

A chip of genetics

Designing molecules on the computer

SANDHYA RAO
in Madras

ANSWERS in search of questions. That is how Venkat Venkatasubramanian describes computer-aided molecular design (CAMD). Applying Charles Darwin's principle of evolution and natural selection to artificial intelligence, this professor of chemical engineering and his team at Purdue University, U.S., have been using computers to design complex molecules for the creation of new materials. The programme would benefit industries concerned with pharmaceuticals, agrochemicals, fertilizers, pesticides, herbicides, motor oils, paints, varnishes, even perfumes.

"It's like doing chemistry on the computer," says Venkatasubramanian. The process of hypothesising, synthesising, testing and revising several hundred times a new drug, for example, in a laboratory, would take some seven to 10 years before all the parameters are met and at phenomenal cost.

"If I can cut down the time, cost and effort involved by a factor of two, it will be a tremendous advancement," he says. And that is what his Genesys computer - with its biblical allusion to the creation of the world - has been showing progressively in the last 10 years of work in this direction.

There are two ways of looking at problems - forward and inverse. The former raises questions and looks for answers and is the more widely-followed method. "Basically in materials design, you know what you're looking for. If it's a drug, then you know what biochemical properties it should have in order to cure a disease. The answer is known, the question is not. Because the answer is the property. You know how it should behave, but you don't know what kind of structure will give you those kinds of properties. This is known as the inverse or reverse problem in molecular design."

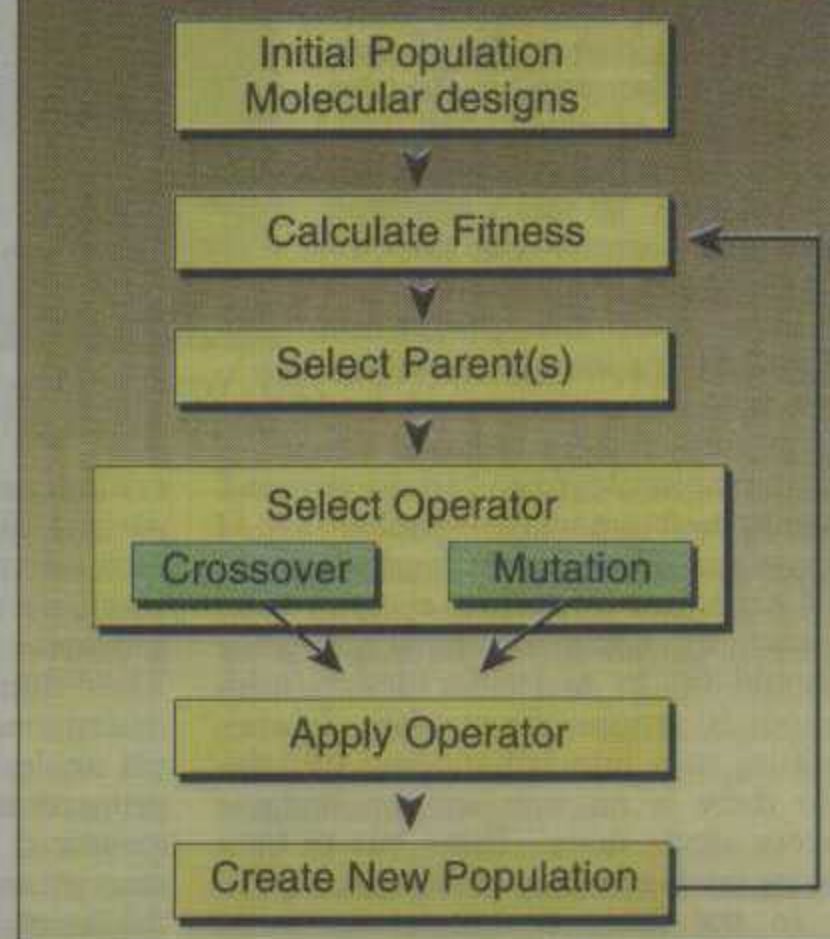
In 1986, when he was teaching at Columbia University after his Ph. D., Venkatasubramanian asked himself: How does nature design? How did complex biological molecules evolve? Soon he arrived at the survival of the

fittest theory - and decided to test it in the computer.

