Computers in Pharmacy

Martin Garrison

COMPUTERS IN PHARMACY

COMPUTERS IN PHARMACY

Martin Garrison

Computers in Pharmacy
by Martin Garrison

Copyright© 2022 BIBLIOTEX

www.bibliotex.com

All rights reserved. No part of this book may be reproduced or used in any manner without the prior written permission of the copyright owner, except for the use brief quotations in a book review.

To request permissions, contact the publisher at info@bibliotex.com

Ebook ISBN: 9781984664341

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

1

Computed Tomography

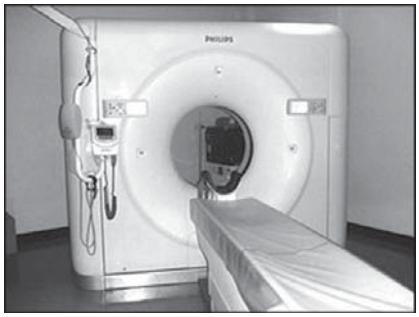


Fig. A Multislice CT Scanner.

Computed tomography (CT), was originally known as "EMI scan" as it was developed at a research branch of EMI, a company best known today for its music and recording business. It was later known as *computed axial tomography* (CAT or CT scan) and *body section roentgenography*.

Computed tomography is a medical imaging method employing tomography where digital geometry processing is used to generate a three-dimensional image of the internals of an object from a large series of two-dimensional X-ray

images taken around a single axis of rotation. The word "tomography" is derived from the Greek *tomos* (slice) and *graphein* (to write).

CT produces a volume of data which can be manipulated, through a process known as *windowing*, in order to demonstrate various structures based on their ability to block the X-ray beam.

Although historically the images generated were in the axial or transverse plane (orthogonal to the long axis of the body), modern scanners allow this volume of data to be reformatted in various planes or even as volumetric (3D) representations of structures.

Although most common in healthcare, CT is also used in other fields, for example nondestructive materials testing. Another example is the DigiMorph project at the University of Texas at Austin which uses a CT scanner to study biological and paleontological specimens.

History

In the early 1930s the Italian radiologist Alessandro Vallebona proposed a method to represent a single slice of the body on the radiographic film. his exam was known as tomography.

The idea is based on simple principles of projective geometry: moving synchronously and in opposite directions the X-ray tube and the film, which are connected together by a rod whose pivot point is the focus; the image created by the points on the focal plane appears sharper, while the images of the other points annihilate as noise.

This is only marginally effective, as blurring occurs only in the "x" plane. There are also more complex devices which

can move in more than one plane and perform more effective blurring.Tomography has been one of the pillars of radiologic diagnostics until the late 1970s, when the availability of minicomputers and of the transverse axial scanning method, this last due to Godfrey Newbold Hounsfield and Allan McLeod Cormack, gradually supplanted it as the modality of CT. The first commercially viable CT scanner was invented by Sir Godfrey Newbold Hounsfield in Hayes, United Kingdom at EMI Central Research Laboratories using X-rays. Hounsfield conceived his idea in 1967, and it was publicly announced in 1972.

Allan McLeod Cormack of Tufts University, Massachusetts, USA independently invented a similar process, and both Hounsfield and Cormack shared the 1979 Nobel Prize in Medicine.The original 1971 prototype took 160 parallel readings through 180 angles, each 1° apart, with each scan taking a little over five minutes. The images from these scans took 2.5 hours to be processed by algebraic reconstruction techniques on a large computer. The scanner had a single photomultiplier detector and operated on the Translate/ Rotate principle.

The CT scanner was "the greatest legacy" of The Beatles, with the massive profits resulting from their record sales enabling EMI to fund scientific research, including into computerised tomography.

The first production X-ray CT machine (in fact called the "EMI-Scanner") was limited to making tomographic sections of the brain, but acquired the image data in about 4 minutes (scanning two adjacent slices) and the computation time (using a Data General Nova minicomputer) was about 7

minutes per picture. This scanner required the use of a water-filled Perspex tank with a pre-shaped rubber "head-cap" at the front, which enclosed the patient's head.

The water-tank was used to reduce the dynamic range of the radiation reaching the detectors (between scanning outside the head compared with scanning through the bone of the skull). The images were relatively low resolution, being composed of a matrix of only 80 x 80 pixels.

The first EMI-Scanner was installed in Atkinson Morley's Hospital in Wimbledon, England, and the first patient brain-scan was made with it in 1972.In the U.S., the first installation was at the Mayo Clinic. As a tribute to the impact of this system on medical imaging the Mayo Clinic has an EMI scanner on display in the Radiology Department.

The first CT system that could make images of any part of the body, and did not require the "water tank" was the ACTA (Automatic Computerized Transverse Axial) scanner designed by Robert S. Ledley, DDS at Georgetown University.

This machine had 30 photomultiplier tubes as detectors and completed a scan in only 9 translate/rotate cycles, much faster than the EMI-scanner. It used a DEC PDP11/34 minicomputer both to operate the servo-mechanisms and to acquire and process the images.

The Pfizer drug company acquired the prototype from the university, along with rights to manufacture it. Pfizer then began making copies of the prototype, calling it the "200FS" (FS meaning Fast Scan), which were selling as fast as they could make them. This unit produced images in a 256x256 matrix, with much better definition than the EMI-Scanner's 80x80.

Previous Studies

Tomography

CT's primary benefit is the ability to separate anatomical structures at different depths within the body. A form of tomography can be performed by moving the X-ray source and detector during an exposure. Anatomy at the target level remains sharp, while structures at different levels are blurred.

By varying the extent and path of motion, a variety of effects can be obtained, with variable depth of field and different degrees of blurring of 'out of plane' structures.Although largely obsolete, conventional tomography is still used in specific situations such as dental imaging (orthopantomography) or in intravenous urography.

Tomosynthesis

Digital tomosynthesis combines digital image capture and processing with simple tube/detector motion as used in conventional radiographic tomography - although there are some similarities to CT, it is a separate technique. In CT, the source/detector makes a complete 360 degree rotation about the subject obtaining a complete set of data from which images may be reconstructed. In digital tomosynthesis, only a small rotation angle (e.g. 40 degrees) with a small number of discrete exposures (e.g. 10) are used.

This incomplete set of data can be digitally processed to yield images similar to conventional tomography with a limited depth of field. However, because the image processing is digital, a series of slices at different depths and with different thicknesses can be reconstructed from the same acquisition, saving both time and radiation exposure.

Because the data acquired is incomplete, tomosynthesis is unable to offer the extremely narrow slice widths that CT offers. However, higher resolution detectors can be used, allowing very-high in-plane resolution, even if the Z-axis resolution is poor. The primary interest in tomosynthesis is in breast imaging, as an extension to mammography, where it may offer better detection rates, with little extra increase in radiation exposure.

Reconstruction algorithms for tomosynthesis are significantly different from conventional CT, as the conventional filtered back projection algorithm requires a complete set of data. Iterative algorithms based upon expectation maximization are most commonly used, but are extremely computationally intensive. Some manufacturers have produced practical systems using commercial GPUs to perform the reconstruction.

Diagnostic use

Since its introduction in the 1970s, CT has become an important tool in medical imaging to supplement X-rays and medical ultrasonography. Although it is still quite expensive, it is the gold standard in the diagnosis of a large number of different disease entities. It has more recently begun to also be used for preventive medicine or screening for disease, for example CT colonography for patients with a high risk of colon cancer.

Although a number of institutions offer full-body scans for the general population, this practice remains controversial due to its lack of proven benefit, cost, radiation exposure, and the risk of finding 'incidental' abnormalities that may trigger additional investigations.

Cranial

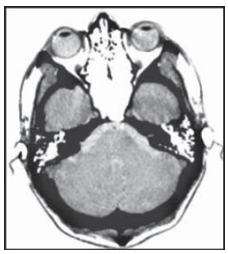


Fig. Normal CT Scan of the Head; this Slice Shows the Cerebellum, a Small Portion of Each Temporal Lobe, the Orbits, and the Ethmoid Sinuses.

Diagnosis of cerebrovascular accidents and intracranial hemorrhage is the most frequent reason for a "head CT" or "CT brain". Scanning is done with or without intravenous contrast agents.

CT generally does not exclude infarct in the acute stage of a stroke. However, CT is useful to exclude an intra-cranial hemorrhage as a cause, or complication of the stroke. For the detection of acute hemorrhage, especially subarachnoid hemorrhage, CT is the test of choice as it is more sensitive than MRI.For detection of tumors, CT scanning with IV contrast is occasionally used but is less sensitive than magnetic resonance imaging (MRI).

CT has an important role in evaluation of the functioning of a ventriculoperitoneal shunt by demonstrating the volume of the ventricular system. Although CT cannot assess intracranial pressure, it can demonstrate several important causes of raised intracranial pressure. Despite the limitation that CT may not detect raised intracranial pressure, it may help in the clinical decision to perform lumbar puncture and is often performed in this context. CT is also useful in the setting of trauma for evaluating facial and skull fractures.In the head/neck/mouth area, CT scanning is used for surgical planning for craniofacial and dentofacial deformities,

evaluation of cysts and some tumors of the jaws/paranasal sinuses/nasal cavity/orbits, diagnosis of the causes of chronic sinusitis, and for planning of dental implant reconstruction.

Chest

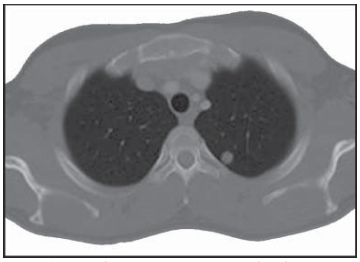


Fig. Chest CT: Axial Slice.

