RESEARCH INTERESTS

Dr. Judith Klein-Seetharaman

The long-term objective of my present and future research is to study the relationship between sequence, structure and function of proteins. My interests and approaches in this area are outlined in Figure 1. They can be broadly categorized as (1) understanding the molecular mechanisms of signal transduction at cell membranes and (2) exploring the analogy between natural languages and biological data.
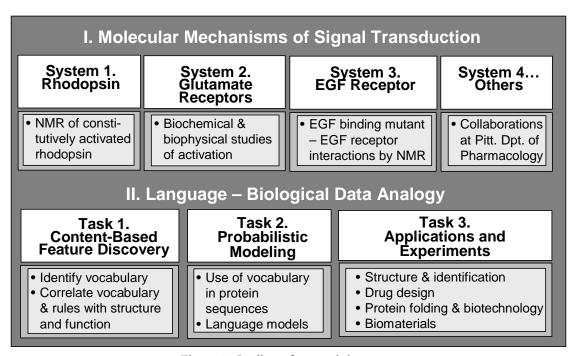

I. Molecular Mechanisms of Signal Transduction

System 1. Rhodopsin	System 2. Glutamate Receptors	System 3. EGF Receptor	System 4… Others
• NMR of constitutively activated rhodopsin	• Biochemical & biophysical studies of activation	• EGF binding mutant – EGF receptor interactions by NMR	• Collaborations at Pitt. Dpt. of Pharmacology

II. Language – Biological Data Analogy

Task 1. Content-Based Feature Discovery	Task 2. Probabilistic Modeling	Task 3. Applications and Experiments
• Identify vocabulary • Correlate vocabulary & rules with structure and function	• Use of vocabulary in protein sequences • Language models	• Structure & identification • Drug design • Protein folding & biotechnology • Biomaterials

Figure 1. Outline of research interests.

I. Molecular Mechanisms of Signal Transduction

My specific research interest (I in Figure 1) is to study by biochemical, biophysical and computational approaches the mechanisms of action of signal transduction protein complexes at cell membranes. I am specifically interested in three specific systems, which are outlined in Figure 1, and described in more detail below.

(1) Rhodopsin
The primary photoreceptor molecule in visual signal transduction.
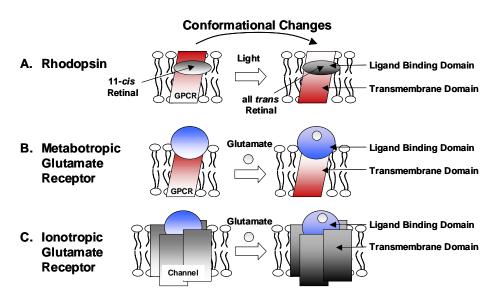
Rhodopsin is the prototype for the G-protein coupled receptor (GPCR) family, the largest family of cell surface receptors known. Recently, the three-dimensional structure of dark state rhodopsin has been obtained by X-ray crystallographic methods. However, the mechanism of signal

transduction by rhodopsin still remains to be elucidated with molecular detail. This requires precise determination of the changes in conformation when forming the activated state. Solution NMR spectroscopy of isotope labeled rhodopsin will be used to evaluate structural differences between rhodopsin in the dark and in rhodopsin mutants with (a) elongated half-lives of the light-activated species and (b) with amino acid substitutions causing constitutive activity. Due to the high level of structural and functional information known for rhodopsin, this system will serve as the standard for the other systems proposed to be studied, especially the following, system 2.

(2) Glutamate receptors
Primary excitatory receptor systems in the brain involved in synaptic plasticity and memory storage.

