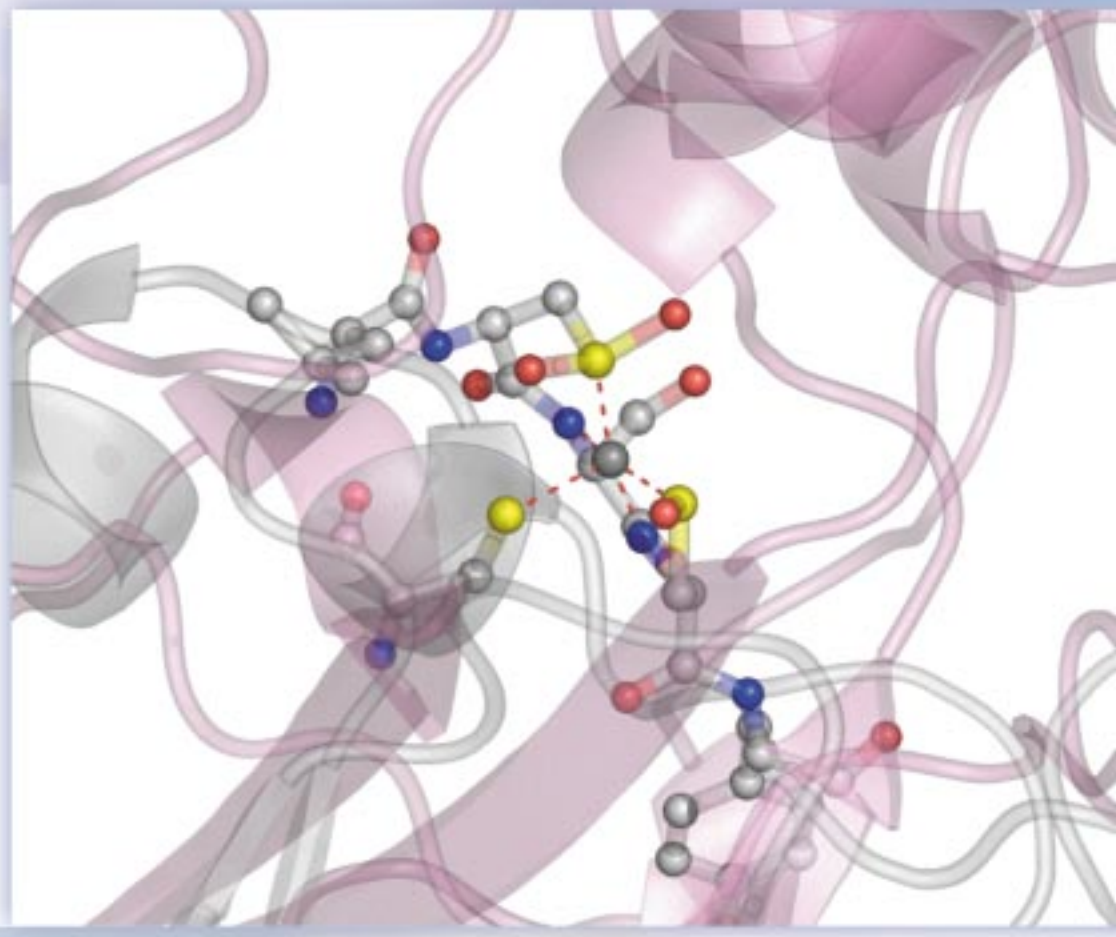


*Right: The active site of a nitrile hydratase molecule showing the binding of cobalt (grey) to three sulphur (yellow) atoms from cysteine amino acids. Nitrogen atoms are shown in blue and oxygen in red.*
Picture: Serah Wangari Kimani using PyMol

There's huge commercial potential within the molecules of life, but scientists first have to understand the structures of proteins at the deepest atomic level. Unlocking these secrets, says Trevor Sewell, gives South African biotechnology a key to the treasure chest of global economic success.

# Treasure from the molecules of life

'First-generation' biotechnology has flourished proudly in South Africa. Using organisms in a targeted way, it has enabled the country to set up and grow one of the world's largest brewing companies, a globally renowned wine industry, and important programmes in plant and animal breeding. 'Second-generation' biotechnology went a step further by breeding useful characteristics into microorganisms. Now the pressure is on to realize the benefits of 'third-generation' biotechnology, which engineers molecules and cells to create products such as industrial enzymes, therapeutic agents, pesticides, and vaccines.