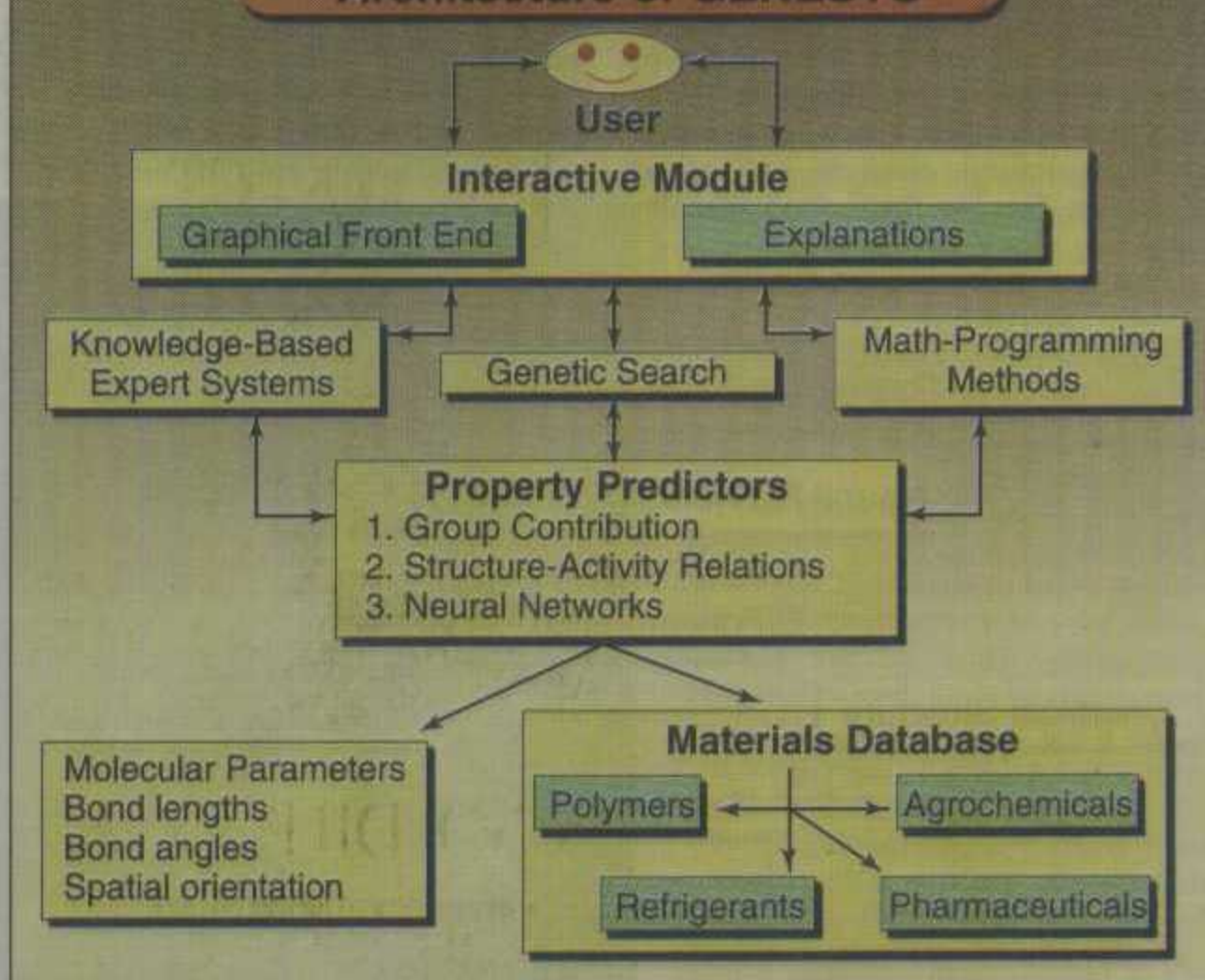
"When you put the Darwinian model into the computer it becomes genetic algorithm," he explains. His group is the first in the U. S. to demonstrate the use of genetic algorithm - a mathematical process which has been around for 25 years - in chemical engineering. Beginning with designing known molecules, the Purdue team successfully moved on to unknown ones.

