Beyond Recognition

Treating ligand–receptor recognition as a process rather than a structure is a first step toward developing rules of how a combination of hydrogen bonds, salt bridges and hydrophobic contact sites between a ligand and its receptor causes a conformational change in the receptor.

Lukas K. Buehler
is the founder of SciScript Inc., in San Diego, California, USA. He also teaches at the University of San Diego's Extension Bioscience Program. He can be reached at lbuehler@ucsd.edu.

Despite the heavy investment of intellectual capital in drug discovery, predicting physiological effects based upon structural models of drug–receptor complexes remains an elusive goal. Take, for example, the difference between receptor agonist and antagonist when defined as the efficacy of a drug. When ligands bind to their receptors, they do so through specific arrangements of non-covalent bonds. Mechanistically, ligand–receptor recognition is a self-assembly process, no different than protein folding or micelle formation. But what makes a ligand an agonist rather than an antagonist?

I would call this the efficacy problem.

Clearly, we want to set rules of behavior at the molecular level during binding to predict the efficacy of a drug and further our general understanding of molecular interactions in biological systems. Part of the answer to what determines the efficacy of a ligand will come from understanding rules governing the formation, stability and behavior of self-assembly systems such as protein complexes and lipid bilayers. What could be the nature of such rules?

A Numbers Game

In 1943, Erwin Schroedinger wrote that biology is a small-numbers game (1). At the time, he was referring to the still hypothetical structure of a gene, describing it as an asymmetric crystal with a relatively small number of atoms (fewer than one million). As we know, changes affecting just a few dozen atoms (mutations) can have a huge impact on the behavior of biological macromolecules. Schroedinger went on to suggest that we are in need of a new physics to explain how an "incredibly small group of atoms — much too small to display exact statistical laws — do play a dominating role in the very orderly and lawful events within a living organism." Schroedinger asked for a physical and chemical accountability of the spatial and temporal processes occurring within the boundaries of a cell, outlining a physicist's view of life. The "naive" conclusion from statistical physical laws is that for a complex system to be orderly, it had better be large. This simply reflects the observation that addition or removal of a single unit from a system affects the overall system much more if there are few molecules (small "N") than if there are billions of molecules (large "N"). The random distribution away from the mean follows the square root of N and not N itself, making large systems predictable, while probabilities have to be assigned for small ones. This distinction between large numbers and small numbers is eminently important to understand biological systems because they are small-number systems rather than the convenient large-number systems that physicists prefer.

The Efficacy Problem

My take is that rules governing self-assembly processes in biological systems should address the gap between purely macroscopic and microscopic laws. Current approaches to drug design merely blend thermodynamic with structural data. To give an example, the enthalpy–entropy compensation is a well-known macroscopic effect that puts practical limits on the structural modification of high-affinity ligands. When designing novel drugs, we manipulate lead compound structures with the goal of lowering the potential energy of the bound complex. Within the constraint of a fixed area of interaction, adding hydrogen bonds comes at the expense of Van der Waals contact sites. Lowering the number of hydrophobic interac-

The growing list of diseases caused by misfolded proteins is an indicator of the intimate relationship between folding and control of function.

tions reduces the entropy contribution of displaced water molecules. As a result, the affinity changes very little.

While this effect might be seen as a limitation to drug design, it can teach us an important lesson about efficacy that goes beyond molecular recognition. Let us for a moment think of the protein as a "solvent." Instead of treating the ligand–receptor complex as two solutes dimerizing in water, the ligand can be viewed as a solute stabilized at the protein–water interface. The problem now is that we have a system with two solvent phases and one solute type rather than one single phase with two solute species. The ligand can be considered a substance affecting the surface tension of the protein–water interface.

Could such a view — or any other view — of receptor systems address the efficacy problem? Whatever the solution, it should be able to answer several important questions: Why do some ligands cause conformational changes in receptors while others don't? Why do most drugs share common amphipathic features, including aromatic and electronegative functional groups? Why are there only several thousand protein structures, and even fewer motifs, instead of millions or billions, as combinatorial calculations suggest? In other words, what is unique about the structures of proteins found in all forms of life?

Searching for Needles in Sequence Space

Computer-aided methods to identify novel ligand structures from established pharmacophore models are, without a doubt, outstanding scientific advances. Applying them to modeling ligand efficacy (e.g., G-protein activation in the presence of an agonist but not an antagonist) requires the inclusion of a receptor's conformational changes. Doing so simply by calculating possible structural solutions in sequence space is akin to searching for a needle in a haystack.