Exploiting these third-generation possibilities depends on understanding how proteins are structured. The new Programme in Structural Biology, initiated jointly by the University of Cape Town (UCT) and the University of the Western Cape (UWC), plans to make certain that South Africa has the expertise to be competitive in this rich and rapidly growing field.

Generously funded in its initial stages by the Carnegie Corporation of New York, the Programme in Structural Biology arose through the collaboration of academics at the Universities of Cape Town, Western Cape, Stellenbosch, and the Witwatersrand. Currently, its 10 master's, one doctoral, and one postdoctoral student are working on projects related both to drug design and industrial enzymes.

## Why examine protein structures?

In the last ten years, the ready availability of genetic information has revolutionised biological science. Quickly and affordably, geneticists can obtain the sequences of entire genomes. These sequences provide a catalogue of the proteins that make up individual organisms, as well as details of the proteins' composition and how they relate to one another. Scientists in the field of bioinformatics analyse this wealth of information and make it accessible using powerful computers and communication networks.

Now scientists are seeing that all this information gives only part of the answer. Although gene sequences make sense to the cell, we humans barely understand them. For more insight into how cells actually work, we have to study the proteins within them – no easy task.

The information contained in genes is used to create proteins. Proteins do most of the 'work' of the biological cell: they catalyse reactions, transport molecules around, detect and respond to the environment in which the cell is situated, and control the way genetic material is expressed in the formation of cells (*see box*). They make cells into exquisite, well-organized, and efficient machines.

Enzymes are proteins that act as chemical 'reactors', converting one substance into another. This makes them interesting to people for two reasons. First, we can use them to make new chemicals of our own – this gives rise to protein engineering and industrial enzymology. Second, we can use our knowledge of them to design molecules that interfere with the life-giving processes of harmful pathogens and insect pests – this leads to the creation of drugs and pesticides.

## The making ('expression') of a protein

gene (DNA)

Made of DNA, the gene contains information (coded as a sequence of chemical elements called 'bases') needed for a particular protein to be made or 'expressed'. That information is then 'transcribed' (or, converted) into a form that can be carried by messenger RNA to the ribosome.

messenger RNA

ribosome ← transfer RNA

A multimolecular complex made of protein and RNA, the ribosome acts as a 'machine' for making proteins, and:

- positions an adapter molecule called transfer RNA in such a way that it docks with the messenger RNA on one side and locates the amino acid appropriately on the other side
- builds the protein up, amino acid by amino acid, by forming peptide bonds.

protein

A protein is a biological macromolecule made up of a specific sequence of amino acids linked by peptide bonds[1]. Proteins vary considerably in length and amino acid composition. Most of them fold spontaneously into a 3-D form that is characteristic of the amino acid sequence. Though many thousands of protein folds are known, it is not yet possible to predict the shape of a protein unless the structure of a protein with a very similar sequence is already known.

Many pathogens and pests in Africa are unique. In addition, the biodiversity of our continent creates great possibilities for new discoveries. It's the responsibility of Africans to overcome the problems posed by our pathogens and pests, and also to exploit the opportunities that our biodiversity offers. So far, we have been slow to take up these challenges. To help South Africa improve in these fields, the UCT and UWC groups have focused on the study of protein structures.

### What's involved?

Exploring the inner space of protein structures – the foundation of third-generation biotechnology – is as complex and difficult as it is fascinating.

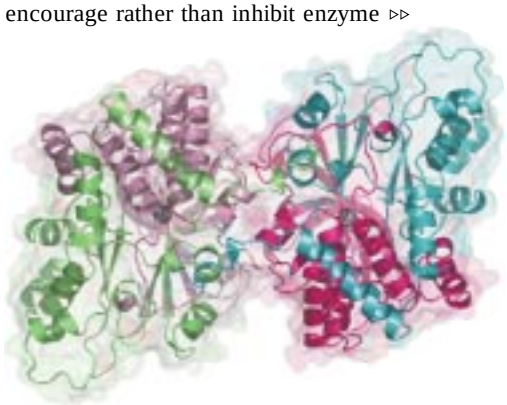
Nature uses up to 20 amino acids to make proteins. Each has a different side chain[1] with its own special chemical properties. The amino acids that together make up a protein are linked in long chains, and every different type of protein has its own specific sequence of amino acids. These long chains generally fold in such a way that a specific sequence produces a corresponding 3-D (three-dimensional) formation. This folded formation is what gives the proteins their interesting properties.