If all this seems esoteric, take the example of refrigerants. Freon - a gas at room temperature that can be liquefied under pressure - has long been used as the cooling agent in refrigerators and air-conditioners. But it is a

Overview of Genetic Algorithms



Architecture of GENESYS



chlorofluorocarbon (CFC) and attacks the ozone layer in the atmosphere. In their experiments, therefore, Venkata-subramanian and his team were looking for non-CFCs.

"We told Genesys that we were looking for an eco-friendly refrigerant with all the good things of Freon but not the bad," he says. "It searched and came up with a number of structures which are feasible refrigerants." When the researchers called the industry in the U. S., they found that some of these had been discovered in the laboratory by scientists and were being developed as commercial refrigerants. "It was a confirmation that Genesys was not producing nonsense," he adds.

How does the programme work? "Let's imagine I create a big flask in computer, where I randomly create a hundred different molecules from various groups," Venkatasubramanian explains. "I know that these are the kind of groups I'll need for a refrigerant except I don't know in what combinations they should fit. In molecular design, with about 20 groups, the number of possibilities runs into billions and trillions. So there is no way we can imagine every single thing. There has to be a more intelligent way of doing it."

In the flask, the molecules swim around and bump into one another – this can be displayed on the computer – and produce reactions which are all guided by chemistry. The products are all chemicals. Some molecules break apart and recombine with parts of other molecules, resulting in large structural changes called crossovers. Small structural changes are called mutations. It is like the mating of molecules.

The new molecules are the first gen-



V. Venkatasubramanian... "a promising start where the potential has been demonstrated".

eration, or offspring of the primordial parents. By looking at what each new molecule is made of and how it is put together, it is possible to predict the properties these new structures have. The "fitness" of the molecule is then determined by comparing with the target molecule, which satisfies all the properties of a particular structure. It is measured on a scale of zero to 100, zero meaning the molecule is way off the target, and 100, it is just right.

The survival of the fittest idea is pressed to select only the top few (with higher fitness percentages) and eliminate all the other molecules. These are made to interact and produce the next generation of molecules. Once again the fitness test is applied and the process is repeated generation after generation, saving only the best. Thus, the fitness of the molecules evolves. By randomly improving their ability ("... mathematically you cannot prove that

each generation will necessarily improve in fitness. This is known as the convergence problem. You may start with very bright parents, but the kids are all stupid..."), they finally hit 100 per cent, which means they find the target.

Sometimes this happens in 100 generations, that is, the 100th pair is the molecule they are looking for. Sometimes it takes 1,000 generations. Sometimes it happens quickly and sometimes it may never happen. The heuristic approach, in which there is no guarantee, is the one drawback of CAMD. But, says Venkatasubramanian, the up side is that it is likely to work 95 per cent of the time.

