once the station is ready," says Baldwin.

One pressing problem is how to maintain interest in a program with few research opportunities until the space station is ready to do science in 2005. "We're at a very critical stage in building the space station," says Gaffney, "and not to do good science once it is ready would be an absolute tragedy."

–ANDREW LAWLER

HUMAN GENOME PROJECT

Chromosome 21 Done, Phase Two Begun

Chalk up two achievements—one public relations, the other scientific—for the international consortium in the race to sequence the human genome. This week, on the eve of an annual genome meeting at Cold Spring Harbor Laboratory in New York—and, perhaps, to preempt an expected statement from rival Celera Genomics of Rockville, Maryland—the international consortium announced on 8 May that it has completed (almost) "phase one" of the project, the rough draft of the human genome. (85% of the promised 90% of the draft sequence is now available in GenBank, says the consortium.) On 9 May, the consortium entered "phase two" and turned its collective sequencing firepower to "finishing" the human genome—that is, producing a 99.99% accurate sequence. Celera

Gold standard. The highly accurate sequence of chromosome 21 should help illuminate Down syndrome.

CREDIT: (RIGHT) RCSB

head J. Craig Venter dismissed the announcement as an "artificial" milestone and "science by press release"—a tactic he has used himself.

But Venter lauded the consortium's other announcement: the complete sequence of chromosome 21, only the second chromosome to be finished. Already, this chromosome, published electronically this week and in print in the 18 May issue of *Nature*, has reached the gold standard for which the consortium is striving for the entire genome, says Yoshiyuki Sakaki, who directs the human genome sequencing effort at the Institute of Physical and Chemical Research (RIKEN) outside Tokyo and whose team did half the sequencing.

Together, the 62 scientists from 13 labs have determined the identities of 33.5 million bases of the long arm and another 280,000 bases of the short arm of chromosome 21. Just three clone gaps remain—stretches, each about 30,000 bases long, that could not be determined with current technology. By contrast, chromosome 22, which is roughly the same size and was completed last December, has 10 such gaps (*Nature*, 2 December 1999). In another technical tour de force, the chromosome 21 team can also claim the longest contiguous stretch of DNA ever sequenced, at 25.5 million bases, says Sakaki.

Sequencing proceeded so quickly in part because several groups interested in Down syndrome had begun mapping chromosome 21 in the 1980s, before the Human Genome Project even existed. In Down syndrome, an extra copy of chromosome 21 results in mental retardation, heart problems, and other abnormalities. Now the task of figuring out just what goes wrong in this devastating disease will be far easier. Instead of finding the expected 800 to 1000 genes, gene prediction programs came up with just 225. "We now have a real definition of who the [genetic] players are," says Roger Reeves, a geneticist at Johns Hopkins University School of Medicine.

The paucity of genes on chromosome 21—one 7-million-base stretch contains just one gene—may have broader implications as well: If the gene prediction programs prove correct, then the entire human genome could have less than the 100,000 genes previously estimated.

As sequencers enter the home stretch, competing claims of genome accomplishments should be coming fast and furious. But, says Reeves, it will be quite a while before any sequence beats the accuracy and completeness of chromosome 21.

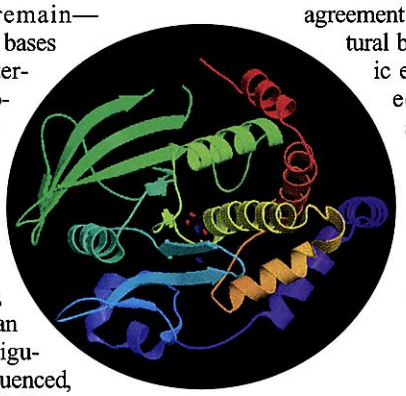
–ELIZABETH PENNISI

With reporting by Leslie Roberts.

STRUCTURAL GENOMICS

Protein Data Justice for All

Scientists who crack protein structures and colleagues who want to decipher what these proteins do are on the verge of a watershed agreement that would usher structural biology into the genomic era. The carefully crafted guidelines are designed to help coordinate international financing of publicly funded protein structure efforts and ensure prompt release of structure data so that no team has an unfair advantage in working out the functions of unknown proteins.



Fast and fair. New guidelines urge teams to quickly submit structures to repositories such as the Protein Data Bank, which last week released this image of human protein tyrosine phosphatase–1B.

The guidelines, being finalized as *Science* went to press, come at a time when robotics and computer automation promise to transform structural biology into a high-speed effort, dubbed "structural genomics," in which researchers will churn out thousands of protein structures in the next 5 years. Nurturing this souped-up approach, the National Institutes of Health (NIH) this fall plans to fund up to six structural genomics pilot centers to establish and test techniques for high-throughput x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, the workhorse technologies of structural biology. Similar approaches are being adopted or considered in Japan, the United Kingdom, France, Brazil, and Germany (*Science*, 17 March, p. 1954).

To help coordinate these efforts, officials at NIH and Britain's Wellcome Trust last month brought some 50 leading protein specialists to the Wellcome Trust Genome Campus in Cambridge, U.K., for a brainstorming session on how to release data quickly and fairly. They had a gulf to bridge.