yet it is just now the community as a whole has reached the point where we can say, "We can do a simulation on this system and the next lab over can do the same simulation and ought to get the same statistical result." The same can be said of oligopeptide that might be a little bit longer than that because there are fewer rotatable bonds per residue in amino acids than there are in oligonucleotides, so you can do a little bit longer peptide.

Even now these are heroic calculations to sample the phase space of relatively short peptides by brute force methods. Until you can easily do that, I think that using folding free energy as a mean of force field development will not be very attractive to me and, I think, to anybody else. We are not yet there where we can say, "Well, let us tweak the force field and rerun the folding simulation and just find out what happens," because it is still too expensive a calculation to do that routinely. In a way, yes, it is a grand challenge to really understand protein stability, but it is not a grand challenge that, I think, the NIH or the NSF should pour money into right now. There are many intermediate challenges that we have to solve which are hard enough without dealing with this problem. But that is my personal bias and I am happy that other people try other things. There is no reason to always think that you can only do solvable problems – looking at things that seem to be unsolvable is a good way to approach science as well.

**VII. Does this mean that for a while we may need to live with two versions of force fields – one for folded proteins and another one for unfolded proteins?**

I talked earlier about how the predictions for folded proteins for OPLS or CHARMM or AMBER are actually more close to each other than they are different from each other. Of course, that is a subjective judgment, kind of a view. This is not so true if you look at disordered proteins. If you try to estimate computationally, for example, what is the unfolding temperature of a small protein – that can be very different from one force field to another. But I did not mean to imply that I am happy that we can have a disordered protein force field and an ordered protein force field. I think that is not a very good situation. Partially, for the reason that I have mentioned before: you have little elements of local unfolding that go on all the time, even in the conformation that you think of as a folded protein. So I think there really is a sense that we want the behavior of intrinsically disordered proteins – as measured by radius of gyration, by small angle scattering, as measured by NMR propensities (in what you call a disordered protein there is still some fraction of partial secondary structure and that is often measurable experimentally) – we want structural aspects like that to be faithfully reproduced. We have not solved this problem as a community yet, but I think it is a legitimate and relatively short-term goal to say that a single force field ought to be able to do a reasonable job on both folded and unfolded proteins.

Actually getting the free energy difference between two different states, i.e. reliably predicting thermodynamics in the unfolded state, as opposed to qualitatively reproducing the right radius of gyration or the right fraction of helical structure, is the grand challenge that I think is one step too far. I will be long-retired before someone solves that problem. But I think we can make good progress and learn a lot about what is wrong with "folded" force fields by looking at disordered proteins. As I said, in my lab we tried to fix some of the problems that we know about in unstructured proteins and overshot the mark. We ended up over-stabilizing the unfolded part and then when we applied that same force field back to a nicely folded protein under conditions where it should be nicely folded – it was not so nicely folded anymore. But we and other people in the community will probably learn how to handle those problems. Maybe those problems will require new types of potentials as well as new parameters – I do not think the answer to that is known yet.

**VIII. Considering different water models (TIP3P, SPC/E, TIP4P-Ew, TIP5P-E, etc.), how should one approach the choice of water model?**

There are different protein force fields (we have already talked about a few of them) – and then there are lots and lots of different water models. Chemical physicists, as well as biochemists, have been engaged in making force field models for pure water, and simple ionic solutions in water, and so on for a long time. You can think about mixing and matching the various water models with various protein potentials and the question comes up in a practical sense: if you are not just interested in the water, but you want a faithful representation of the protein, what should you choose? And again it does not make a lot of difference.

I think that the really old but still popular water models, such as TIP3P, are kind of past their design date by now. That particular water model [TIP3P], is tremendously well used and studied, but has both diffusional behavior and time-dependent behavior that are quite a ways away from experiment. It was designed back in an era where one did not handle long-range electrostatics with Ewald-based methods in the way that we do now. And it is no cheaper than other models that are more recent (the fact that they are recent is not the point, it is that they address some of these problems). There are three- and four-point water models that capture a lot better the time dependence of water motion. So if you are interested in motional behavior, these newer models are much better.

Still it is not so much the water model per se that is of interest, but what is the protein-water interaction energy under certain circumstances. And not only that absolute number, but the relative number of that interaction energy vs. a protein-protein interaction energy that comes about. Having said that, we are very happy with TIP4P type models. Even three-point models that are usually based upon the SPC/E type model often seem to give really good results. Or results that you cannot tell are bad in any obvious way if you apply them to a biomolecule. I think the challenge for the chemical physicists or the people that are interested in water structure is to do a more thorough job of seeing what happens with non-polar solutes dissolved in these water models and how you get that to work out. How you find the right experiments to compare to and how you really optimize that potential? So it is not as important just to say, "I have a better pure water potential," as it is to say, "I have developed better ways of optimizing the protein-water interactions vs. the protein-protein interactions." We have thoughts along those lines, but we do not have any magic bullet that helps one do that sort of thing.