Studying protein folds and their consequences is one of the jobs of structural biologists. Understanding the 3-D structure provides unique insights into how proteins work, how they interact with one another and with other biomolecules, and how these interactions produce higher-order complexes and, ultimately, cells.

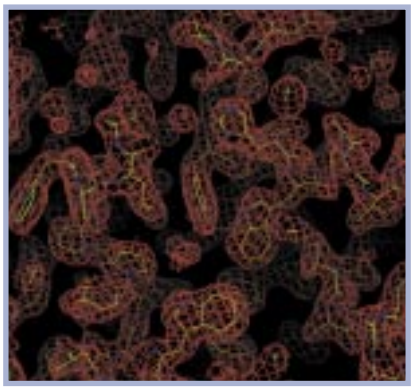
Knowledge of protein structure satisfies our curiosity. But it's also shaping our technology in two important and different ways. First, it helps us to design molecules that can be used as drugs and pesticides, and, second, it helps us to tailor Nature's enzymes to produce valuable industrial chemicals.

Once we know the shape of the active site[2] of an important molecule of a pathogen that causes disease, for instance, we can, in principle, manufacture a chemical compound that inhibits it in some way – by slowing down the activity of the molecule or even stopping it altogether. Peter Colman and his team in Australia used this approach when they designed the first 'rational' anti-influenza drug, Relenza, about a decade ago. Influenza viruses produce an enzyme, neuraminidase, that weakens the host cells of an infected person or animal so that the virus can gain entry. Once they had worked out the protein structure of neuraminidase, Colman and his colleagues were able to create Relenza, which specifically targets the active site of neuraminidase and in this way renders the virus harmless.

Designing industrial enzymes means solving the inverse problem of how to encourage rather than inhibit enzyme ▹▹



*Computer generated 'cartoon' showing the 3-D structure of nitrile hydratase and the overall fold of the molecule. The spirals represent alpha helical regions stabilized by hydrogen bonds. Broad arrows represent strands that are linked by hydrogen bonds in a pattern called beta sheets. These fundamental building blocks of protein structure were discovered over 50 years ago by Nobel laureate Linus Pauling.*
Picture: Serah Wangari Kimani using PyMol



*Part of the map of the electron density of nitrile hydratase (*see next page*). The red 'chickenwire' contours are chosen at a single electron density level so as to determine the location of the backbone polypeptide and amino acid side chains. Interpreting the map involves locating the atoms within these contours. Sophisticated computer programs allow experienced users to position the atoms (here shown linked by yellow, green, red, and blue lines) within the density. This map is a very good one, since the volume enclosed by the contours is continuous, allowing for unambiguous positioning of the atoms.*

1 *Side chain*: Each amino acid consists of a carbon atom to which are bonded an amino group ($NH_2$) and a carboxy group (COOH) (as well as a central hydrocarbon group and a variable group). In making a protein, the carboxy group of one amino acid joins the amino group of the next. That joining is called a *peptide bond*. The term 'polypeptide' is used where there are many such bonds: all proteins are polypeptides. Joining onto the carbon is another group of atoms called a side chain. There are 20 different side chains in Nature, each with different properties that form part of the particular amino acid. When the amino acid takes its place in a protein, it is the side chains that do the protein's work.

2 *Active site*: the special area on a protein where the chemical reaction of the enzyme takes place.

▹activity. Many important chemicals – used in various processes, such as drug, plastic, or paint manufacture, for instance – are difficult to make using traditional synthetic chemistry. But in Nature there exist chemicals similar in structure and function that are produced by a series of reactions catalysed by enzymes. The enzymes have been honed by the process of natural selection to operate with exquisite selectivity on natural substrates (that is, on the molecules upon which enzymes act) within the cell environment. The challenge is to understand how these enzymes work and then modify them to operate on industrially important substrates in the more difficult industrial environment. Knowing the enzyme structure is the first important step in this process.