CT is excellent for detecting both acute and chronic changes in the lung parenchyma. A variety of different techniques are used depending on the suspected abnormality. For evaluation of chronic interstitial processes (emphysema, fibrosis, and so forth), thin sections with high spatial frequency reconstructions are used - often scans are performed both in inspiration and expiration. This special technique is called High resolution CT (HRCT).

Note: HRCT is normally done with thin section with skipped areas between the thin sections. Therefore it produces a sampling of the lung and not continuous images. Continuous images are provided in a standard CT of the chest.

For detection of airspace disease (such as pneumonia) or cancer, relatively thick sections and general purpose image reconstruction techniques may be adequate. IV contrast may also be used as it clarifies the anatomy and boundaries of the great vessels and improves assessment of the mediastinum and hilar regions for lymphadenopathy; this is particularly important for accurate assessment of cancer.

CT angiography of the chest is also becoming the primary method for detecting pulmonary embolism (PE) and aortic dissection, and requires accurately timed rapid injections of contrast (Bolus Tracking) and high-speed helical scanners. CT is the standard method of evaluating abnormalities seen on chest X-ray and of following findings of uncertain acute significance.

Pulmonary Angiogram

CT pulmonary angiogram (CTPA) is a medical diagnostic test used to diagnose pulmonary embolism (PE). It employs computed tomography to obtain an image of the pulmonary arteries.

Diagnostic Use

It is a preferred choice of imaging in the diagnosis of PE due to its minimally invasive nature for the patient, whose only requirement for the scan is a cannula (usually a 20G).

Before this test is requested, it is usual for the referring clinician to have carried out a D-dimer blood test and requested a chest X-Ray to rule out any other possible differential diagnosis.

Acquisition

- MDCT (multi detector CT) scanners give the optimum resolution and image quality for this test
- Images are usually taken on a 0.625mm slice thickness, although 2mm is sufficient.
- 50 - 100 mls of contrast is given to the patient at a rate of 4 ml/s.

- The tracker/locator is placed at the level of the Pulmonary Arteries, which sit roughly at the level of the carina.
- Images are acquired with the maximum intensity of radio-opaque contrast in the Pulmonary Arteries. This is done using bolus tracking.

CT machines are now so sophisticated that the test can be done with a patient visit of 5 minutes with an approximate scan time of only 5 seconds or less.

Interpretation

A normal CTPA scan will show the contrast filling the pulmonary vessels, looking bright white. Ideally the aorta should be empty of contrast, to reduce any partial volume artefact which may result in a false positive.

Any mass filling defects, such as an embolus, will appear dark in place of the contrast, filling/ blocking the space where blood should be flowing into the lungs.

Cardiac

With the advent of subsecond rotation combined with multi-slice CT (up to 64-slice), high resolution and high speed can be obtained at the same time, allowing excellent imaging of the coronary arteries (cardiac CT angiography).

Images with an even higher temporal resolution can be formed using retrospective ECG gating. In this technique, each portion of the heart is imaged more than once while an ECG trace is recorded.

The ECG is then used to correlate the CT data with their corresponding phases of cardiac contraction.

Once this correlation is complete, all data that were recorded while the heart was in motion (systole) can be ignored and images can be made from the remaining data that happened to be acquired while the heart was at rest (diastole). In this way, individual frames in a cardiac CT investigation have a better temporal resolution than the shortest tube rotation time.

Because the heart is effectively imaged more than once, cardiac CT angiography results in a relatively high radiation exposure around 12 mSv.

For the sake of comparison, a chest X-ray carries a dose of approximately 0.02 to 0.2 mSv and natural background radiation exposure is around 0.01 mSv/day. Thus, cardiac CTA is equivalent to approximately 100-600 chest X-rays or over 3 years worth of natural background radiation.

Methods are available to decrease this exposure, however, such as prospectively decreasing radiation output based on the concurrently acquired ECG (aka tube current modulation.) This can result in a significant decrease in radiation exposure, at the risk of compromising image quality if there is any arrhythmia during the acquisition.

The significance of radiation doses in the diagnostic imaging range has not been proven, although the possibility of inducing an increased cancer risk across a population is a source of significant concern.

This potential risk must be weighed against the competing risk of not performing a test and potentially not diagnosing a significant health problem such as coronary artery disease. It is uncertain whether this modality will replace invasive coronary catheterization.

Currently, it appears that the greatest utility of cardiac CT lies in ruling out coronary artery disease rather than ruling it in. This is because the test has a high sensitivity (greater than 90 per cent) and thus a negative test result means that a patient is very unlikely to have coronary artery disease and can be worked up for other causes of their chest symptoms.

This is termed a high negative predictive value. A positive result is less conclusive and often will be confirmed (and possibly treated) with subsequent invasive angiography. For the record, the positive predictive value of cardiac CTA is estimated at approximately 82 per cent and the negative predictive value is around 93 per cent.

Dual Source CT scanners, introduced in 2005, allow higher temporal resolution by acquiring a full CT slice in only half a rotation, thus reducing motion blurring at high heart rates and potentially allowing for shorter breath-hold time. This is particularly useful for ill patients who have difficulty holding their breath or who are unable to take heart-rate lowering medication.

The speed advantages of 64-slice MSCT have rapidly established it as the minimum standard for newly installed CT scanners intended for cardiac scanning. Manufacturers are now actively developing 256-slice and true 'volumetric' scanners, primarily for their improved cardiac scanning performance.

The latest MSCT scanners acquire images only at 70-80 per cent of the R-R interval (late diastole). This prospective gating can reduce effective dose from 10-15mSv to as little as 1.2mSv in follow-up patients acquiring at 75 per cent of

the R-R interval. Effective doses at a centre with well trained staff doing coronary imaging can average less than the doses for conventional coronary angiography.

Abdominal and Pelvic

CT is a sensitive method for diagnosis of abdominal diseases. It is used frequently to determine stage of cancer and to follow progress. It is also a useful test to investigate acute abdominal pain.

Renal/urinary stones, appendicitis, pancreatitis, diverticulitis, abdominal aortic aneurysm, and bowel obstruction are conditions that are readily diagnosed and assessed with CT. CT is also the first line for detecting solid organ injury after trauma.

Oral and/or rectal contrast may be used depending on the indications for the scan. A dilute (2 per cent w/v) suspension of barium sulfate is most commonly used. The concentrated barium sulfate preparations used for fluoroscopy e.g. barium enema are too dense and cause severe artifacts on CT.

Iodinated contrast agents may be used if barium is contraindicated (e.g. suspicion of bowel injury). Other agents may be required to optimize the imaging of specific organs: e.g. rectally administered gas (air or carbon dioxide) for a colon study, or oral water for a stomach study.

CT has limited application in the evaluation of the *pelvis*. For the female pelvis in particular, ultrasound and MRI are the imaging modalities of choice. Nevertheless, it may be part of abdominal scanning (e.g. for tumors), and has uses in assessing fractures.

CT is also used in osteoporosis studies and research along side DXA scanning. Both CT and DXA can be used to assess bone mineral density (BMD) which is used to indicate bone strength, however CT results do not correlate exactly with DXA (the gold standard of BMD measurement).

CT is far more expensive, and subjects patients to much higher levels of ionizing radiation, so it is used infrequently.

Extremities

CT is often used to image complex fractures, especially ones around joints, because of its ability to reconstruct the area of interest in multiple planes. Fractures, ligamentous injuries and dislocations can easily be recognised with a 0.2 mm resolution.

Advantages and Hazards

Advantages Over Projection Radiography

First, CT completely eliminates the superimposition of images of structures outside the area of interest. Second, because of the inherent high-contrast resolution of CT, differences between tissues that differ in physical density by less than 1 per cent can be distinguished.

Third, data from a single CT imaging procedure consisting of either multiple contiguous or one helical scan can be viewed as images in the axial, coronal, or sagittal planes, depending on the diagnostic task. This is referred to as multiplanar reformatted imaging.

Radiation Exposure

CT is regarded as a moderate to high radiation diagnostic technique. While technical advances have improved radiation

efficiency, there has been simultaneous pressure to obtain higher-resolution imaging and use more complex scan techniques, both of which require higher doses of radiation.

The improved resolution of CT has permitted the development of new investigations, which may have advantages; e.g. Compared to conventional angiography, CT angiography avoids the invasive insertion of an arterial catheter and guidewire; CT colonography may be as useful as a barium enema for detection of tumors, but may use a lower radiation dose.

The greatly increased availability of CT, together with its value for an increasing number of conditions, has been responsible for a large rise in popularity.

So large has been this rise that, in the most recent comprehensive survey in the UK, CT scans constituted 7 per cent of all radiologic examinations, but contributed 47 per cent of the total collective dose from medical X-ray examinations in 2000/2001.

Increased CT usage has led to an overall rise in the total amount of medical radiation used, despite reductions in other areas.The radiation dose for a particular study depends on multiple factors: volume scanned, patient build, number and type of scan sequences, and desired resolution and image quality. Additionally, two helical CT scanning parameters that can be adjusted easily and that have a profound effect on radiation dose are tube current and pitch.

CT scans of children have been estimated to produce non-negligible increases in the probability of lifetime cancer mortality leading to calls for the use of reduced current settings for CT scans of children.

A 2007 report in the New England Journal of Medicine suggested that the radiation from current CT-scan use may cause as many as 1 in 50 future cases of cancer.According to the USAToday, and members of the American Heart Association, an average CT scan can expose a patient to between 1,000 to 10,000 millirems of radiation, depending on the exact machine and the examination being performed. However, Japanese people who were 1 mile from ground zero received only 3,ooo millirems of radiation, on average.

Adverse Reactions to Contrast Agents

Because CT scans rely on intravenously administered contrast agents in order to provide superior image quality, there is a low but non-negligible level of risk associated with the contrast agents themselves. Certain patients may experience severe and potentially life-threatening allergic reactions to the contrast dye.