**IX. Oftentimes people would argue that e.g. CHARMM has been developed with TIP3P water and hence we should stick with it. Is this a valid argument – or is it ok to "mix and match"?**

This is a really contentious issue and different people have different opinions on this. There are good arguments on both sides. You can say, "Well, this potential has developed in the context of a TIP3P model." AMBER, in fact, was developed in the context of that model (a little less explicitly TIP3P-centric than was CHARMM, but the same ideas were there). And so yes, you can say, "Well, maybe we should stick with that." On the other hand, as we test things more and more, the tests that we have done lead us to say, "Does it really make a big difference if you swap out one water model for a different water model in most practical simulations?" It seems not to make a big difference except for time scales on things like TIP3P and so on.

But I tend to be agnostic. In general, you cannot just say one can willy-nilly mix and match protein potentials. It kind of goes back to the question that we discussed before: if you change your water potential, you should go back to ground zero and change everything else. And, in fact, these machines that develop new potentials that I was talking about actually do that. It is not just that if the protein potential changes then you turn the crank and rerun all the parameterization. If the water model changes, you also redo everything. So we may eventually have separate potentials for different water models. And this first came about with ion-water interactions because those are clearly quite dependent upon the water model. But proteins have charged side chains. And interactions between non-polar groups and water are as sensitive as interactions between polar groups and water. So the field is increasingly going to say, "Okay, if I want to have a different water model because it has better water properties, then I will redo everything."

**X. There are a number of enhanced sampling schemes that can be used in MD simulations of biomolecules, e.g. REMD, metadynamics, etc. When are these schemes useful – and what are their limitations?**

Yes, there are a variety of advanced sampling methods. I talked earlier about just wanting to do a brute-force simulation and maybe that is wrong because I should not give up on schemes that improve sampling. I think of advanced sampling (or enhanced sampling) schemes as being in two categories. One is where you know the collective variables that you want to explore. You want to look at the opening or closing of a pocket, you have particular torsion angles, or you have a reaction coordinate that you want to explore, or something along those lines. So you usually have a small number of important variables: collective variables,

progress variables, whatever you want to call them. And there are a variety of techniques – umbrella sampling was one of the earliest ones, but now there are other adaptive schemes, various other schemes that will accelerate sampling in those particular directions. And those schemes are really important and are increasingly widely used.

The second broad category is where you do not know what you do not know: you do not know what sorts of transitions you are not sampling. Generally, if there is a barrier that is higher than 5 or 6 kilocalories per mole, then you will not sample it in a given time frame with unenhanced sampling. So there are a variety of schemes that do not require you to know in advance or to specify in advance what sorts of coordinates you want to enhance sampling on, but improve the sampling broadly. Temperature replica exchange is one of those. Temperature is a very global variable – if you increase the temperature, you increase the rate of sampling of all kinds of interactions. So why do not people always use that? I think there are two main reasons. First, you usually pay a big computational price to do that. You may only be interested in the behavior of your protein at 300 Kelvin, but you are doing simulations at many temperatures. In the end, you are throwing away all the results of the other temperatures. So you are paying a big computational price in order to do that sort of thing. And the second, more practical, reason is that until recently it has been easiest to run replica exchange simulations in a synchronous fashion where all the different temperatures or different replicas are running in an identical time and they can swap information back and forth (it is still mostly true, but not completely true). That kind of implies that you are working on a supercomputer, you are working on a big cluster where you have access to large numbers of cores at the same time. And that is sometimes the case: there are supercomputer centers and clusters that are available increasingly and individual institutions can handle that, but that still is a bigger commitment of computer time than people are often willing to make.

I will just add the platitude here: the harder you accelerate something, the harder it is to get back, to correct that accelerated simulation back to the original thermodynamic simulation that you want. Everybody knows this. The more you go away from Boltzmann sampling in your raw simulation, the more trouble you have when you correct it back to thermal equilibrium and the greater is the sensitivity to errors. And the tradeoffs there are a really interesting area of global exploration that the entire community is going through right now.

I have personally been a little more interested in exploring things where you do not know what variables you want to explore. As an example, if you are doing crystal simulations you have these channels of water, relatively small channels of water in a crystal – they are in between proteins. The water molecules in there can get trapped into bad configurations and it can be hard for them to relax into a really good configuration from a bad starting point. Harder than in solution where you just have a single protein and the water bath around it actually diffuses relatively quickly. Diffusion of water molecules inside a protein channel in a crystal is much slower. But we do not know the collective variable to say, "I want to enhance water motion there." So we have to look at different types of things.

