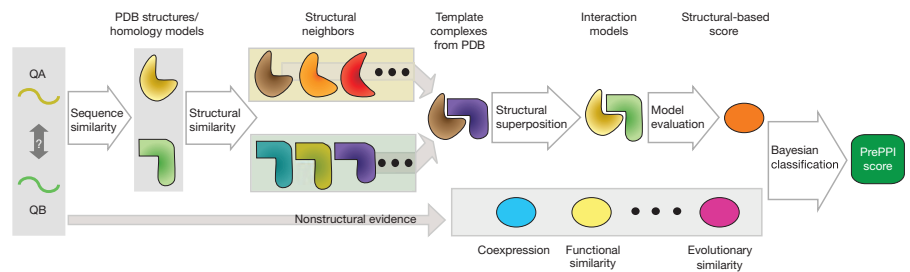**BIOINFORMATICS**

Predicting PPIs

An algorithm to predict protein-protein interactions using structural information yields results comparable in accuracy to those of high-throughput experiments.

The PrePPI algorithm uses structural and functional evidence, coupled with Bayesian statistics, to evaluate the likelihood that two query proteins (QA and QB) interact. PDB, Protein Data Bank. Reprinted from *Nature*.

When researchers in two established fields come together, remarkable things can happen. Sometimes they can lead to new, powerful approaches that tackle a problem of common interest. This was the case at Columbia University in New York, where systems biologist Andrea Califano and computational structural biologist Barry Honig joined forces to develop a tool, dubbed PrePPI, to predict protein-protein interactions using available structural information. "We have created an environment at Columbia where researchers working in the two areas, including ourselves, interact on a regular basis," say Honig and Califano in a joint statement. "The concepts and methods that underlie PrePPI are taken from both fields, but their combination required the sort of interdisciplinary cross-fertilization that we hoped would occur when we started working together some years ago."

The goal of large-scale interactomics projects is to understand, on a proteomic level, who interacts with whom in the cell. Such data allow researchers to interpret protein function and generate hypotheses for individual follow-up experiments. A cornucopia of methods is available for generating interactome data, ranging from the classical yeast two-hybrid method, to affinity purification–mass spectrometry, to prediction tools. However, to date, structural information has played almost no part in helping to generate interactome data—despite the rapidly growing availability of structural representatives of all known protein families, boosted by structural genomics efforts, and despite the recent progress in the ability to produce good homology models from experimental structures. "Homology models sometimes get a bad rap," says Honig, "but it should be obvious that many of them contain very useful information."

The PrePPI algorithm takes advantage of this available structural information, as well as functional evidence, and uses robust Bayesian statistics to predict protein interactions. The algorithm first uses sequence alignments to search far and wide for structural 'representatives'—either experimentally derived structures or homology models—for a pair of putative interacting 'query' proteins. Next it uses structural alignment to identify both close and remote structural 'neighbors' of the representatives. If such a complex can be found in the Protein Data Bank, this serves as a template for modeling the interacting query proteins, which is achieved by superimposing the structural representatives on the template. This process generates millions of interaction models, which are then evaluated by a five-pronged empirical scoring system. Functional evidence such as coexpression, functional similarity and evolutionary similarity is used to further refine the results. The scores are then combined using Bayesian statistics to generate a likelihood ratio that the query proteins represent a true interaction.

Applying PrePPI, the team predicted 30,000 high-confidence protein interactions for yeast and 300,000 for human. Not only was PrePPI's performance superior to that of previous prediction algorithms that do not rely on structural information, it was also comparable to and even somewhat better than experimental high-throughput methods to identify protein interactions. "We were both delighted and surprised by PrePPI's performance," says Califano. Although there was little overlap between the PrePPI results and those from various experimental methods, this is not surprising, he notes, because orthogonal methods are known to produce complementary results. And a little bonus feature provided by PrePPI over experimental methods is that PrePPI provides a crude model of the interaction interface.

The researchers validated 19 of PrePPI's novel predictions using co-immunoprecipitation experiments across four individual labs. Of the 19 predicted interactions, 15 were found to be true. Many of the PrePPI predictions are likely to be false positives, but this is the nature of the method, the researchers say. "Our results will improve as more structural information becomes available," says Honig. "Continuing improvements in homology modeling technologies will also have significant impact." PrePPI is also not yet trained to handle unstructured regions in proteins, which, the structural biology field is coming to realize, mediate many protein interactions.

PrePPI's main value is as a hypothesis generator, note Honig and Califano. They hope that other researchers will use their yeast and human results, as well as the tool itself, to focus investigations into protein function.
Allison Doerr

RESEARCH PAPERS
Zhang, Q.C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **490**, 556–560 (2012).