The contrast agent may also induce kidney damage. The risk of this is increased with patients who have preexisting renal insufficiency, preexisting diabetes, or reduced intravascular volume.

In general, if a patient has normal kidney function, then the risks of contrast nephropathy are negligible. Patients with mild kidney impairment are usually advised to ensure full hydration for several hours before and after the injection.

For moderate kidney failure, the use of iodinated contrast should be avoided; this may mean using an alternative technique instead of CT e.g. MRI. Perhaps paradoxically, patients with severe renal failure requiring dialysis do not require special precautions, as their kidneys have so little

function remaining that any further damage would not be noticeable and the dialysis will remove the contrast agent.

Process

X-ray slice data is generated using an X-ray source that rotates around the object; X-ray sensors are positioned on the opposite side of the circle from the X-ray source. The earliest sensors were scintillation detectors, with photomultiplier tubes excited by (typically) sodium iodide crystals.

Modern detectors use the ionization principle and are filled with low-pressure Xenon gas. Many data scans are progressively taken as the object is gradually passed through the gantry. They are combined together by the mathematical procedures known as tomographic reconstruction. The data are arranged in a matrix in memory, and each data point is convolved with its neighbours according with a seed algorithm using Fast Fourier Transform techniques.

This dramatically increases the resolution of each Voxel (volume element). Then a process known as Back Projection essentially reverses the acquisition geometry and stores the result in another memory array. This data can then be displayed, photographed, or used as input for further processing, such as multi-planar reconstruction.

Newer machines with faster computer systems and newer software strategies can process not only individual cross sections but continuously changing cross sections as the gantry, with the object to be imaged, is slowly and smoothly slid through the X-ray circle.

These are called *helical* or *spiral CT* machines. Their computer systems integrate the data of the moving individual

slices to generate three dimensional volumetric information (3D-CT scan), in turn viewable from multiple different perspectives on attached CT workstation monitors.

This type of data acquisition requires enormous processing power, as the data are arriving in a continuous stream and must be processed in real-time.In conventional CT machines, an X-ray tube and detector are physically rotated behind a circular shroud; in the electron beam tomography (EBT) the tube is far larger and higher power to support the high temporal resolution. The electron beam is deflected in a hollow funnel shaped vacuum chamber. X-rays are generated when the beam hits the stationary target.

The detector is also stationary. This arrangement can result in very fast scans, but is extremely expensive.The data stream representing the varying radiographic intensity sensed at the detectors on the opposite side of the circle during each sweep is then computer processed to calculate cross-sectional estimations of the radiographic density, expressed in Hounsfield units. Sweeps cover 360 or just over 180 degrees in conventional machines, 220 degrees in EBT.

CT is used in medicine as a diagnostic tool and as a guide for interventional procedures. Sometimes contrast materials such as intravenous iodinated contrast are used. This is useful to highlight structures such as blood vessels that otherwise would be difficult to delineate from their surroundings. Using contrast material can also help to obtain functional information about tissues.Pixels in an image obtained by CT scanning are displayed in terms of relative radiodensity.

The pixel itself is displayed according to the mean attenuation of the tissue(s) that it corresponds to on a scale from -1024 to +3071 on the Hounsfield scale. Pixel is a two dimensional unit based on the matrix size and the field of view. When the CT slice thickness is also factored in, the unit is known as a Voxel, which is a three dimensional unit.

The phenomenon that one part of the detector cannot differ between different tissues is called the *"Partial Volume Effect"*. That means that a big amount of cartilage and a thin layer of compact bone can cause the same attenuation in a voxel as hyperdense cartilage alone.

Water has an attenuation of 0 Hounsfield units (HU) while air is -1000 HU, cancellous bone is typically +400 HU, cranial bone can reach 2000 HU or more (os temporale) and can cause artifacts. The attenuation of metallic implants depends on atomic number of the element used: Titanium usually has an amount of +1000 HU, iron steel can completely extinguish the X-ray and is therefore responsible for well-known line-artifacts in computed tomograms.

Artifacts are caused by abrupt transitions between low- and high-density materials, which results in data values that exceed the dynamic range of the processing electronics.

Windowing

Windowing is the process of using the calculated Hounsfield units to make an image. The display device, as well as the human eye, can only resolve 256 shades of gray. These shades of gray can be distributed over a wide range of HU values to get an overview of structures that attenuate the beam to widely varying degrees.

Alternatively, these shades of gray can be distributed over a narrow range of HU values (called a "narrow window") centered over the average HU value of a particular structure to be evaluated.

In this way, subtle variations in the internal makeup of the structure can be discerned. This is a commonly used image processing technique known as contrast compression. For example, to evaluate the abdomen in order to find subtle masses in the liver, one might use liver windows.

Choosing 70 HU as an average HU value for liver, the shades of gray can be distributed over a narrow window or range. One could use 170 HU as the narrow window, with 85 HU above the 70 HU average value; 85 HU below it. Therefore the liver window would extend from -15 HU to +155 HU.

All the shades of gray for the image would be distributed in this range of Hounsfield values. Any HU value below -15 would be pure black, and any HU value above 155 HU would be pure white in this example.

Using this same logic, bone windows would use a *"wide window"* (to evaluate everything from fat-containing medullary bone that contains the marrow, to the dense cortical bone), and the centre or level would be a value in the hundreds of Hounsfield units.

To an untrained person, these window controls would correspond to the more familiar "Brightness" (Window Level) and "Contrast" (Window Width).

Artifacts

Although CT is a relatively accurate test, it is liable to produce artifacts, such as the following.

- *Aliasing Artifact or Streaks:* These appear as dark lines which radiate away from sharp corners. It occurs because it is impossible for the scanner to 'sample' or take enough projections of the object, which is usually metallic. It can also occur when an insufficient X-ray tube current is selected, and insufficient penetration of the x-ray occurs. These artifacts are also closely tied to motion during a scan. This type of artifact commonly occurs in head images around the pituitary fossa area.
- *Partial Volume Effect:* This appears as 'blurring' over sharp edges. It is due to the scanner being unable to differentiate between a small amount of high-density material (e.g. bone) and a larger amount of lower density (e.g. cartilage). The processor tries to average out the two densities or structures, and information is lost. This can be partially overcome by scanning using thinner slices.
- *Ring Artifact:* Probably the most common mechanical artifact, the image of one or many 'rings' appears within an image. This is usually due to a detector fault.
- *Noise Artifact:* This appears as graining on the image and is caused by a low signal to noise ratio. This occurs more commonly when a thin slice thickness is used. It can also occur when the kV or mA of the X-ray tube is insufficient to penetrate the anatomy.
- *Motion Artifact:* This is seen as blurring and/or streaking which is caused by movement of the object being imaged.

- *Windmill:* Streaking appearances can occur when the detectors intersect the reconstruction plane. This can be reduced with filters or a reduction in pitch.
- *Beam Hardening:* This can give a 'cupped appearance'. It occurs when there is more attenuation in the centre of the object than around the edge. This is easily corrected by filtration and software.

Three Dimensional (3D) Image Reconstruction

The Principle

Because contemporary CT scanners offer isotropic, or near isotropic, resolution, display of images does not need to be restricted to the conventional axial images. Instead, it is possible for a software programme to build a volume by 'stacking' the individual slices one on top of the other. The programme may then display the volume in an alternative manner.

Multiplanar Reconstruction

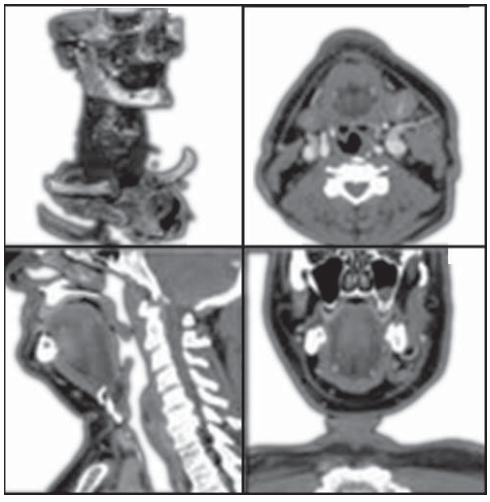


Fig. Typical Screen Layout for Diagnostic Software, Showing One 3D and Three MPR Views

Multiplanar reconstruction (MPR) is the simplest method of reconstruction. A volume is built by stacking the axial slices. The software then cuts slices through the volume in a different plane (usually orthogonal). Optionally, a special projection method, such as maximum-intensity projection (MIP) or minimum-intensity projection (mIP), can be used to build the reconstructed slices.

MPR is frequently used for examining the spine. Axial images through the spine will only show one vertebral body at a time and cannot reliably show the intervertebral discs. By reformatting the volume, it becomes much easier to visualise the position of one vertebral body in relation to the others.

Modern software allows reconstruction in non-orthogonal (oblique) planes so that the optimal plane can be chosen to display an anatomical structure. This may be particularly useful for visualising the structure of the bronchi as these do not lie orthogonal to the direction of the scan.

For vascular imaging, curved-plane reconstruction can be performed. This allows bends in a vessel to be 'straightened' so that the entire length can be visualised on one image, or a short series of images. Once a vessel has been 'straightened' in this way, quantitative measurements of length and cross sectional area can be made, so that surgery or interventional treatment can be planned.

MIP reconstructions enhance areas of high radiodensity, and so are useful for angiographic studies. mIP reconstructions tend to enhance air spaces so are useful for assessing lung structure.

3D Rendering Techniques

Surface Rendering

A threshold value of radiodensity is chosen by the operator (e.g. a level that corresponds to bone). A threshold level is set, using edge detection image processing algorithms. From this, a 3-dimensional model can be constructed and displayed on screen.