**XI. What is the value of a good force field in solving an NMR structure of a biomolecule – or an x ray structure of a biomolecule?**

One of the real things that force fields are used for are as adjuncts to refinement either in NMR or in crystallography. The value of doing that better *vs.* doing that worse depends on what you are going to do with the structure. Most NMR structures and crystal structures are actually used in a very qualitative

fashion. People visualize them, they turn them around, they look at things without thinking in great detail about their origins – or even knowing which force field might have been involved in their calculation. If someone like me is going to use one of those structures to do things, the very first thing I do is equilibrate it in my favorite force field anyway. Then I have accommodated into my calculation whatever imperfections there might have been in how the structure was refined – and they are gone.

Having said that, in the NMR field and the crystallographic field a number of measures of quality really do get better if you use modern force fields and explicit solvent and things like that in the refinement process. You can tell without even knowing, without looking at the paper where it came from, an NMR structure that was refined in explicit water with a certain force field. By local geometric standards, lack of bad contacts, favorable regions of side-chain torsions and so on, the structures that are done in more modern (and I think generally better) force fields, and the ones that are refined in explicit solvent rather than implicit solvent, turn out to be statistically better. We are finding this in crystals as well. You can actually dramatically improve your score on a validation test, which is primarily looking to see if you have any bad contacts or unfavorable local regions of space, by merging crystallographic refinement with force fields in a way that people do not commonly do now. Now, whether you have really learned anything more about the protein than you knew before without that extra validation checkmark (confirming that it looks good locally) is an interesting question. For most purposes, you probably have not learned anything biochemically interesting – you have just adjusted average positions of atoms by a few tenths of an angstrom here and there, so that statistically it looks better. And as I said, if some person like me is going to really use that structure anyway, the first thing they are going to do is equilibrate it in their favorite force field and adjust out those things anyway.

But that may be understating the benefits to be gained by having better local structure. The better your model is (in either NMR or crystallography, let us say crystallography), the better your electron density map is – because you are taking the phases that go into the map from the model. And the clearer your map is, the more easily you are able to see alternate side chain conformations or other things that you might not have been able to see on a lower quality map. So doing these things to make structural models better – even though it does not sound like it is biochemically important – can have payoffs in making the maps better and easier to interpret. I think we still have to see how useful this will really be.

**XII. How to model/parameterize the transition metal ions in MD simulations – what is road forward?**

There is a lot of work in that and I think of this in two different ways. One is when you have metal ions, whose configurations do not change very much as a function of time. Inorganic iron-sulfur clusters look like iron-sulfur clusters in a protein. Zinc ions in a zinc finger structure do not change their ligands or swap out with waters or so on. Their first coordination shell is relatively rigid and you do not have the transition metal problem with Jahn-Teller distortions or interactions of electronic structure with nuclear structure. Those kinds of things most conventional force field models potentially ought to be able to handle. And there are increasingly automated and reliable and testable procedures to do that.

If you have cases where the first shell near the transition metal (or near any metal) is variable, e.g. a magnesium ion is swapping out and exchanging a water for an oxygen on a protein molecule, then that is much harder to model and probably cannot be modeled without going beyond the functional forms that we have now. I think there is a subset of metal ion problems that we are equipped to handle and another

subset of problems (and I do not know which one is bigger) that requires a lot more thinking or maybe QM/MM or other types of procedures that go well beyond what I am calling conventional force fields.

**XIII. Is it possible to integrate certain chemical reactions into AMBER (cf. reactive force fields)?**

That is certainly possible. AMBER already has (and other programs do as well) things like empirical valence bond models, which enable you to parameterize reactive conditions in a pretty intuitive way. You can parameterize those against experimental data, against quantum chemical data and so on – in much the same way that you parameterize non-reactive sorts of things. That is easiest done if there is just a single bond-breaking / bond-making step; if you have lots of steps where you are making and breaking chemical bonds then that procedure becomes much more difficult. But once you have parameterized your reactive system (and it may take a long time to do expensive quantum chemical calculations and set up the parameterization) then the simulation is fast.

Making a part of the system be intrinsically quantum-mechanical is another approach. We have played some with that. That still is very expensive, but it is also really illuminating when thinking about chemical reactions. My career has been such that I have worried a lot less than maybe I should have about chemical mechanisms, making and breaking chemical bonds. So I am probably not the best person to really give a good overview of that area. But the initial parameterization approach – where you say, "Okay, I am going to create a force field that looks very much like a fast force field, but allows bonds to be made and broken and I am going to see what that implies" – is a very attractive way. It is maybe kind of underappreciated relative to the more intrinsically obvious choice, "I will make this part of a system quantum-mechanical and quantum mechanics can do everything." Well, quantum mechanics can in principle do everything, but mostly it cannot do the things that you really want it to do in the time that you have computationally to apply it.

**XIV. Is it true that MD reproduces motional amplitudes in biomolecules more accurately than the motional time scales?**

I am not sure that it is true. I think in the past people might have been left with this impression. We have mentioned before the problem of getting the diffusional characteristics of water, which is the solvent of interest for proteins, to be correct. And it is a challenge to say, "Get time scales of rotational tumbling and overall motion of proteins in a calculation to match the experiment." But I do not think that is an unbridgeable challenge. I think we have a lot more reliable NMR evidence experimentally about motional amplitudes than we do about time scales, which are harder to extract from experiment. The time scales are more involved and they are model-dependent in some cases and so on. So we know more about amplitudes than we do about time scales. But I am not at all convinced that there is any fundamental difference – and there should not be. If we are getting the time scale wrong because the heights of barriers between different states are wrong – then we are going to get amplitudes wrong as well.