## Journey into the interior

The Mitsubishi Rayon Company in Japan manufactures large quantities of acrylamide, a chemical used in the production of an important polymer, from the readily available 'intermediate' organic chemical, acrylonitrile. As many as 100 000 tons of acrylamide a year are needed globally to make polyacrylamide, which is used mainly for treating municipal and industrial wastewater. A UWC team in the Department of Biotechnology, led by Donald Cowan, has been studying a nitrile hydratase, an enzyme that converts nitriles (such as acrylonitrile) to their corresponding amides (such as acrylamide).

These studies aim to produce enzymes that will make other amides from the corresponding nitriles and – equally important – enzymes that can withstand the rigours of an artificial industrial environment, which is far more hostile than the protein-friendly conditions in the cell. (A living cell has its own salvage processes, for instance, whereas, in an industrial environment, an enzyme must operate in isolation, without the helpful cellular infrastructure that Nature provides.)

As with any hunt for the treasure of economically useful knowledge, the research process is complex and takes place in stages, each depending on the successful completion of the last, and each posing its own problems.

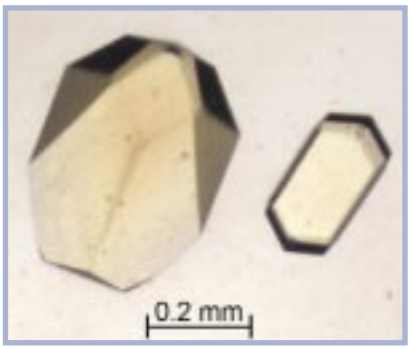
The first challenge is to discover a naturally occurring organism that contains a potentially useful enzyme. To find a suitable one, Cowan looked especially at bacteria that could survive in extreme conditions, such as those in polar ice or volcanic pools. In the end, he selected *Bacillus* sp. RAPc8, a promising bacterium containing an enzyme with nitrile hydratase activity and found in a geothermal pool in New Zealand.

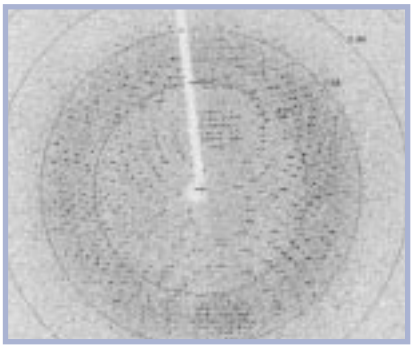
The next problem is that interesting proteins are made in small amounts in their organisms of origin and, because their behaviour can be temperamental, are often difficult to work with. For industrial use, we need a material that's pure and a supply that's abundant. So, borrowing technology from molecular biologists, structural biologists clone the gene for the protein of interest into an 'expression plasmid'[3]. Then, using the plasmid, they transform cells of a more 'laboratory-friendly' bacterium – a specially engineered variety of the bug found in our gut, *Eschericia coli* – to make them capable of producing the protein. These transformed cells can then become 'factories' to make sufficient quantities of the protein of interest; these, in turn, can be further purified. Postdoctoral researcher, Rory Cameron, accomplished this step in our journey.

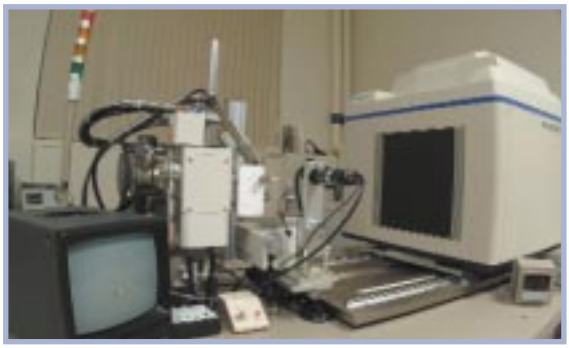
Duly encouraged, the team's immediate goal now was to produce a visual image of the atomic structure of nitrile hydratase – something that, at the beginning of 2004, had never yet been accomplished anywhere in Africa. It seemed a daunting task[4]. We decided to use a technique called X-ray crystallography, in which an intense beam of X-rays is focused on a crystal of nitrile hydratase, resulting in a pattern of diffracted beams that can be recorded on a special electronic device called an image plate.