Multiple models can be constructed from various different thresholds, allowing different colors to represent each anatomical component such as bone, muscle, and cartilage. However, the interior structure of each element is not visible in this mode of operation.

Volume Rendering

Surface rendering is limited in that it will only display surfaces which meet a threshold density, and will only display the surface that is closest to the imaginary viewer. In volume rendering, transparency and colors are used to allow a better representation of the volume to be shown in a single image - e.g. the bones of the pelvis could be displayed as semi-transparent, so that even at an oblique angle, one part of the image does not conceal another.

3D Rendering Software

Image Segmentation

Where different structures have similar radiodensity, it can become impossible to separate them simply by adjusting volume rendering parameters. The solution is called segmentation, a manual or automatic procedure that can remove the unwanted structures from the image.

Example

Some slices of a cranial CT scan are shown below. The bones are whiter than the surrounding area. (Whiter means higher radiodensity.) Note the blood vessels (arrowed) showing brightly due to the injection of an iodine-based contrast agent.

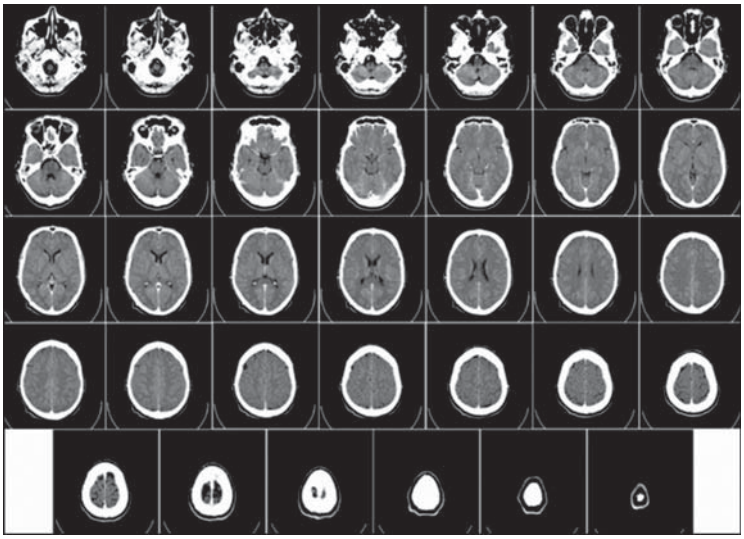


Fig. Computed Tomography of Human Brain, From Base of the Skull to Top. Taken with Intravenous Contrast Medium.

Types of Modern CT acquisition

Scout/Pilot/Topogram

A Scout image is used in planning the exam and to establish where the target organs are located. The beginning and end of the scan are set by the target region and the location of the patient on the table.

Once the Scout image is created it is used to determine the extent of the desired Axial/Helical scan. During the Scout scan the gantry is rotated to a fixed position and the table is translated as x-ray is delivered. The image appears similar to a radiograph.

Axial

In axial "step and shoot" acquisitions each slice/volume is taken and then the table is incremented to the next location. In multislice scanners each location is multiple slices and represents a volume of the patient anatomy. Tomographic reconstruction is used to generate Axial images.

Cine

A cine acquisition is used when the temporal nature is important. This is used in Perfusion applications to evaluate blood flow, blood volume and mean transit time. Cine is a time sequence of axial images. In a Cine acquisition the cradle is stationary and the gantry rotates continuously. Xray is delivered at a specified interval and duration.

Helical/Spiral

Helical is a very fast way to examine the target anatomy. The volume is scanned very quickly because the table is in constant motion as the gantry rotates continuously. There is no interscan delay between slices as in a Axial acquisition.

DRR

A Digitally Reconstructed Radiograph is a simulation of a conventional 2D x-ray image, created from computed tomography (CT) data. A radiograph, or conventional x-ray image, is a single 2D view of total x-ray absorption through the body along a given axis. Two objects (say, bones) in front of one another will overlap in the image.

By contrast, a 3D CT image gives a volumetric representation. (Earlier CT data sets were better thought of as a set of 2D cross sectional images.) Sometimes one must

compare CT data to a classical radiograph, and this can be done by comparing a DRR based on the CT data.

An early example of their use is the beam's eye view (BEV) as used in radiotherapy planning. In this application, a BEV is created for a specific patient and is used to help plan the treatment.DRRs are created by summing CT intensities along a ray from each pixel to the simulated x-ray source.Since 1993, the Visible Human Project (VHP) has made full body CT data available to researchers. This has allowed several universities and commercial companies to try and create DRR's.

These have been suggested as useful for training simulations in Radiology and Diagnostic Radiography. It takes a significant number of calculations to create a summative 2D image from a large amount of 3D data.

This is an area of medical science and education that has benefited from the advancing of graphics card technology, driven by the computer games industry.Another novel use of DRR's is in identification of the dead from old radiographic records, by comparing them to DRR's created from CT data.

Electron Beam CT

Electron beam tomography (EBCT) was introduced in the early 1980s, by medical physicist Andrew Castagnini, as a method of improving the temporal resolution of CT scanners. Because the X-ray source has to rotate by over 180 degrees in order to capture an image the technique is inherently unable to capture dynamic events or movements that are quicker than the rotation time.

Instead of rotating a conventional X-ray tube around the patient, the EBCT machine houses a huge vacuum tube in

which an electron beam is electro-magnetically steered towards an array of tungsten X-ray anodes arranged circularly around the patient. Each anode is hit in turn by the electron beam and emits X-rays that are collimated and detected as in conventional CT.

The lack of moving parts allows very quick scanning, with single slice acquisition in 50-100 ms, making the technique ideal for capturing images of the heart. EBCT has found particular use for assessment of coronary artery calcium, a means of predicting risk of coronary artery disease.

The very high cost of EBCT equipment, and its poor flexibility (EBCT scanners are essentially single-purpose cardiac scanners), has led to poor uptake; fewer than 150 of these scanners have been installed worldwide. EBCT's role in cardiac imaging is rapidly being supplanted by high-speed multi-detector CT, which can achieve near-equivalent temporal resolution with much faster z-axis coverage.

Helical or Spiral CT

Helical, also called spiral, CT was introduced in the early 1990s, with much of the development led by Willi Kalender and Kazuhiro Katada. In older CT scanners, the X-ray source would move in a circular fashion to acquire a single 'slice', once the slice had been completed, the scanner table would move to position the patient for the next slice; meanwhile the X-ray source/detectors would reverse direction to avoid tangling their cables.

In helical CT the X-ray source (and detectors in 3rd generation designs) are attached to a freely rotating gantry. During a scan, the table moves the patient smoothly through the scanner; the name derives from the helical path traced

out by the X-ray beam. It was the development of two technologies that made helical CT practical: slip rings to transfer power and data on and off the rotating gantry, and the switched mode power supply powerful enough to supply the X-ray tube, but small enough to be installed on the gantry.

The major advantage of helical scanning compared to the traditional shoot-and-step approach, is speed; a large volume can be covered in 20-60 seconds. This is advantageous for a number or reasons:

- Often the patient can hold their breath for the entire study, reducing motion artifacts,
- It allows for more optimal use of intravenous contrast enhancement,
- The study is quicker than the equivalent conventional CT permitting the use of higher resolution acquisitions in the same study time.

The data obtained from spiral CT is often well-suited for 3D imaging because of the lack of motion mis-registration and the increased out of plane resolution. These major advantages led to the rapid rise of helical CT as the most popular type of CT technology.

Despite the advantages of helical scanning, there are a few circumstances where it may not be desirable - there is, of course, no difficulty in configuring a helical capable scanner for scanning in shoot-and-step mode. All other factors being equal, helical CT has slightly lower z-axis resolution than step-and-shoot (due to the continual movement of the patient).

Where z-resolution is critical but where it is undesirable to scan at a higher resolution setting (due to the higher

radiation exposure required) e.g. brain imaging, step-and-shoot may still be the preferred method.

Multislice CT

Multislice CT scanners are similar in concept to the helical or spiral CT but there are more than one detector ring. It began with two rings in mid nineties, with a 2 solid state ring model designed and built by Elscint (Haifa) called CT TWIN, with one second rotation (1993): It was followed by other manufacturers.

Later, it was presented 4, 8, 16, 32, 40 and 64 detector rings, with increasing rotation speeds. Current models (2007) have up to 3 rotations per second, and isotropic resolution of 0.35mm voxels with z-axis scan speed of up to 18 cm/s.This resolution exceeds that of High Resolution CT techniques with single-slice scanners, yet it is practical to scan adjacent, or overlapping, slices - however, image noise and radiation exposure significantly limit the use of such resolutions.

The major benefit of multi-slice CT is the increased speed of volume coverage. This allows large volumes to be scanned at the optimal time following intravenous contrast administration; this has particularly benefitted CT angiography techniques - which rely heavily on precise timing to ensure good demonstration of arteries.

Computer power permits increasing the postprocessing capabilities on workstations. Bone suppression, volume rendering in real time, with a natural visualization of internal organs and structures, and automated volume reconstruction really change the way diagnostic is performed on CT studies and this models become true volumetric scanners.

The ability of multi-slice scanners to achieve isotropic resolution even on routine studies means that maximum image quality is not restricted to images in the axial plane - and studies can be freely viewed in any desired plane.

Dual Source CT

Siemens introduced a CT model with dual X-ray tube and dual array of 64 slice detectors, at the 2005 Radiological Society of North America (RSNA) medical meeting. Dual sources increase the temporal resolution by reducing the rotation angle required to acquire a complete image, thus permitting cardiac studies without the use of heart rate lowering medication, as well as permitting imaging of the heart in systole.