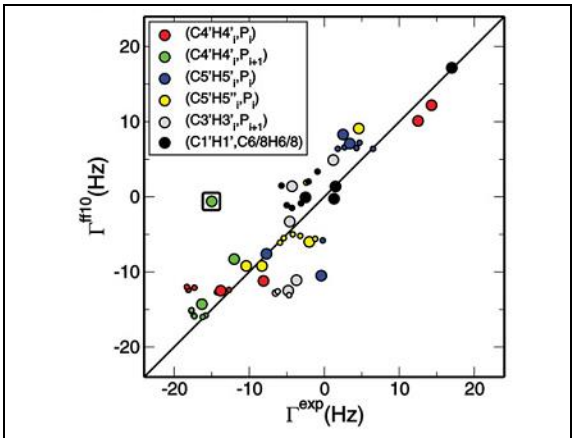


Figure 1. Experimental vs. simulated cross-correlated relaxation rates in the UUCG tetraloop RNA hairpin [Giambaşu, York & Case *RNA* **2015**, *21*, 963].

Otherwise if you are looking at very small amplitude motions that do not involve barriers, you should get the amplitudes right – but then you should get the time scales of those motions right as well. If we look at very fast picosecond-type motions, there is every reason to think that the simulations are getting that sort of dynamics correctly. So there should not be a distinction there and I do not think in practice there really is a distinction – beyond the historical fact that many simulations of proteins in water used water models that are not faithful enough about their viscosity to correctly predict protein motional time scales. But I think that is a solvable problem.

**XV. What are the NMR observables that are most valuable as benchmarks for MD simulations?**

"Which NMR observables are valuable for benchmarking and development of force fields and improving our simulations?" is a different question than "Which types of NMR observables are the most useful for helping us understand the biochemistry of a particular protein system?" Getting nuclear Overhauser enhancements and the distance-type restraints or residual dipolar couplings that give you information about longer-range structure can be extremely important. But as your question implies, they are less obviously important for benchmarking of MD simulations or the fine-tuning of force fields. Residual dipolar couplings are generally plagued by the fact that we do not have a good physical model for what is causing the anisotropy in the motion. That is not always true, e.g. it is not true for residual dipolar couplings (or chemical shift anisotropies) that are caused by magnetic anisotropies of paramagnetic iron ions, where we understand the physics. But in most cases where we have an oriented medium, we do not have a good model for what is actually causing the orientation. So that makes residual dipolar couplings less valuable.

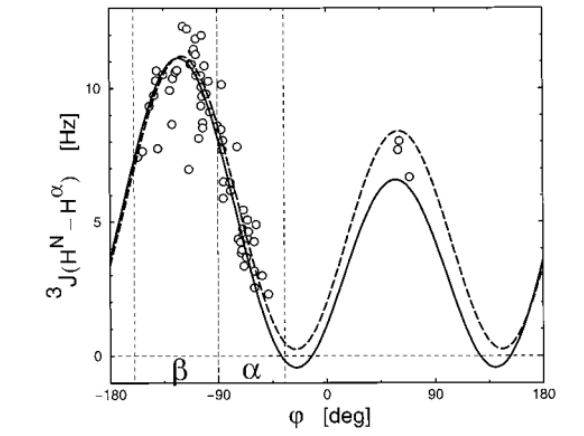


Figure 2. DFT results (open circles) of the $^3J_{HN,H\alpha}$ couplings using MD simulations of ubiquitin; superimposed are the computed Karplus curves [Case, Scheurer & Brüschweiler *J. Am. Chem. Soc.* **2000**, *122*, 10390].

Certainly, *J*-couplings are a nice feature in MD models and are widely used. The measured *J*-couplings on oligopeptides are increasingly demanded by reviewers as a test of the feasibility of a proposed new force field. And they give you good information. Chemical shifts are another one, even though we cannot calculate chemical shifts from structure as well as we can calculate spin-spin couplings. But still they provide valuable information about averages and conformational averaging. As I mentioned earlier, i.e. about ten years ago, comparisons of relaxation behavior (especially the fast picosecond-nanosecond type of relaxation behavior and the corresponding motional amplitudes that can be measured by NMR) became a really important piece of evidence that drove the entire field toward modifying force fields to make proteins more rigid because NMR has been saying repeatedly that motion of atoms about their average coordinates was smaller than we were seeing in solution MD simulations. That continues to be a valuable source of information. And although it has not been used as much, overall tumbling of proteins is a valuable piece of information as well. It reports upon fluctuations about the average structure not quite so directly as the fast motions do, but reports upon the strength of the protein-water interactions on average. Gaining more confidence in our ability to describe overall tumbling as well as local motions

with MD simulations is a good spot. So those are the measurables from the NMR side that have been the most useful in my field and I hope will remain useful in the future.