Glutamate receptors (GluR) are characterized by a soluble extracellular ligand binding domain with homology to bacterial periplasmic binding proteins but diverse transmembrane (TM) domains: (a) ligand-gated ion-channels in ionotropic GluR, (b) K^+channel in prokaryotic GluR and (c) seven TM helical bundle characteristic for GPCR in metabotropic GluR. Ligand-induced conformational changes in the extracellular domain are well understood but how these are transmitted to the TM and cytoplasmic (CP) domains – central to the understanding of the mechanism of signal transduction by GluR – remains unknown. In particular, the relation of the mechanism in metabotropic GluR to that in rhodopsin (Figure 2A,B) and the relation of the mechanism in metabotropic to ionotripic GluR (Figure 2B,C) are intriguing questions.

Figure 2: Schematic Representation of Ligand-Induced Protein Conformational Changes in Different Signal Transduction Proteins Discussed in this Research Statement.
A. Rhodopsin. **B.** Metabotropic Glutamate Receptors. **C.** Ionotropic Glutamate Receptors.
In rhodopsin (A.) and metabotropic glutamate receptors (B.), the transmembrane domains are conserved and presumed to be similar, while the ligand-binding domain is different. In metabotropic glutamate receptors (B.) and ionotropic glutamate receptors (C.), the ligand binding domain is conserved, while the transmembrane domains are different.

To address these questions, high-level expression systems and purification and folding schemes for active receptor proteins will be developed. GluR proteins, with initial focus on GPCR type GluR, in different ligand-bound forms will be subjected to cysteine reactivity measurements,

disulfide cross-linking, EPR spectroscopy of spin labeled derivatives, fluorescence spectroscopy and NMR spectroscopy of ^{19}F, ^{31}P or ^{13}C labeled derivatives and of isotope labeled proteins containing $^{15}N/^{13}C/^{2}H$-labeled amino acids. Computational biolinguistics approaches (see II below) will be used to predict sites important for folding and function of the receptors. The studies will contribute to the understanding of the molecular basis for different pharmacological subtypes of GluR and may help drug design for treatment of neuronal and psychiatric illnesses in which dysfunction of glutamate signal transduction pathways has been implemented.

(3) Epidermal growth factor – receptor interactions
Important for proliferation and differentiation of many cell types.

Epidermal growth factor (EGF) mutants with enhanced binding affinities to the extracellular domains of the EGF receptor (EGFR) exist. The structure of these EGF-mutants and EGFR alone and in complex will be studied by approaches as in section 2. Since EGFR is over-expressed and/or constitutively activated on many tumors, including brain, mammary and ovarian, the generation of antagonist EGF ligands that bind to the receptor without inducing activation could have potent cancer therapeutic applications. Implications of the results of this project in biomaterials research will be explored.

II. Language - Biological Data Analogy

Whole-genome sequencing projects have made a large amount of gene sequence data available, increasing the urgency in addressing the challenge of understanding the structure and function of the proteins that these genes encode. Recently, I have been focusing on acquiring computational skills to approach this goal. In this effort, the following central hypothesis is proposed:

*The collective of **protein sequences** encoded by the genomes of different organisms are "written" representations of different "**languages**" or "dialects".*
*The ability of each protein to **fold** into a three-dimensional structure able to participate in functional networks of proteins in a cell corresponds to the **meaning** of a text.*

The hypothesis provides a means to approach the above challenge by application of sophisticated computational tools developed for the analysis of human languages to protein sequences. The first task in this project (Figure 1) is to identify a vocabulary of a genome sequence. What are the equivalents of words and what are the word boundaries in protein sequences? What are the rules that allow creating meaningful sentences using this vocabulary? Increasing understanding in these areas will be reflected by the ability to predict protein sequences through probabilistic models (Task 2, Figure 1). This exploration of the analogy between language and biological data has various applications. For example, "word phrases" that a pathogen uses but not its host may lead to identification of drug targets. "Rare words" seem to indicate the presence of initiators of protein folding. It is tempting to speculate that a sequence "written" in the "language" of one organism could be "translated" into that of another organism, thereby allowing expression of functional e.g. human protein in bacteria where the the wild-type sequence would result in misfolded aggregates.