


FIRST ASIAN GENOME SEQUENCED

 Individual genomes get boost from Chinese effort.
www.nature.com/news

The shape of protein structures to come

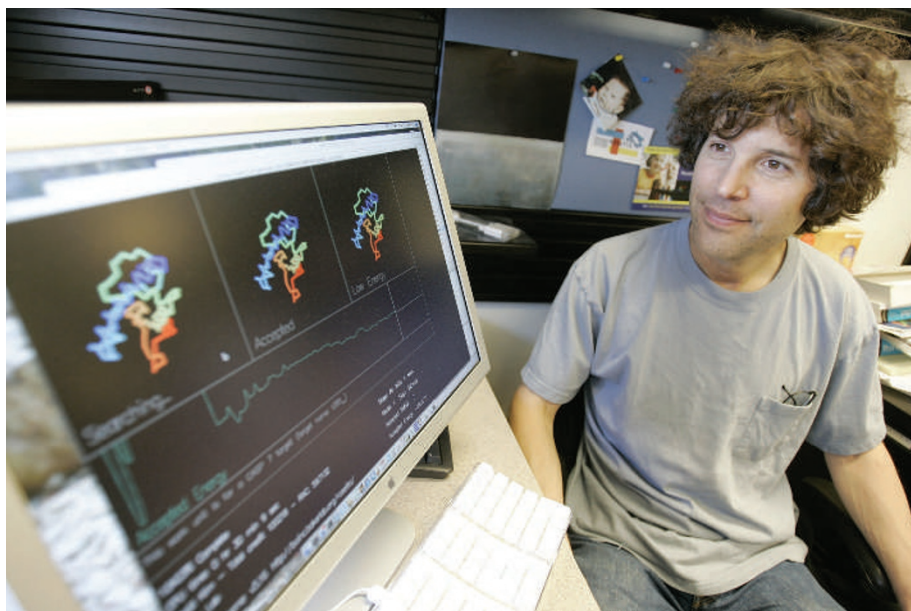
By exploiting millions of hours of computing time donated by the users of 150,000 home computers, scientists have predicted the structure of a protein using just its sequence of amino acids. The project marks a significant advance in a field that's been flush with hope yet short on tangible results, experts say.

Determining the shape of a protein is normally a matter of firing X-ray beams at its crystalline form and measuring their diffraction, and protein chemists have long been sceptical of attempts to replace this practice with modelling or theory. "Modelling right now has a terrible name in the field," says Michael Levitt, a computational biologist at Stanford University. But in a *Nature* paper published online on 14 October, David Baker, a biochemist at the University of Washington in Seattle, and his colleagues report results that may do much to dispel that scepticism (B. Qian *et al. Nature* doi:10.1038/nature06249; 2007).

A protein's shape — and therefore its activity — is determined by the precise way its constituent string of amino acids folds up. "If you think of the whole thing as like an utterly flexible snake, there are hundreds of degrees of freedom at every site," says Eleanor Dodson, a structural biologist at the University of York, UK. The final shape depends on the molecular interactions each amino acid has with its neighbours, with surrounding water molecules and with other amino acids that are a long distance away in terms of sequence, but which become close as a result of folding. It's a horrendous problem to model.

Rather than trying to solve the problem from first principles, Baker's technique combines information from the sum of what is already known about protein structures with the vast computing power available through the Berkeley Open Infrastructure for Network Computing. This software, developed at the University of California, Berkeley, allows people to contribute spare computing power on their desktops to scientific projects (most famously, the search for extraterrestrial intelligence, in the form of SETI@home); 150,000 volunteers used it to download a copy of the Baker lab's Rosetta@home program.

Rosetta breaks a protein's sequence into short stretches that can be matched to identical stretches from proteins with known structures. These shapes offer many ways to sew the protein under study back together, and



David Baker admires the handiwork of 150,000 computers.

the program chooses those that minimize the free energy of the structure (a measure of its stability). By running the program over and over again on thousands of computers, the researchers lurched towards an ever more accurate protein model.

When fed the sequence of T0283, a 112-amino-acid protein from a bacterium, the network spat out several million structures after a million hours of computing time. Those millions were whittled down to five by further repeated computer analysis, and one of the structures was spot on, correlating with the structure as determined from its crystal.

Although the structure's precision was not on a par with high-resolution crystal models, it was good enough for researchers to think that the technique could simplify the process of obtaining X-ray structures in the future. To

turn X-ray patterns into structures, researchers must produce patterns from crystals that have either been spiked with heavy-metal 'markers' or come with some indication of what the final structure will look like, for example from the shape of a related protein. Rosetta's structure for T0283 was good enough to function in this way. The program should now be able to provide such reference points for proteins for which there are already X-ray data, but which lack useful relatives and whose structures thus

languish unsolved. "You will be able to solve a whole bunch of these structures rather quickly," says Adrian Roitberg, a protein modeller at the University of Florida in Gainesville.

There is still room for improvement, though, says Rhiju Das, a postdoc working on the Rosetta@home project and a co-author on Baker's paper. Each home computer works in isolation, he explains. If the program could be rewritten to run on the many parallel processors in a supercomputer, Rosetta might become considerably more powerful.

One pay-off of better structure prediction would be the prospect of custom-made proteins, says Baker, who uses Rosetta to hunt for sequences that correspond to desired structures. His lab and that of University of Washington biochemist Bill Schief are currently working on redesigning the gp120 protein of HIV to make a vaccine that could stimulate the immune system in a different way from the natural virus. The reshaped protein should elicit antibodies that attack the virus more effectively than antibodies created after infection.

The days when protein modellers thought they could make crystallization obsolete are long gone, Baker adds. But melding the two techniques could offer biologists insight into many more proteins — and faster. "If you really care about the structure of your protein, you should get some experimental data and combine it with modelling," he says.

Ewen Callaway

"If you think of the whole sequence as like an utterly flexible snake, there are hundreds of degrees of freedom at every site."