Computing conformational changes in atomic resolution means juggling enormous groups of solutions that, by necessity, include a small fraction of correct ones. Here, it is important to appreciate the uniqueness of the structures of biologically active proteins with regard to protein "sequence space." The number of protein structures found in nature actually is very small, compared to what the entire sequence space has to offer (this is simply a combinatorial problem). Looking at the number of known genes and related structural motifs of their products is like looking at the universe and realizing the immense emptiness of space, dotted with the occasional life-supporting planet. A focus on high-resolution structures, therefore, appears to be too narrow when it comes to predicting a ligand's efficacy, and the reasons for this are conceptual, not computational.

Instead of computing everything in order to find the exception (e.g., rational drug design is actually limited to "nearby" homologues), wouldn't we be better off finding rules that explain the uniqueness of existing receptor systems (e.g., the process by which proteins — the structures of which currently are deposited in the protein data bank — find their global energy minima within seconds)? This also is known as the protein folding problem — a simple paradigm that states that sequence determines structure. The complete prediction of structure from sequence has yet to be solved satisfactorily, as we still lack a full understanding of the folding process. What we do know is that protein–ligand interactions must follow the same rules that govern protein folding. For example, an agonist is but a small, extraneous ligand affecting folding, and we have good reason to think of agonist binding as the last step in the folding pathway of a native receptor structure. In this view, agonists jolt a receptor into a global energy minimum, while antagonists prevent this from happening. The growing list of diseases caused by misfolded proteins is an indicator of the intimate relationship between folding and control of function.

Molecular Rules of Engagement

So how can we find "rules of engagement," or systems properties that would de-emphasize atomic models and reduce computational demand yet capture the dynamics of molecular interactions? I would like to compare the challenge of this situation to that of an alien learning to play chess by watching a match between famed chess champion Kasparov and his famed computerized counterpart, Deep Blue. Machine and man operated by different mechanisms but on average each would make a number of indistinguishable moves. To begin, our alien would likely catalog the moves, measuring the way each figure was being moved and the frequency of each such move. Equipped with these instructions, this creature could now be ready to

move figures by employing the "machine" approach of calculating every possible combination thereafter. Or, the alien could instead focus on learning tactical rules — how certain series of moves and the grouping of figures correlate with the outcome of the game. Applied to drug design strategies, what we want to accomplish is to distinguish between computing the entire sequence space to select the most stable docking model (Deep Blue's strategy) and the dynamic behavior of molecules (Kasparov's strategy).

We have yet to see by what rules heterogeneous molecular ensembles are governed, an important finding for modeling receptor activation through complex formation. What can we say about their nature? They will likely reflect system's properties that cannot be understood based upon the properties of isolated interactions between individual components. The hydrophobic effect illustrates this point, describing some rules governing the separation of oil from water. Though it is perceived commonly as an act of repulsion between oil and water, hydrophobicity is not a quality possessed by individual oil molecules. Being non-polar, these molecules do not repel water. On the contrary, oil and water molecules are weakly attracted to each other through Van der Waals bonds. It is the preference of water molecules for the bulk phase that leaves the non-polar solutes to interact with themselves. So why the bulk effect? Phase separation, protein folding and micelle formation work much like a bootstrapping mechanism where random association of the non-polar solutes is entropically stabilized by the growing bulk phase of the solvent, making the aggregation of hydrophobic molecules irreversible.

The hydrophobic effect illustrates an important lesson regarding the already mentioned gap between the macroscopic (bulk properties) and the microscopic (molecular interactions). Phase separation and self-assembly processes are systems properties of molecular ensembles brought forth by small differences in the strength of attractive forces among molecules and increased degrees of freedom within the system. Much the same way hydrophobicity refers to solubility properties rather than molecular interactions, efficacy refers to allosteric regulation of ternary complexes involving ligand, receptor and accessory molecules rather than binding.

Most drugs work as antagonists. So why bother with dynamics? First, pharmaceutical companies will have to rely more and more on finding drugs able to rescue rather than inhibit function of proteins. Second, solving the efficacy problem will be important for proteomic efforts to help researchers understand how select combinations of protein interactions determine a cell's state in health and disease. In my last column for *PharmaGenomics* (2), I discussed how molecular structures explain the mechanisms behind already known functions. If function follows structure, why is it so difficult to predict one from the other? For biological macromolecules, function follows changes in structure. These structural changes serve to switch metabolic activity and signaling pathways on and off. The control over whether changes occur, and for how long if they do, is a hallmark of biological macromolecules. These conformational changes are not random and, therefore, they must carry information. Together, they produce activity pattern-defining metabolic and regulatory pathways; in other words, the proper functioning of cells. The molecular biology revolution of the past 50 years promised to predict these patterns from sequence alone, extolling DNA as the "molecule of life." Future developments in proteomics and systems biology will further address this problem.

References

1. E. Schroedinger, *What is Life?* (Cambridge University Press, Cambridge, United Kingdom, 1944, 1967, 1996).
2. L.K. Buehler, *PharmaGenomics* 3(5), 20–21 (2003).PG