The use of two x-ray units makes possible the use of dual energy imaging, which allows an estimate of the average atomic number in a voxel, as well as the total attenutaion. This permits automatic differentiation of calcium (e.g. in bone, or diseased arteries) from iodine (in contrast medium) or titanium (in stents) - which might otherwise be impossible to differentiate. It may also improve the characterization of tissues allowing better tumor differentiation.

256+ Slice CT

At RSNA 2007, Philips announced a 256 slice scanner, while Toshiba announced a “dynamic volume” scanner based on 320 slices. The majority of published data with regard to both technical and clinical aspects of the systems have been related to the prototype unit made by Toshiba Medical Systems. The recent 3 month Beta installation at Johns Hopkins Press Release using a Toshiba system tested the

clinical capabilities of this technology JHU Gazette. The technology currently remains in a development phase but has demonstrated the potential to significantly reduce radiation exposure by eliminating the requirement for a helical examination in both cardiac CT angiography and whole brain perfusion studies for the evaluation of stroke.

Inverse Geometry CT

Inverse geometry CT (IGCT) is a novel concept which is being investigated as refinement of the classic third generation CT design. Although the technique has been demonstrated on a laboratory proof-of-concept device, it remains to be seen whether IGCT is feasible for a practical scanner. IGCT reverses the shapes of the detector and X-ray sources.

The conventional third-generation CT geometry uses a point source of X-rays, which diverge in a fan beam to act on a linear array of detectors. In multidetector computed tomography (MDCT), this is extended in 3 dimensions to a conical beam acting on a 2D array of detectors.

The IGCT concept, conversely, uses an array of highly collimated X-ray sources which act on a point detector. By using a principle similar to electron beam tomography (EBCT), the individual sources can be activated in turn by steering an electron beam onto each source target.

The rationale behind IGCT is that it avoids the disadvantages of the cone-beam geometry of third generation MDCT. As the z-axis width of the cone beam increases, the quantity of scattered radiation reaching the detector also increases, and the z-axis resolution is thereby degraded - because of the increasing z-axis distance that each ray must

traverse. This reversal of roles has extremely high intrinsic resistance to scatter; and, by reducing the number of detectors required per slice, it makes the use of better performing detectors (e.g. ultra-fast photon counting detectors) more practical. Because a separate detector can be used for each 'slice' of sources, the conical geometry can be replaced with an array of fans, permitting z-axis resolution to be preserved.

Synchrotron X-ray Tomographic Microscopy

Synchrotron X-ray tomographic microscopy is a 3-D scanning technique that allows non-invasive high definition scans of objects with details as fine as 1,000th of a millimetre, meaning it has two to three thousand times the resolution of a traditional medical CT scan.

Synchrotron X-ray tomographic microscopy has been applied in the field of palaeontology to perform non-destructive internal examination of fossils, including fossil embryos to be made.

Scientists feel this technology has the potential to revolutionize the field of paleontology. The first team to use the technique have published their findings in Nature, which they believe "could roll back the evolutionary history of arthropods like insects and spiders."

Archaeologists are increasingly turning to Synchrotron X-ray tomographic microscopy as a non-destructive means to examine ancient specimens.

X-ray Tomography

X-ray Tomography is a branch of X-ray microscopy. A series of projection images are used to calculate a three dimensional

reconstruction of an object. The technique has found many applications in materials science and later in biology and biomedical research. In terms of the latter, the National centre for X-ray Tomography (NCXT) is one of the principal developers of this technology, in particular for imaging whole, hydrated cells.

2

Computer Assistant Drug Discovery

Modern drug discovery is a complex, risky, time consuming, and costly process. Despite increased investment in research and development, the numbers of drugs launched has declined in recent years. In the so-called post-genomic era, rapid progress is being made in structural genomics, functional genomics, proteomics, and phamacokinetics and drug metabolism.

This has provided information about protein structure and function, cellular profiles of proteins, molecular distributions, and metabolism. The volume of public sequence databases, target databases, and compound database are increasing exponentially these days.

Drug discovery process, therefore, is steadily becoming more information driven and a data-centric problem where new drug discovery is based on analysis and data mining to unveil the information hidden behind the large genomic, proteomic, and small molecular databases.

To improve productivity, knowledgebase-guided decisions must be incorporated into the drug discovery and development process. Computational methodologies have become a crucial component of many drug discovery processes, from hit identification to lead optimization and beyond, and approaches such as structure- or ligand-based virtual screening techniques and information-based method are widely used in many drug discovery efforts.

It is well know that relational databases have become critically important in business applications, but they have played a relatively minor role in drug discovery and development process, which has generally been concerned with modeling and simulation activities.

In recent years, massively parallel database architectures are beginning to offer the ability to quickly search through terabytes of data with hundred-fold or even thousand-fold speedup over server-based architectures, and the Netezza Performance Server system, a massively parallel database machine originally designed for such analytic searches, provides a powerful tool that integrates multiple computational algorithms for meeting the data mining needs of biologists, chemists, and pharmacologists to speed up the new drug discovery. In this paper, the various uses of Netezza in the drug discovery process have been investigated.

THE NETEZZA ARCHITECTURE

The Netezza Performance Server (NPS) system's architecture is a two-tiered system designed to handle very large queries from multiple users, which consists of a closely coupled server with many parallel Snippet Processing Units,

each with their own disk and streaming database logic chip to perform fast pattern matching.

The second tier consists of dozens to hundreds or thousands of Snippet Processing Units (SPUs) operating in parallel which are connected by Gigabit Ethernet to both the host server and to the other SPUs.

Each SPU is an intelligent query processing and storage node, and consists of a powerful commodity processor, dedicated memory, a disk drive and a fieldprogrammable disk controller with hard-wired logic to manage data flows and process queries at the disk level. The massively parallel, shared-nothing SPU blades provide the performance advantages of massively parallel processors.

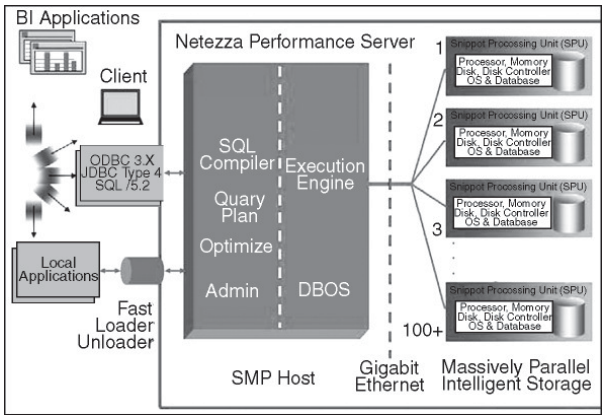


Fig. The Netezza Performance Server system

Nearly all query processing is done at the SPU level, with each SPU operating on its portion of the database. All operations that easily lend themselves to parallel processing (including record operations, parsing, filtering, projecting, interlocking and logging) are performed by the SPU nodes, which significantly reduces the amount of data moved within the system. Intelligent Query Streaming is performed on each SPU by a Field-Programmable Gate Array (FPGA) chip that

functions as the disk controller, but which is also capable of basic processing as data is read from the disk.

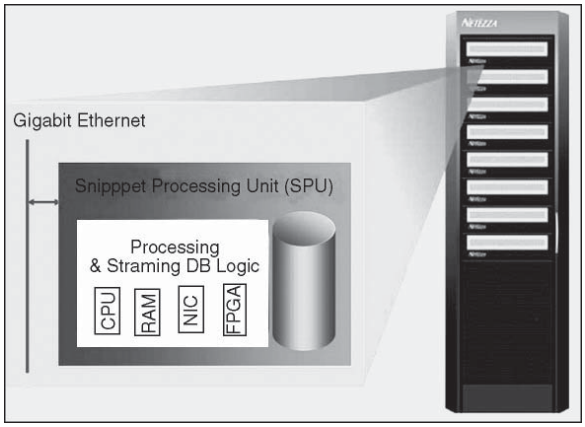


Fig. Snippet Processing Units

The system is able to run critical database query functions such as parsing, filtering and projecting at full disk reading speed, while maintaining full ACID (Atomicity, Consistency, Isolation, and Durability) transactional operations of the database.

To achieve high performance, the storage interconnection, which is a bottleneck with traditional systems, is eliminated by directly attaching the disks so that data can stream straight into the FPGA for initial query filtering.

Then, to further reduce the workload on the central server, the intermediate query tasks are performed in parallel on the SPUs.

APPLICATIONS IN COMPUTER-ASSISTANT DRUG DISCOVERY

Computational methodologies have become an integrate and crucial component of many drug discovery processes. Netezza provides a powerful tool that integrates multiple computational algorithms for meeting the data mining needs

of drug discovery process and could make new drug discovery faster, cheaper and smarter.

Structure-based Drug Discovery

Structure-based virtual screening (VS) method, for drug discovery is typically carried out by computationally docking a large amount of compounds into the active site of protein target. The completion of the Human Genome Project and recent advances in structural genomics and proteomics have identified a large number of human proteins as drug targets, which were estimated about 5000.

However, the explosion in the protein target available is yet to create a commensurate increase in the efficiency of the drug discovery process. The success of structure-based method strongly depends on the amount and quality of information available about the system under investigation.

It can be argued that obtaining protein target is no longer the key issue facing early drug discovery, and that challenge has now becomes using proper tools and suitable methods to represent, store, retrieve, and analyse the protein targets and apply the information to select right target to screening.

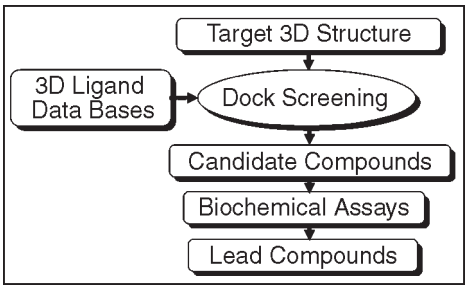


Fig. Structure-based Drug Discovery

Furthermore, during the past few years, a great deal of effort has gone into the development of computational methods for filtering screening databases. More and more

people are skeptical of the need to engage in ultra-VS, where many hundreds of thousands of compounds are screened.