First, though, we had to grow crystals of the enzyme that would be suitable for X-ray crystallography – that is, crystals of the right shape and size to diffract to a high enough resolution for the atoms to be visible on a computer screen. This was the job of Tsepo Tsekoa. It took him thousands of experiments to obtain crystals suitable for the next step – obtaining X-ray diffraction patterns.

The X-ray beam used for protein crystallography is created by focusing





*Top left: Crystals of nitrile hydratase obtained at room temperature in a solution containing polyethylene glycol (PEG), magnesium chloride, and MES (pH 6.5). Typical dimensions of crystals selected for diffraction studies were approximately 0.2 mm × 0.1 mm × 0.1 mm.*

*Lower left: X-ray diffraction pattern from nitrile hydratase crystals. Many hundreds of such patterns are combined and the intensity of the spots measured. Such spot patterns cannot be interpreted directly but must be converted into a map of the electron density of the molecule that was crystallized.* Pictures: Trevor Sewell

*Top right: Protein crystallographer Muhammed Sayed and master's student James Onyemata prepare to load crystals for data collection.*

*Lower right: The Protein Crystallography Unit at UWC contains state-of-the art diffraction equipment made up of a high-powered X-ray generator and an image-plate detector that records the diffraction data. Another device, the X-Stream, cools the crystal to cryogenic temperatures, preventing radiation damage to the crystals (which would greatly reduce the quality of the data).* Pictures: Cameron Ewart-Smith

3 *Expression plasmid*: a circular piece of DNA containing the gene for the protein of interest as well as the minimum number of additional components necessary for E. coli to make the protein. Normally these plasmids also make (or 'express') a protein that confers antibiotic resistance so that only bacteria containing (or 'transformed' by) the plasmid survive in the presence of the antibiotic.

4 Mastering the various technologies to determine protein structures at the atomic level was a late-20th-century triumph, resulting in the award of at least 10 Nobel Prizes and laying the foundation for 21st-century biological science.

electrons onto a rotating, water-cooled copper block and then focusing the resulting X-rays with a glass capillary. The crystals (cooled to −160°C to preserve them from the damaging effects of the high-energy radiation) are then placed in the beam and, as the crystal is slowly rotated, a series of several hundred diffraction spot patterns is recorded.

The data, which comprise the intensities and positions of the spots, can be read directly off the image plate into the computer. The next step is to convert the many thousands of intensity readings into a 3-D image of the crystal's atomic structure that scientists can understand and interpret. This is no trivial exercise, because, although the readings give us amplitude (that is, the intensity or brightness of the spots), the equipment has not yet been invented that would give us all the information we need concerning the phases of the diffracted beams.

Some cunning techniques invented in the last century make it possible to overcome these limitations. In our case, structural biologist Muhammed Sayed solved the problem by searching the world's databases for a molecule that was similar to our nitrile hydratase and whose structure was already known, which he could use as a template. He was lucky. He found what he was looking for, which meant he could apply a technique called 'molecular replacement' to obtain a first estimate of the missing phases.

With the phases determined, the next stage was to create a computer-generated 3-D electron density map[5], represented on the computer screen by a 3-D mesh of contour lines – which looks like chicken wire – linking points of the same electron density. Because the electron density is centred on the atoms in the molecule, you can locate the positions of each of these atoms.

Seeing our map was an extraordinary and wonderful experience for everyone in the programme. We knew at once that we had generated a picture whose resolution was high enough for us to be able to interpret. This was a triumphant, major landmark of the project.

The huge task that followed, of locating the many thousands of atoms in the protein, fell to post-doctoral researcher Ozlem Tastan-Bishop.

What's visible on the map depends on the resolution. Our map, with its resolution of 0.25 nanometres, meant that we could clearly distinguish individual side chains. This fact, combined with knowledge of the geometry of proteins and the laws of chemistry, made it possible for us to interpret the map using a combination of computer programs, insight, and some very hard work.

An atomic model was built to interpret the map. The model needed to be verified against the original data and also against our knowledge of protein structure based on a database of many thousands of known protein structures. It passed the tests: now we could be confident that our insights from exploring the structure of nitrile hydratase were correct – and potentially useful.

