New and Notable

Protein Structures from Domain Packing—A Game of Twenty Questions?

Dennis J. Underwood
Molecular Design and Diversity, Merck Research Laboratories,
West Point, Pennsylvania USA

G-protein coupled receptors (GPCRs) are ubiquitous. Their role in communicating extracellular events across the cell membrane is essential and has not been overlooked by those interested in the processes of signal transduction and by those attracted to manipulating signal pathways for disease therapy. The nature of GPCRs makes them suitable targets for drug therapy for a number of reasons. Depending on the signal pathway and the desired physiological response, they can be activated (agonists) or blocked (antagonists). The ligand binding site is accessible from the extracellular environment, thereby not requiring drugs to be able to enter the cell. Receptor subtype homology enables the development of ligands having a balanced effect by interacting equally with each subtype. In addition, receptor subtype heterology holds promise for the ability to exquisitely tune ligand specificity for particular tissues and particular physiological responses. During the last 5 years or so, there have been significant advances in our understanding of the details of the manner in which hormones such as angiotensin II and epinephrine bind to GPCRs. The liaison between molecular biology and medicinal chemistry in this area has given birth to the concept of a ligand binding site, which is conserved among a great number of GPCRs and can be used by antagonists to block the effects of the endogenous agonist. In addition, we are beginning to understand the biochemical processes involved with agonist activation, signal transduction, constitutive activity, and so forth. However, there is a paucity of structural information on this important class of integral membrane receptors. Our current understanding in the areas of ligand binding and signal transduction in GPCRs comes from the familiar scientific process of deductive reasoning. This understanding has been developed without the benefit of atomic resolution detail obtained from x-ray diffraction or NMR studies. Recent work by Herzyk and Hubbard entitled "An Automated Method for Modeling Seven-Helix Transmembrane Receptors from Experimental Data" published in this issue of the *Biophysical Journal*, makes significant advances in building three-dimensional models of GPCRs that are consistent with the results of seemingly disparate experiments, a process not unlike the popular guessing game "Twenty Questions" (von Heijne, 1995).

Scientific opinion from some quarters maintains that the biological mysteries of ligand-receptor interaction and the chemical and conformational response of the receptor to these precise but subtle events must await the determination of a high resolution x-ray or NMR structure. However, structural determination of GPCRs is difficult, mainly because large membrane-bound proteins depend so critically on the surrounding lipid bilayer and the lipid-water interface for both their structure and their function. In a very real sense, the function of integral membrane proteins such as GPCRs requires the intimate association with the lipid bilayer that complicates the more traditional methods of structure determination and interpretation. It would be difficult to appreciate the scale of the heroic efforts that would be required from hordes of x-ray crystallographers and NMR spectroscopists to understand the details of such complex events as agonist binding, G-protein binding, the allosteric reaction of the receptor to the presence of either agonist or G-protein, and the processes of signal transduction. Nonetheless, research efforts in these areas are ongoing and undoubtedly will eventually bear fruit. Fortunately, we may not need to await the outcome of this effort. Recent advances in the area of protein homology modeling have enabled the development of structural hypotheses that attempt to unify much of the available experimental data, and they can be used as a framework for designing additional experiments and can provide a basis for understanding the biochemical processes involved. The key to building such models lies in the interpretation of experimental data, such as the effects on ligand binding of chimeric receptors and site-directed mutant receptors, fluorescence quenching, NMR (REDOR), site-di-rected spin labeling, etc. In general, these are simpler experiments than those required for high-resolution structure determination, but they will (almost) always yield less three-dimensional information; the task confronting us then is to weave these threads of information into a fabric of understanding. Herzyk and Hubbard describe a strategy whereby they interpret the results of mutational experiments, biophysical experiments, sequence analysis, and protein structure comparison as distance restraints or constraints between the helical domains of GPCRs. By using a rule-based automated method employing a Monte Carlo, simulated annealing procedure, they derive "footprints" of the heptahelical protein that conform to these data. As described in their article, this method satisfactorily generates the observed electron diffraction footprints of bacteriorhodopsin and rhodopsin. Considering the amount of information used in the model generation, their results are at the same time startling and heartening.

The challenge is clear: advances in the structure-based design and discovery of drugs demand that we invest intellectually in ways of exploiting the deluge of genetic information from various genome-sequencing projects through automated structure-building methods. Although the work of Herzyk and Hubbard addresses the structures

Received for publication 29 September 1995 and in final form 30 September 1995.

0006-3495/95/12/2183/02 $2.00

of bacteriorhodopsin and rhodopsin, the implications for their methodology to protein homology modeling are far-reaching. Doolittle (1995) recently reviewed domain organization of protein structure, and he defines domains as those parts of a protein that can fold independently of neighboring sequences. Put another way, these structural domains provide building blocks for protein structure; in turn, helices, β sheets, turns, and loops can be considered building blocks of the structural domains. The current rate of discovery of new protein structural families and structural domains suggests that although the number of protein sequences seems almost unending, a finite number of structural and functional domains can be assembled in various ways to form the cellular protein machinery (Chothia and Taylor, 1994; Murzin et al., 1995). Therefore automated methods of homology modeling such as those described by Herzyk and Hubbard show great potential in enabling worthy protein models to be built from their building blocks. These methods help lay the foundation for the next big revolution, from sequence to protein structure (other important contributions in this area have been reviewed recently by Eisenhaber et al., 1995). The significance of the work of Herzyk and Hubbard lies in the manner in which they use disparate experimental data to refine a protein model to within 1.9 Å of the electron diffraction structure, whereas each datum poorly determines the structural possibilities; when taken together, they provide powerful structural restraints. With these models in hand, experiments can be designed to test the various structural hypotheses or models. What is lacking are the experimental strategies that directly address ways of recognizing and defining structural domains and that probe the many ways in which these domains might assemble in three dimensions. The all-or-none philosophy of many structural biologists might have to be tempered as the structural problems become more difficult, as the biological processes become more complex, and importantly, as the potential of three-dimensional models is recognized.

REFERENCES

Chothia, C. and W. Taylor. 1994. Sequences and topology. *Curr. Opin. Struct. Biol.* 4:381–382.

Doolittle, R. 1995. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* 64:287–314.

Eisenhaber, F., B. Persson, and P. Argos. 1995. Protein-structure prediction: recognition of primary, secondary and tertiary structural features from amino-acid sequence. *Crit. Rev. Biochem. Mol. Biol.* 30:1–94.

Murzin, A., S. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247: 536–540.

von Heijne, G. 1995. Membrane protein assembly: rules of the game. *Bioessays.* 17: 25–30.