Using Netezza's massively parallel database architecture, we can filter the databases and screen a small but highly diverse collection, such as the compounds that comprise drug-like synthetics, natural products, and FDA-approved drugs.

Information-based Drug Discovery

Since 1990, the United States National Cancer Institute (NCI) has conducted an anticancer drug discovery programme in which approximately 10 000 compounds are screened every year in vitro against a panel of 60 human cancer cell lines from different organs.

Available are screening results of compounds that are not covered by a confidentiality agreement, and the compound list is updated at least once a year , which provides us with a valuable source for computer-based virtual screening of anticancer drugs using a bioinformatics-based approach.

A number of studies have shown that although growth inhibitory activity for a single cell line is not informative, the activity patterns across the 60 cell lines provide incisive information on the mechanism of action of screened compounds and also on molecular targets and modulators within the cancer cells. Several algorithms have been introduced to use the activity information for discovery of anticancer drugs and for understanding of the molecular pharmacology of cancer, which have proven to be very useful for finding agents with activity patterns similar to that of a "seed" compound and for finding compounds with activity patterns that correlate well across the 60 cell lines with the

expression levels of particular molecular targets. An "information-intensive" approach has been developed to use this anticancer database for studies of molecular pharmacology of cancer and for the identification of potential protein targets of an anticancer drug. The application of Netezza would be of great help in speeding up this "informationintensive" approach.

Ligand-based Drug Discovery

Ligand-based approach for drug discovery begins typically with a collection of molecules known to bind to a set of related target. This collection of compounds is then used to perform similarity searching, pharmocophore searches or property profiling against one or more databases which might contain several million chemicals.

In recent years, the number of chemicals generated by traditional and contemporary approaches has increased dramatically. In principle, there could be as many as 1047 quadrillion chemicals that can be made to interact with human protein targets. Although screening methods and scoring contribute to a successful screen, no factor has a larger role than the compounds used for the screen.

It is being recognized that increasing the quality of screening databases, rather than their quantity, is likely to be an important determinant for the identification of active compounds that have a chance to make it through the drug discovery pipeline.

There is a current trend to 're-rationalize' drug discovery research, that is, departing from a mere 'numbers game' and carrying out fewer, but 'smarter' screening. Factors such as physical properties, target classes and 'drug-likeness' are all

important considerations, and computational approaches using Netazza can be valuable in addressing several of these issues. During the past few years, a great deal of effort has gone into the development of computational methods for filtering screening databases.

A large number of drug targets, drug candidates, and a paucity of suitable computer tools to represent, store, retrieve, and analyse these information in the drug discovery process have created a 'target-rich and lead-poor' imbalance. The application of Netezza could help resolve the imbalance between target-rich and lead-poor since cheminformatics methods can be applied to extract knowledge from large-scale molecule databases using Netezza in a shorter time periods in order to assure that good properties are achieved before screening.

3

Pharmaceuticals Development

Speed and productivity are mission critical in Pharmaceutical Development. Long-term survival and success in the drug development space will go to those players who figure out how to get promising new products to market quickly while simultaneously controlling development costs. Critical chain provides pieces to the solution not found in traditional project management.

Focus

In late-stage development, teams have literally hundreds of tasks going on in parallel. The problem is that only a few are driving the filing date. Traditional milestone-based approaches obscure work and hide priorities.

Critical chain and PPM fix that. We have seen tremendous gains in speed and productivity come from a combination of the PPM methodology with its emphasis on practicing focused work, and our software's ability to identify key tasks and resource areas.

Decision Support

In order to file sooner with the FDA, there are hundreds of potential actions the team can take that might possibly result in acceleration. But which ones should be followed? The trick is in knowing where the leverage is—what changes are really going to make a difference. By applying PPM teams know which handful of actions is truly leveraged.

Improved Communication

When it comes to managing the project pipeline, we have found that Pharma companies often have a tremendous amount of data but little real information. This results in a lack of both vertical alignment and horizontal synchronization regarding the true status of a project and what schedule outcomes are most likely.

Despite all the data, senior leadership and functional lines are often not as well informed as they need to be with respect to the true status of projects. PPM and ProChain Enterprise introduce an unprecedented level of transparency and visibility to the organization.

Challenges in Neurotherapeutic

INTRODUCTION

The pharmaceutical industry must generate novel, effective, and safe therapeutics that address unmet medical needs at a cost that is palatable to consumers globally, in a time frame that allows effective recovery of the investments in the systems and processes necessary to generate new products, and in a manner compliant with international regulations.

In the long run, return on investments in therapeutics development must pay for the full economic costs of development or the process will fail economically and cease. The seriousness of this challenge can be inferred from the steady attrition of pharmaceutical companies large and small, old and new.

The economic component grows steadily each year because of increasing pricing pressures, costs, generic competition, and regulatory hurdles, along with decreasing productivity. The scientific challenges for therapeutics developers are great in every area but particularly so in neurotherapeutics.

On one level, new scientific discoveries and tools create the possibility for a golden age of therapeutics development, in which many long-term scourges of humankind may be addressed in a fundamental and powerful manner. Yet while we move from symptomatic treatments to mechanistically targeted cures, we struggle to base development on fields of knowledge that are rapidly changing and far from complete and in which our understanding of the interactions between factors is fragmentary at best.

Neurotherapeutics developers face unusually high failure rates and huge attendant expenses that must be covered by the small number of therapeutics that succeed. The combination of pricing constraints and trends of increasing regulatory hurdles, falling productivity, lengthening cycle times, and falling success rates is not sustainable in the long term.

Many developers have focused on strategies to improve scientific and operational performance, such as improved collaboration with academicians and regulatory agencies,

better use of technologies to explore study data and design trials, greater operational effectiveness in selecting and training quality sites, monitoring and harvesting data, producing analyses, and amortization of costs over a wider base by global registration.

Although important, these approaches are not enough. Issues of discovery and clinical development strategy must also be addressed. These issues are particularly pertinent to neurotherapeutics development and even more so to the development of therapeutics for neurodegenerative disorders.

A new and better paradigm must be developed for pharma to succeed economically and medically in the long term. While incremental improvement of current approaches can be useful and necessary, a more powerful and deeply penetrating change is needed to meet the challenges the therapeutics development community faces.

In the last century, the pharmaceutical industry progressed from a strategy of observation and serendipity to the current strategy of rational design, which argues that mechanistic understanding of disease leads to mechanistically targeted molecules, which works if the theory of disease etiology is correct. The conceptual weakness of this strategy is that the map is not the territory.

The systems biology underlying disease is complex, the relevant mechanisms are often multiple, interactive, and variable across species, the individual variability in an outbred human population is great, and the statistical noise generated by imperfect clinical assessment instruments together are a huge barrier to overcome. Narrowly targeted molecular interventions based on simplistic mental maps of

the disorder sometimes work but often run afoul of these complicating factors.

The next strategic approach, Biodesign, potentially surpasses the performance of the rational design strategy by embracing complexity by using the capacity of the immune system to rapidly generate the myriad potential probes needed to assess the therapeutic potential of different interventions in conjunction with *in vivo*assessment technologies targeted to the human disease model and integrated with computerized disease models.

Elements of this proposed paradigm include the use of disease modeling to select better targets and potential therapies, along with information-science-based approaches to enhance small molecule chemistry, exploitation of the potential for biological technologies to rapidly generate mechanistic probes, and development of different strategies for using animal models and enhanced strategies and abilities to study human molecules mechanistically and biologically.

The synergistic use of these approaches will change the overall business model of companies engaged in the development of neurotherapeutics as well as other therapeutic areas. The next-generation pharmaceutical company will be defined by intense, deep, and advanced focus on specific target diseases, working with defined populations and specialists to target disease mechanisms with ever-broadening power.

The model will combine expertise in protein therapeutics and small molecules and use them interactively in devising commercially viable therapeutics. This development will both depend on and generate a new set of relationships and

interdependencies between academicians, regulators, and industry. In so doing, the Biodesign paradigm may revolutionize medicine and solve the economic challenges facing healthcare systems and therapeutics developers alike.

SCIENTIFIC CHALLENGES FACING THERAPEUTICS

Genomics

The mapping of the human genome offers a startling opportunity for the development of new generations of therapies in the long term. All of the understood therapies currently in use derive from a relatively small set of targets, about 500. With the mapping of approximately 25,000 to 50,000 new targets and the recognition that there may be as many as 300,000 (or more) proteins of interest, the number of targets that potentially can be explored has expanded dramatically.

The promise is for fundamentally targeted treatments, likely coupled with improved diagnostics to exactly determine the target defect in a given individual. This is very exciting, but the promise may be long in coming when one considers some of the challenges. Relatively little is known about most of these genes. This may be helped by the NIH Protein Structure Initiative as it identifies 10,000 protein structures over the coming decade. Similarly, the vast computational power of new supercomputers may speed progress in understanding the structure-function properties of proteins and thereby in facilitating the targeting of drugs to proteins.

There is an important caveat about even "simple" genetic abnormalities as mechanistic targets. Even a disease such

as cystic fibrosis that is related to a single gene product has more than 1,000 known mutations, most of them private (related to a single family) and a few more common mutations that cause the disease in a larger number of individuals; this is unsurprising.

A given molecule may be dysfunctional because it is never made or has a different Michaelis constant (Km) well above the natural concentration of its molecular substrate, or has a maximum velocity (Vmax) well below normal, or has a normal Km and Vmax but is unusually subject to proteolysis and has a short half-life, or aggregates with other proteins differently, or is subject to some other disruptive factor.