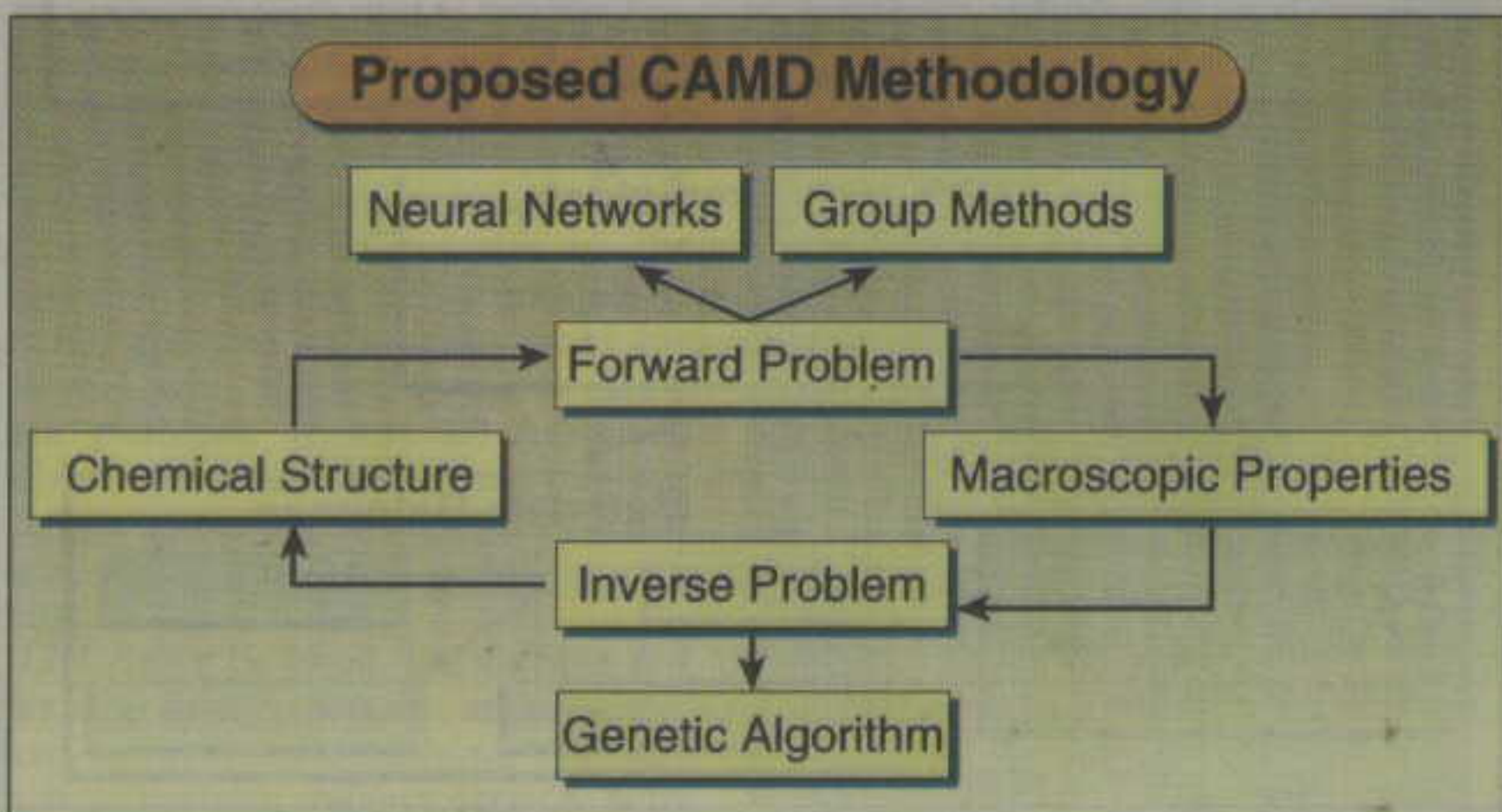
If the simulation is accurate, the results are nearly the same as those arrived at in a laboratory. "Nature can never be pinned down exactly," Venkatasubramanian says. But the results are close enough to be used effectively.

And how soon is fast? In the case of polymers, for example, "typically, it takes less than 500 generations. In real time, it's about 10 minutes". In other words, you can give the computer the problem, go for coffee, come back and find the problem solved.

Of course, it is not as simple as it sounds – but it could become so. "This is a promising start where the potential has been demonstrated," clarifies Venkatasubramanian. "But, for CAMD to become a routine tool, it is about 5-10 years away. We now have to see how effective the system is in large-scale applications and how it fits with other techniques people use. Typically, from invention to complete deployment and acceptance takes 10 to 20 years. The personal computer was invented in the 1970s, but is being used widely only now."

In principle, there are no limits to what can be created by artificial intelligence. And industry comes in when the manufacture of novel materials is to be taken up. Currently, two industries – polymers and pharmaceuticals – are interested in the Purdue programme and expect to produce results in a couple of years.

There is a sense of wonder as Venkatasubramanian speaks of the survival of the fittest. "It's a very simple idea. But, take that idea and give it 500 million years and see what it has produced. It has produced us." But what are the odds that no matter what fiction is factualised by artificial intelligence, nature will always remain one step ahead? ■



LIFE
tain
There a
land-dwe
the land.

From n
distant u
State Bu
tree cano
resent th
now und
as much
Manhatta
its own c
of individ

Nalini
College i
that the i
"from the
feet und
leaves, 30
On the
and bush
Above th
small, no
well as y
the light.

The to
tall, matu
where, it
the worl
According
Institute
Millbrook
forest can
leaves an
the total s

The lif
incredible
plants at
and make
sis plus t
remarkabl
teria an
Bromeliad
cups of ov
ited by o
imals. Dr.
study show
and store v
forest doe
totally on
its roots.

Underst
vital to th
phere, incl
an importa
by a multi
John Grad