Once again, the connection between the microscopic behavior and the observables discussed in the previous paragraph is pretty straightforward. It may be hard to find a force field that actually does a good job, but you kind of know what behavior you want to see in the simulation. What you do not know is how to get there from where you are. Your bad force field does not tell you how to correct it to make a good force field. But it is a fairly straightforward connection between theory and experiment.

**XVI. Calculation of chemical shifts is probably one of the most obvious applications of QM/MM methodology. Why does it lag behind the empirical methods?**

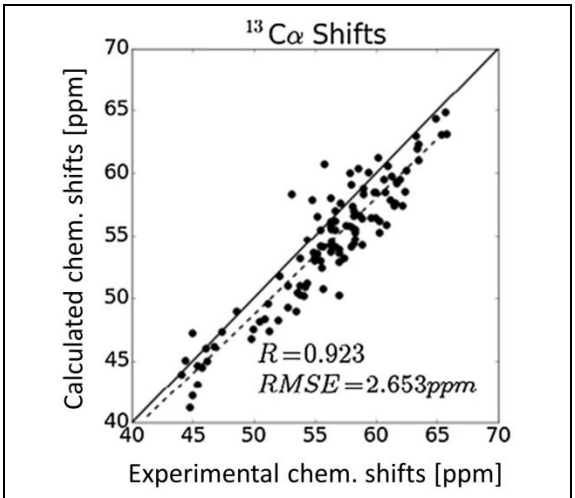


Figure 3. Experimental vs. simulated $^{13}C^{\alpha}$ chemical shifts in the RNA-binding domain of the influenza virus protein NS1; the calculations were conducted using the automated fragmentation quantum mechanics/molecular mechanics approach [Swalis, Zhu, He & Case *J. Biomol. NMR* **2015**, *63*, 125].

The chemical shift, as many people know (but maybe not everyone), is an electronic structure property. It is like a coupling constant in that sense: they are both electronic structure properties. In principle, quantum mechanical methods should be able to calculate chemical shifts; you are actually looking at the response of the electronic cloud to an applied external magnetic field and we know, in principle, how to do that. If you look at small-molecule crystals where you actually know very accurately both the geometry of the molecule and the environment that it is in, i.e. its surroundings, then quantum calculations can be exceedingly accurate under a wide variety of circumstances. Although we are not doing exact quantum calculations by any means, the primary problem of applying quantum chemistry to chemical shifts in biomolecules is not that we do not know how to do good quantum chemistry or we do not know how to fragment things into small pieces, but that we do not know how to describe the solution environment. So even if the molecule that you are looking at, i.e. the protein molecule is relatively rigid, it is actually fluctuating and the water molecules around it are really fluctuating a lot. You have to do a lot of averaging over water positions to get a good answer there. I think learning how to do that has been a tough problem. The empirical methods clearly are fitting to an experiment that already has water averaging done for them. So they are going to better at that.

Going back to force fields and simple (or even four-point) water models, keep in mind that water-protein interactions were really parameterized to get the energetics of protein-water hydrogen bonds correct. Much less has been done about getting the average geometry of that protein-water hydrogen bond correct. The angular dependence, the actual location of the water molecules in the vicinity of a peptide group in MD simulations is probably not all that accurate. So if you take snapshots from MD trajectory and say, "Well, I will do a quantum calculation on that," then that inaccuracy is folded into your calculations. I think the difficulty of averaging over a liquid-state ensemble (i.e. generating the liquid-state ensemble and then averaging over it) is really the most important reason why QM/MM-type methods are not yet competitive with empirical methods.

Having said that, the empirical models, let me say, are only as good as the data that go into them. We recently did collaboration with Jean Baum in this department on collagen peptides which have very unusual configurations. If you just apply standard empirical models of chemical shifts to collagen (where the empirical models are based upon fitting the data from globular proteins) you actually get quite bad answers. And we were able to do some MD sampling and get pretty good answers from quantum mechanical perspective. An empirical model is an empirical model – if you are in an unusual environment that is outside its region of fitting then you may do poorly. So for the immediate future, I think the sweet spot for applying quantum approaches to chemical shift analysis is in cases where we do not have a lot of good empirical data or we have unusual configurations (e.g. a protonated side chain rather than a deprotonated side chain, or nucleic acids because we do not have a big database of chemical shifts of nucleic acids the way we have a big database for folded proteins). There will be the immediate sweet spot where these techniques can pay dividends.