## What next?

The purpose of solving the structure was to develop an industrial enzyme, so we had to answer specific questions. What is the structure in the vicinity of the active site? How does this structure affect the specificity? What factors influence the stability of the enzyme? How susceptible is the enzyme to environmental conditions like pH?

We found, for instance, that our nitrile hydratase has a very unusual active site[6]. But to find out exactly how it works we'll have to understand it better – which, in turn, will help us to improve it and to discover ways to make it even more useful. So although the enzyme is already being used industrially, researchers in Japan and South Africa continue to investigate the details of the mechanism by which this structure converts nitriles to amides. In this way, the Western Cape's successful determination of protein structures is paving the way for new and exciting science, with potentially handsome rewards. □

*Trevor Sewell is extraordinary professor of protein crystallography in UWC's Department of Biotechnology and director of the Electron Microscope Unit at UCT. He is the project leader of the joint Programme in Structural Biology and he co-ordinated the structural research on the nitrile hydratase.*

For more, read Gale Rhodes, *Crystallography Made Crystal Clear* (Academic Press, 2nd ed., 2000) and visit Professor Rhodes's superb web site: www. usm.maine.edu/~rhodes. Other books worth consulting are Carl Branden and John Tooze, *Introduction to Protein Structure* (Garland Science Publishing, 2nd ed., 1999) and David P. Clark and Lonnie D. Russell, *Molecular Biology Made Simple and Fun* (Cache River Press, 1997). For more on transfer RNA visit www.rcsb.org/ pdb/molecules/pdb15_1.html. For more on the ribosome visit www.rcsb.org/pdb/molecules/pdb10_1.html

5 The 3-D map is generated by using a computer to apply a mathematical process called Fourier transformation.

6 In the case of our nitrile hydratase, we found that part of the active site comprises a cobalt ion, held in position by a crown-like arrangement of the sulphurs of three cysteine side chains.

# Q Measuring up

## Listen carefully

Because noise covers a wide range of intensity (loudness), its measurement is based on a logarithmic scale. Decibels (dB) express a relationship between two levels of sound. Librarians get agitated if noise exceeds 30 dB SPL and rockets launch at 180 dB SPL (SPL stands for sound pressure level). The unit is named after the inventor of the telephone, Alexander Graham Bell.

"Decibels have caused untold confusion among audio people," says Florida State University's web site, "and most of this is due to the failure to realize that decibels are not quantities of anything and can represent only power ratios."

## Scanning storms

Sciencedaily.com reports that a new radar instrument, Rapid-Scan Doppler on Wheels (DOW), can assess the threat of tornados much faster than was previously possible. It scans a storm every 5 to 10 seconds, using multiple beams. This is important because tornados change so quickly.

## How do you measure ...

### ... quality of life?

Sciencedaily.com says the 'seventh generation standard' holds that a decision should be judged by its effect on people six generations later. That's difficult to predict in modern times. But the United Nations' Human Development Index (HDI) provides a way of comparing the quality of people's lives in terms of life expectancy at birth, the adult literacy rate, the education enrolment ratio, and gross domestic product per capita. In 2004, the country with the highest HDI ranking was Norway.

### ... change in acceleration?

The unit of change in acceleration is a jerk, equal to 0.3048 $m/s^{-3}$.

### ... the insulating properties of cloth (using the metric system)?

A tog is 10 times the temperature difference (in °C) between the two sides of a piece of cloth, if the flow of heat through the cloth is 1 watt per square metre. A tog is equal to exactly 0.1 $m^2$ K/W.

### ... the height of your audio, video, or computer equipment?

The unit of distance is a U, which is equal to 44.45 mm. For example, a 3U component is 133.35 mm high, and a 20U rack or shelf can house a stack of components 889 mm high.

## UK National Measurement Laboratory definitions of basic units

Second – the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium-133 atom.
Metre – the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.
Kilogram – equal to the mass of the international prototype of the kilogram (made of platinum-iridium and kept at the International Bureau of Weights and Measures in France).
Ampere – that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \times 10^{-7}$ newton per metre of length.
Kelvin – unit of thermodynamic temperature; the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.
Mole – the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of carbon-12.
Candela – the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \times 10^{12}$ hertz and that has a radiant intensity in that direction of 1/683 watt per steradian.