The key is to realise that each of these very different problems may not be solved by the same intervention. One dysfunctional molecule may be best remediated by an agent that stabilizes it, while another may need an allosteric modification to change its affinity for substrate. Similarly, in thalassemias patients with seemingly identical genotypes may differ greatly in severity.

The protein deficiencies responsible for the hemoglobinopathies have been well-characterized for some years, yet the curative therapies are still some time away. Genomics opens up the possibility of more effective therapeutics but for many conditions the needed interventions may in fact require significant individualization.

The development of compounds targeting individualized therapies is likely to be highly un-economic, particularly if each such compound is required to meet current regulatory standards of proof, including the requirement for study in large numbers of patients. This challenge is even greater

when it is noted that most of the large-market nervous system diseases are thought to be polygenic. Additionally, the lack of complete concordance between twins in studies of important nervous system diseases such as schizophrenia, depression, and multiple sclerosis highlights the importance of environmental factors.

In the long run, genomics is of great importance in providing targets and a basis for a better understanding of which targets matter, but the complexity of the challenge is sufficient that the long run may too long for some developers to last.

Specificity

The special challenges of neurotherapeutics development begin with the unusually complicated, anatomically specific, and temporally intense nature of the functioning of the target organ: the brain. Although all organ functions, from heart and liver to kidney and lung, have significant inherent complexity, the complexity of the brain is particularly immense.

In the midst of normal functioning, neurons with patterns of connectivity related to development but also affected by antecedent experience interact via intimate connections and communicate signals based on the fine timing and code of the firing. The excitation of any one neuron is typically the result of convergent stimulation from other neurons and in turn can affect multiple other neurons.

Signals flow simultaneously to excitatory and inhibitory recipients to then be transmitted back to the source neuron as a regulatory signal or to affect the firing of other neurons

in parallel or other pathways. Moreover, the system self-regulates over time to compensate for changes in input.

Because the bursts of stimulation leading to neuronal firing are brief, while pharmaceutical levels typically vary slowly over hours, there is no exogenous chemical delivery system that could directly excite and inhibit specific targeted neurons in the timing pattern to allow direct mimicking of neuronal function. There is no exogenous chemical targeting system at this time that can directly stimulate or blockade only a highly specific cluster of neurons while avoiding effects on other neurons of the same class where the same signal might have a different and unwanted effect.

Instead, there are two basic options: to essentially raise or lower the general signal strength in a given system by the application of agonists or antagonists and allow the fine pattern of signal transmission that was otherwise present to continue in an amplified or subdued manner, or to enhance or retard the release, removal, or degradation of an endogenous transmitter to retain its pattern of release while amplifying or diminishing its signal.

In principle, combinations of invasive mechanical and chemical delivery systems might accomplish greater anatomical specificity but would still act at a gross level. In some cases, the nature of the system or the specificity of the receptors is such that a relatively specific influence over a more general system is possible. For the most part, however, pharmaceuticals circulating to the brain have the potential to produce widespread effects across the brain and the body rather than only focusing the desired action on a particular nucleus or nuclei.

As a further complication, the functioning of many neuronal systems and the pathology in many neurologic diseases involve complex systems influencing more than one neurotransmitter. The implication is that highly specific effector molecules may influence one set of neurons that are part of a disease process but in their very specificity fail to affect other systems that are also a part of the same disease process.

The normal course of discovery assessment of molecules generally proceeds on the principle of creating the most selective molecule—agonist or antagonist—possible. Modification of a given brain system by a specific effector molecule produces potential adverse events by stimulating the same target in other brain regions.

For example, the blockade of dopamine neurons by dopamine antagonists that produces beneficial effects on schizophrenia by affecting the mesolimbic dopaminergic systems also produces adverse effects on prolactin levels by blocking tuberoinfundibular dopaminergic neurons.

But because more than one system may be relevant to a given disease process, the very specificity designed into the molecule by discovery groups also precludes the possibility of affecting those other neuronal systems relevant to the disease. In other words, the effort to create a molecularly clean and specific "bullet" is likely to result in decreased efficacy in a situation where more than one neurotransmitter system is involved in the pathology.

It is perhaps not a coincidence that clozapine, the prototype atypical neuroleptic believed by many to be the most effective agent for treatment-resistant schizophrenia has potent

anticholinergic, antiadrenergic, antihistaminic, and antiserotonergic properties.

In principle, one could develop a series of specific effector molecules to mix and match in polypharmacy to optimally treat a particular disease or to optimize treatment for an individual; such a system could allow optimization of the exact amount of blockage or facilitation of each relevant system in a given individual to optimize response and reduce adverse events. The reality, however, is that to be registered for therapeutic use, each molecule would have to be approved independently as a therapeutic.

Because the more specific molecules are often likely to be at a therapeutic disadvantage to less specific molecules already on the market that serendipitously affect a better combination of targets, these approvals would be very difficult to obtain. Hence the option of producing patient-specific or disease-specific cocktails that combine the right set of agonists and antagonists seems unlikely.

Indeed, by the time all the approvals had been obtained on all the relevant cocktail components, the patent life would likely have expired on some. For small molecules at this time the field instead seems largely stuck on developing molecules that affect multiple systems in some relative ratio fixed in the design of the molecule, in the manner of the design development of the various atypical neuroleptics.

Another aspect of specificity relates to the ability of discovery groups to produce molecules that effectively interact with new genomic targets. In current discovery functioning, as a target is identified the company library of compounds—hundreds of thousands to millions of compounds—is

screened against the target. Compounds that assay well are studied as possible lead structures and the most promising are modified and tested, then further modified iteratively until optimum compounds are developed.

Such compound libraries historically have been generated by prior programs and as such tend to be heavily weighted toward compounds hitting G-protein-coupled receptors, serine proteases, and similar targets and may not always have the structural diversity and range of structural complexities that are optimal to meet entirely new classes of targets uncovered by new genomic approaches.

Improving methods of target-oriented syntheses to find molecules hitting preselected protein targets or diversity-oriented syntheses to find molecules modulating particular pathways without regard to a predefined protein target should enhance the ability of chemists to generate compounds relevant to new mechanisms, but the challenge of library diversity remains real. Additionally, rational methods of matching protein structure to molecular design are based on crystallized proteins that are likely to have a different shape from their configuration *in vivo* and hence are likely to be misleading.

Animal models

Modern pharmaceutical discovery is heavily dependent on various disease models yet serious fundamental questions can be raised about the adequacy of these models in general. A great deal of work has gone into characterizing many different animal models for a variety of CNS diseases in terms of understanding their face validity, predictive validity, the phenomenology of abnormalities observed in the model, and

construct validity, among other qualities; yet as Horrobin has noted, for any animal model to be relevant to the development of treatment for a disease, at least three logical conditions should be met, but they rarely are.

First, one should understand the nature of the animal model in detail (i.e., understanding comprehensively why the pathology of the animal model occurs and the mechanisms and systems biology). Second, one should understand the nature of the human disease in detail (i.e., why the pathology of the human disease occurs and the mechanisms and systems biology).

Third, one should know that the animal model and the target disease in humans are congruent in all important respects, because if the animal model and the target human disease are not congruent in some important aspect, it is very likely that predictions made with the animal model will be wrong. In some cases (e.g., neurotrauma or some stroke models), a plausible case can be made for reasonable congruity between model and human disease.

Head trauma models can mimic the process that induces lesions in humans within the limitations of cross-species differences in reaction to trauma, and various methods of stroke induction, from ligation to introduction of clots or microspheres, can induce lesions similar to those experienced by human stroke patients.

Even in such cases, however, there are still differences in circulation, species-related responses, the size of the organ affected, and the time course of lesion induction that may be relevant to the appropriateness of the model. In most neurologic diseases and almost all psychiatric disorders, the

gap between animal model and human disease is great and the animal models used in neurotherapeutics development do not come close to meeting the three conditions.

Consider the following examples. First, studies of the biologic effects of molecules focus on effects on cells in cell culture, cells with a very different relation to nutritional systems and to their exogenous environment, and cells with a generally different phospholipid composition in their membranes than cells *in vivo*. Yet the phospholipid composition of a membrane can easily affect the shape and hence the binding characteristics of molecules embedded within them and the relations to nutrients and the environment are rather different than in *in vivo* cells with a blood-brain barrier.

Further, *in vitro* models typically remove the relevant region from potentially important regulatory controls. Second, genetic animal models of disease (for example, animals with gene deletions) are often posited as indicators, allowing a more comprehensive understanding of the relevance of a system to a given disorder. Yet when the details are examined, a given gene deletion may have markedly different effects in different strains of mice

If the deletion in a given mouse cannot universally predict the outcome in all other strains of mice, how likely is it to predict the outcome in a human? This is not to dismiss the remarkable science and potential utility of transgenic mice, but rather to suggest that the biologic complexity is such that the congruence of animal and human models is difficult to establish. However useful, these models are maps with potentially important differences from the territory of the human disease.

Third, regulation and metabolism may vary across species even if certain pathways are preserved. Fundamental problems with animal models occur across all therapeutic areas but the very complexity of the CNS and the size and behavioral uniqueness of the human brain make the problem of congruity loom large for neurotherapeutics developers. In this sense, animal models for most human CNS diseases at best can be interpreted as having suggestive value, indicating that a potential hypothesis should be tested in humans.

The clear caveats are that in neurotherapeutics development the distance between model and man is large and that human data are far more likely to be of relevance and use than data from even multiple animal models. A further source of incongruity arises from the fact that interventions in the animal models are often used in a manner very different from the human model.

In terms of face validity of the model, stroke research could be considered a good area for extrapolating from animal data to humans. A stroke intervention may be tested in an animal using an agent administered before stroke induction, but is then tested in a human after a stroke occurs. A stroke intervention in animals may be evaluated by rapid inspection of the volume of brain infarcted, then tested in man by using a behavioral rating scale at a very different assessment time.