**XVII. You have mentioned that MD does not always model protein hydration very accurately. How successful is MD in recovering the positions of crystallographic water molecules?**

The question of how to handle solvent in crystallographic refinement is an important one, and I am not the first person to say that. Crystallographers model water in practice by looking for "ordered water molecules", i.e. the places where there is high electron density so that they can put a water model there. Then they iteratively do that and find the next piece of high density and so on – until they either run out of time, or patience, or good features in their electron density map that can be improved by adding discrete water molecules. And then they model the rest as the bulk solvent with various levels of approximation and sophistication. You may get the feeling that you have this database of water positions from crystal structures (and in some cases those are completely reasonable water positions and good hydrogen bonds). But I am a little unhappy with that because of the very modeling procedure that crystallographers are using (which admittedly has occupational parameters that they can adjust, B-factors that they can adjust for motion, and so on). Yet it is really not a physically intuitive way of modeling waters, it is merely a convenient computational way of doing that.

We know from simulations (even though the simulations are far from perfect) that water molecules move around. The reason you have a high electron density in a solvent region is often not because there is a single water molecule that is staying there. In fact, we know from NMR relaxation that very few water molecules remain bound to a protein for more than a few nanoseconds – there are almost none. I am not sure that the data saying, "Well, we know the positions of a bunch of waters in high-resolution crystal structures," is the best gauge to check if simulation is doing the right thing. As I said earlier, we are doing MD simulations on crystals, computing the average electron density from the simulation and then calculating the scattering and comparing that to experiment. The overall answer that we are going to come up with is not yet known, but you can describe pretty well with MD what the solvent is doing – particularly in the simple case if you keep the protein fixed as average x-ray configuration and just allow the water to move around.

There is probably some combination of high-density positions that we are actually missing in an MD simulation (that should be added). In molecular dynamics we may have a more realistic description of the relatively subtle ups and downs of electron density away from strong hydrogen bonding sites, but we might not be doing the right thing at the strong hydrogen bonding sites. This is a good question for the future and I think there are a lot of opportunities to be pursued here. But I kind of argue that one should not take the

crystallographic models of what waters are doing without taking a grain of salt along in the same way. These data are largely dependent upon the type of protocol that the crystallographers have traditionally chosen to model the scattering that is going on in their crystals.

**XVIII. What is more valuable for predicting chemical shifts in a globular protein – ultra-high-resolution x-ray structure or state-of-the-art MD trajectory?**

The short answer to that right now: let us better have a good structure. Either with the empirical models or with quantum mechanical calculations, it is really clear – the structure is important. We have the interesting situation when, say, you are doing empirical chemical shift calculation: it is generally better to use a crystal structure for the calculation than to use an NMR structure for the calculation, even though the crystal structure is in a crystalline environment and the NMR structure is in a solution environment (in fact, it is from the same sample they used to get the chemical shifts from). The reason is that the local precision of the NMR structure is likely to be lower because NMR does not have as much data about local geometries as crystallography does. And MD simulations deviate from the correct average structure as well. So the utility of knowing good crystal structure will remain high for a quite a while yet. It is going to be some time before any MD simulation comes close to matching the local geometries (chemical shifts are primarily a local property, so that is the reason why local geometry comes in). Right now you want a good crystal structure if you want to predict chemical shifts.

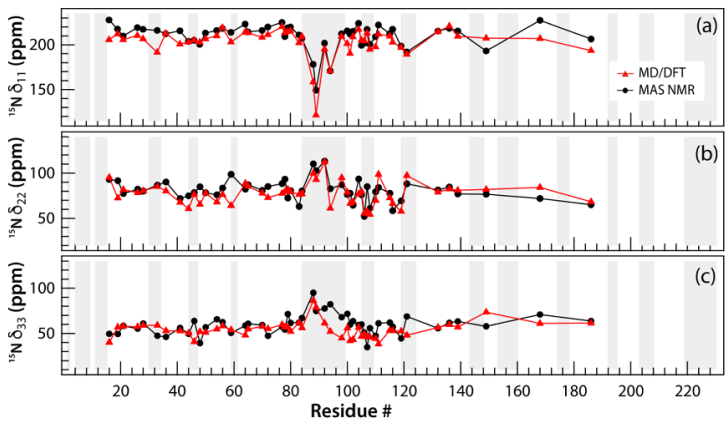


Figure 4. Backbone $^{15}N$ principal components of the CSA tensors in tubular assemblies of HIV-1 capsid protein, plotted as a function of the residue number. The experimental values obtained from 3D solid-state NMR experiment are shown in black; the values calculated by MD/DFT are in red [Zhang et al. *J. Am. Chem. Soc.* **2016**, *138*, 14066].

**XIX. Could you please describe one or two specific examples illustrating predictive power of MD methods in the context of biomolecular structure/function (i.e. the situations where MD does not simply reproduce or rationalize the experimental findings, but makes a major prediction that is subsequently verified by experiment)?**

What is the predictive power of molecular dynamic simulations in biomolecules? One of the things I always tell the students who are thinking about coming to work in my lab is that they have to be comfortable with the fact that, unlike with many experiments, you cannot just automatically trust a simulation that you do. There are so many possible things that proteins or nucleic acids can do that it is actually unlikely that our simulations will capture such a large phase space of possible outcomes of motions – we certainly do not capture all those correctly. To be in this field you have to be willing to constantly be thinking about going back to experiment. But that can create kind of a trap: you have already done the experiment and now you just spend your time showing what a good and high-powered computer person you are to be able to get

the same answer that you already knew. So it is important to think about making predictions, and as the field progresses that happens a lot.

I will bring to the fore a couple of examples where I think predictive molecular dynamics simulations of proteins have been very important. One is in the area of designing small-molecule inhibitors to protein pockets. If you look at a crystal structure you can often see places where small-molecule inhibitors could bind; you can use docking procedures and other computational procedures to try to figure out which molecules might tightly bind and be an inhibitor for a given enzyme active site. One of the things that have come to the fore in the past decade or so is the recognition that if you just use a crystal structure to do that you may be missing some important information. When people began running molecular dynamic simulations – let us say, modeling the apo protein with no ligand bound – and just looking at the shape of the binding pocket as the simulation progressed, they found that the shape was not constant, it would fluctuate. In particular, there would be cases where new parts of the pocket might open up or close down, and you would have a very different shape of the pocket to try to design your inhibitor into compared to what you had just by looking at a static crystal structure. So it has become rather common now to dock against multiple protein conformations, and usually those protein conformations are generated from predictive MD simulations. There are a number of examples of pretty important successes in seeing new shapes of pockets – or even new binding pockets – and designing molecules into that.

The other example that I will just cite (of many that I could) is in the protein design and enzyme design field. We are increasingly getting closer to what you call "protein engineering". You can have a design goal related to enzyme activity, or binding affinity, or something like that. It is the opposite problem of what I was just talking about – rather than trying to find a small molecule that binds to a given protein, you are turning the problem around and saying, "I have an epitope, e.g. a small-molecule toxin, and I want to design a protein to bind the small molecule." Almost all of that is now done first in silico and then tested in experiment – partially because experiments, especially large-scale experiments, are more difficult to do with proteins than with small molecules.

There are also blind tests of protein folding, which just say, "Here is the amino-acid sequence, what does it fold into?" You just want to find the qualitative fold that the single-domain protein folds into. This is a problem that is not completely solved, but is very largely solved. We can predict the general nature of the fold and even the particular details of the fold of the protein from its amino-acid sequence. Going further, we may not only want the general fold to be correct, but we may want to have the particular side chains oriented in the correct fashion around certain substrate (such that they could do chemistry). The antibodies do this naturally in the immune system – the new proteins are evolved to bind to a foreign antigen. Can humans similarly design new proteins? This is a predictive case, this is really a case where you are not just saying, "I know the answer in advance" – you do not know the answer in advance and you are trying to become a protein engineer. And the field is such that most things are really rationally designed with imperfect, but still predictive calculations (although there are certain alternatives, e.g. you could use random mutagenesis and hope to find something that could still be competitive).

Maybe those are the cases where the bar is too low. The more accuracy you need, the higher the quality of your prediction, the smaller the change that you want to be able to discriminate against – the harder it is to be predictive. But that is kind of true in almost every modeling field. The original question was: what is the predictive power of MD? But that is not the only question worth asking, because the interpretive power of molecular dynamics is at least as strong as the predictive power. So people who look

at movies and drug companies that spend their time looking at three-dimensional structures (whether generated by crystallography or by molecular dynamics or by ohter methods) get an intuitive feel about what those molecules are doing. This is very important in their understanding of protein function (even though they did not directly predict a number which would say that the binding affinity is 7 kilocalories per mole plus or minus 0.1). So the interpretive power of understanding things in three-dimensional space, which is supported by experiment, but also in a very large extent supported by simulation, generates ideas in peoples' brains that help explain function.

**XX. You have talked about protein-ligand docking. Docking programs are usually developed in conjunction with certain specific force fields. Is it possible to use a docking program with another force field, different from its "native" force field?**

Suppose I have a docking program that was designed in conjunction with a particular molecular mechanics force field, whether it was HADDOCK or DOCK 6 or other programs. Can I use that docking program with some different force field? I think mostly the answer is "yes". The searching for plausible binding configurations, i.e. the so-called conformational search problem, is still largely uncoupled from the conformational scoring problem. Let us say, I am docking a small molecule into a druggable active site. Can I actually find the right configuration (given that there is a right configuration, I just do not know it)? Does the computer ever look for it? If the computer stumbles across it, does it know that it is a good configuration? Or does the computer think that it is just like every other configuration? There have been decades of work on that suggesting that conformational search and conformational scoring are to a large extent uncoupled problems (although not completely uncoupled). So if you have a search algorithm that does a good job, you can then take those same configurations and re-score them with different force fields and different scoring functions – and see what you get. Coming back to the original question, I personally would be happy reading about people combining HADDOCK with whatever force field they wanted or combining DOCK with other force fields. I do not think the connection between the scoring function (or the force field that you use) and the search function is all that strong.

**XXI. How would you compare the field of MD simulations of nucleic acids with the field of protein simulations?**

Certainly a lot of people in the field appreciate that our ability to do simulations of nucleic acids is a lot less far advanced than our ability to do simulations of proteins. A lot of that has to do with electrostatics. Proteins have charged amino acids, but they have positively charged amino acids and negatively charged amino acids. They may have a net charge, but unlike DNA they do not have negative charges uniformly arranged every few angstroms along their backbone. In the case of nucleic acids describing the ion atmosphere around them and the response to salt conditions is hard. The thing that makes it really hard for RNA (and makes it even difficult to say that we are only ten years behind proteins) is the dramatic effect of divalent ions like magnesium. In cells (and in in-vitro experiments as well) the concentration of sodium and potassium chloride is roughly one hundred millimoles or so, but the concentration of magnesium ions is below ten millimoles and sometimes below one millimole (it depends on the environment). And there is a combination of specific ion binding sites and general neutralization aspects for magnesium. The observed result of this complexity is that RNA folding, for example, is tremendously sensitive to magnesium concentration – to an extent that it is even hard to capture RNA folding in in-vitro experiments because you do not know the proper salt concentration to even think about using. There are many of those examples of

the RNA that is unfolded in the absence of magnesium and then suddenly folds up in the presence of even small amounts of magnesium.

Eventually that problem has to be solved and MD simulations, in principle, have to be able to solve that sort of problem. But we are a long way from understanding the principles of that behavior. In magnesium-free solution, we can get closer to understanding at least small pieces of oligonucleotides and their conformational preferences and so on – even as a function of monovalent salt concentration. But divalent ions are the real nasty, thorny problem that has not yet been solved. One of the things that go on there we already briefly mentioned when we were talking about metal ions. Most of the time the magnesium ion will have six waters in its first coordination shell – but not all the time. In fact, some of those waters will swap out and be replaced by nitrogen or oxygen atoms on the nucleotide. At the force field level, getting the equilibrium constant for that sort of "swapping out" of one ligand for another is very hard. We do not even know the right targets very well, i.e. we do not know what our solution simulations ought to be doing in that case.

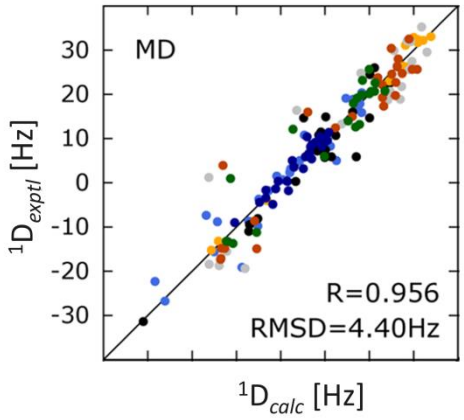


Figure 5. Agreement between measured and MD-generated residual dipolar coupling constants in cUUCGg RNA apical loop; the calculated values have been obtained using ten different elongated kinked RNA constructs (color coded in the plot) that were pre-selected to satisfy the experimental data [Salmon et al. *J. Am. Chem. Soc.* **2015**, *137*, 12954].

In the practical sense that a biochemist would like to see, I am less optimistic about nucleic acids than proteins. In a fundamental sense, we will continue to learn more about what nucleic acids do. It is a fun field to be in and, increasingly, my work has moved from proteins into the nucleic acid field. In the longer run, the extreme sensitivity to divalent ions is a problem. Maybe there will be some great progress on that, which will surprise everybody – but it is a tough thing. So yes, the nucleic acids field is certainly behind the proteins, but it is hard to say by how much. I think the avenues of progress in this area will look very different than they did for the protein field.

**Acknowledgement**

While working on this paper N.R.S. was supported by the RSF grant 15-14-20038.

**Biographies**

**David Case** received his Ph.D. degree in chemical physics at Harvard University working with Dudley Herschbach and Martin Karplus. He has held faculty positions at the University of California, Davis, The Scripps Research Institute and Rutgers University. His research program is centered on MD simulations of biomolecules and he is the leader of the development team for Amber simulation platform. Prof. Case has used extensively the methods of molecular dynamics and quantum chemistry to explore and rationalize various spectroscopic effects, e.g. in the area of vibrational spectroscopy, fluorescence, alpha scattering, and x-ray scattering. However, his most significant interest by far is the NMR spectroscopy. In this field he has made a number of fundamentally important contributions in the areas of spin relaxation, chemical shifts, spin-spin couplings, and structure determination.

**Nikolai Skrynnikov** obtained his Ph.D. degree in physical chemistry from McGill University. He did his postdoctoral work at ETH Zurich and then University of Toronto, before joining the faculty of Purdue University. His main interests lie in the area of NMR methodology, protein biophysics, and MD simulations.