Important and practical reasons are always cited for why the animal assessment has to be done in a different way than the human disease assessment but the reality is that these differences rarely benefit the predictive validity of the model. In the case where the animal model does potentially have clear significance to the human disorder, investment of

effort in developing the model to allow testing similar to the planned timing in humans is appropriate.

Many of the most important neurotherapeutic challenges, such as depression, psychosis, and dementia, derive from disorders with relatively subtle pathology that are detectable only by detailed neuropsychiatric evaluation. Conversely, many of the animal models used to assess compounds for neuropsychiatric disorders have at best a passing plausibility in reflecting on a given disease and no solid mechanistic foundation to suggest congruity.

Models that have shown an ability to detect an effect generated by a molecule known to be clinically effective in treating a disease may well be used as screens for subsequent potential therapeutic molecules even if the connection between the effect detected and the mechanisms causing the disease is unproven. The flaw is that the effector molecule may produce the effect in the model for a reason different from the mechanism that actually was responsible for its therapeutic effect in humans.

The animal model may be a model for detecting the molecular perturbation caused by the class of compounds in which it is initially validated, but that utility says nothing about its ability to detect the utility of a class of molecules operating by a novel mechanism. Looking outside the nervous system to progress in diseases such as asthma, what initially was conceived as a single syndrome is slowly evolving into a complex cluster of diseases with a variety of underlying genetic susceptibilities, molecular mechanisms, and environmental triggers that are leading to a focus on distinct underlying disease states such as atopy and airway inflammation.

One could easily foresee a similar pathway for a syndrome such as depression or schizophrenia, with subsequent fractionation of the population into distinct causal subcategories. In such cases, animal models that are genuinely mechanistically relevant to the disease of interest could then be developed.

Trial Conduct

The challenges in the conduct of neurotherapeutic trials are similar to the challenges experienced in trials in other specialties. Clinical trials themselves are in general more similar as project management exercises, but neurotherapeutic trials do present some unusually intense challenges compared with other areas.

First, extra science is needed with neurotherapeutics to have confidence that the drug is being tested in adequate doses in the target organ. With most therapeutic trials the target organ is in free communication with the circulation and the free concentrations of the molecule generally reach the target organ. In neurotherapeutic trials, however, the blood-brain barrier acts as a distorting influence capable of transporting or blocking a particular therapeutic.

If arterial concentrations are sustained for a sufficiently long period, the central compartment should come into equilibrium with the free unbound plasma concentration, but a variety of sink conditions maintained by pH gradients across cell membranes, metabolism, and active transport, among others, can prevent it. Neurotherapeutics developers must remember that while peripheral compartment pharmacokinetic and pharmacodynamic parameters are useful in understanding adverse events, the time course and

extent of exposure of the central compartment are the main areas of interest for efficacy.

Animal models of blood-brain barrier effectiveness can be useful in assessing the extent of this effect, but vary in their predictions and are typically not done in a manner that allows understanding of the relationship between the concentrations of free compound in the peripheral compartment and the simultaneous concentrations in the central compartment.

Second, the assessment of treatment effect can be particularly complicated in neurotherapeutic trials. The brain is complex. An intervention may have different effects on multiple particular brain regions, reflected in differential impact on *N*different dimensions of brain function. For example, a dopamine agonist may produce psychiatric effects by its mesolimbic activity, but may also increase prolactin secretion and produce galactorrhea by its tuberoinfundibular blocking activity and motor effects by its activity on the basal ganglia.

More subtly, a D2 superfamily dopamine antagonist may differentially bind D2, D3, and D4 receptors centered in different brain regions, producing complex interacting cognitive and affective effects. An assessment instrument such as a rating scale may have more items that measure one aspect of brain functioning than another or may weight them differently, and different rating instruments for the same disease may also give more or less weight to one aspect of brain function than another.

Hence each treatment has *N*-dimensional effects and each assay instrument potentially samples some of those dimensions better than others. Consequently the assay

instrument that best detects the effects of one compound may not be the best one to detect the effects of another compound affecting the disease in a different way. Thus early depression rating scales developed around optimizing assessment of tricyclic antidepressants may not be optimal for assessing the effects of a novel treatment, even if that treatment is more effective on critical aspects of the disease.

In understanding a molecule, this phenomenon points out the importance of using a range of assessment instruments during early development to better understand the parameters of function influenced by the therapeutic agent. A skilled focus on exploratory data analysis evaluating where the various instruments agreed and where they detected differences may help understand how the compound affects the disease and what assessment instruments best detect the compound's effects.

As a practical matter, for a novel therapeutic it is desirable to have an efficiently executed trial with near real-time availability of data from sophisticated assessments to facilitate ongoing exploratory analysis of the effects of the compound. A trial design for non-pivotal trials that is double-blind (patient and investigator blinded) but not triple-blind (company and academic advisors able to conduct unblinded exploratory analysis) might optimally allow prolonged thought, analysis, and discussion to assist downstream design of subsequent trials without adversely affecting the critical path timelines that are essential to submission and eventual retention of patent life.

Third, determination of dose, formulation, and regimen is particularly challenging in neurotherapeutic trials. In anti-

infective trials, the general level needed to suppress or kill a particular organism *in vivo* may closely mimic the relevant concentrations from*in vitro* studies. In neurotherapeutic trials, however, in addition to the aforementioned concerns with penetration to the target organ, assessment of effect is also problematic.

Assessments are done at a particular time and often are made with instruments that are statistically highly variable because of circadian factors, the interaction between the investigator and patient, and the fact that compound effects on the brain may have a different time course than serum concentrations.Finding the optimal dose for a compound is very difficult. Recent developments in statistical design, combined with the availability of central randomization controls create the possibility of improving this problem by using adaptive designs in trials.

Normally in dose-ranging trials some number of doses and a placebo are allocated across a target number of patients either in equal proportion or in some fixed ratio. Modern technological developments combined with near real-time databases allow for an alternative approach. A trial could be conducted with some fraction of the total number of patients being allocated to each arm.

Those patients can then be rapidly assessed for effectiveness and/or safety. If the target is effectiveness, for example, after the initial cohort is treated, some relative efficacies may be noted in the various arms. These effects may be real or artifacts seen in small samples because of significant random variability. Nonetheless, the randomization schema could then be modified to have a

higher probability of assigning patients to arms that are showing an effect and a decreased but nonzero probability of assigning patients to arms that show futility.

As the trial proceeds, arms that initially look positive would be subjected to more stringent testing by preferentially having more patients assigned to them. If the compound is genuinely efficacious, these arms will likely hold up and show a more robust and well-tested effect. If these arms appeared more positive purely by chance, the additional number of test subjects would likely dilute the chance effect and show it as spurious.

As ineffective arms reached the point of futility, i.e., having no significant chance of showing an effect that would reach significance, they would be closed or have fewer patients assigned to them, allowing the other patients to be assigned to arms that have a better chance of detecting an effect. In some cases such an approach could be completely automated and blinded and likely fully acceptable for pivotal registration trials. More complex modifications may require guidance by a monitor and could potentially be useful, but unblinding and human intervention could also lead to regulatory rejection as biased.

The regulatory and statistical risks of automated or manually assisted adaptive randomization clearly require careful study-by-study consideration, but the benefits—in time, money, and reduction of risk to human subjects—are potentially great.

The Biodesign Strategy

A mix of innovative approaches being tried in drug development as separate pieces have the potential to act

synergistically to resolve or ameliorate some of the economic and scientific challenges discussed above. Considered together, a new strategy, Biodesign, has potential to allow us to progress past the limitations of the current rational design strategy centered entirely on small molecules.

The term Biodesign does not equally apply to all components of this new strategy but it seems apt nonetheless in that the focus is on using probes generated biologically to assess human disease. Four components of this strategy will be discussed separately, then the synergies of their combination will be considered along with an assessment of the major impact this model may have on the future of the therapeutics development partnership between academicians, industry, and regulators.

DISEASE MODELS

In biologic reality, every pathological process is the physical result of a dysfunction or of multiple interactive dysfunctions of one or more processes. Diseases are thought initially to be the result of one dysfunctional system, which is explored and found to have some degree of relevance. As the science evolves, other systems affecting the disease are discovered.

Interactions between the systems are explored and the literature becomes dense with the myriad details and hypotheses underlying the integrated pathophysiological process of the disease. Scientists focused on a particular disease have their own mental maps of the key processes and interaction points based on their own scholarly reviews, the reviews of others, and the particular window on the disease opened by their own research experience.

Predicting which interventions will actually work to change the manifestations of a disease is a difficult exercise. Systems show biological redundancy and regulatory control mechanisms can be multiple and complex. Blocking one path may not produce the desired result because alternative paths compensate. In no therapeutic area is this more problematic than in CNS disorders. The time scales of drug effects may be very different from the time scales relevant to the pathology underlying the disease.

Integrating these and other relevant variables to predict response is a major challenge. Disease models essentially attempt to scour the literature for all relevant information bearing on known aspects of the pathophysiology of a disease and explicitly synthesize the lines of causal events to create a model that dynamically describes all relevant regulatory mechanisms of the disease process and links elements by differential equations.

By integrating clinical, genomic, proteomic, physiological, and other biological data *in silico* in a computerized platform, it should be possible to capture the control principles of the system with ever-higher degrees of completeness. The model can be tested by assessing predictions against experimental data and can be enriched by the ongoing addition of experimental data.

In a sense, the models derive from "reverse engineering" of the disease, identifying known manifestations, reasoning back to the relevant causal mechanistic pathways, nesting detailed subsystems with control and context defined by the behaviour of the larger disease, incorporating new relevant pathways as they are discovered, and explicitly mapping how

perturbations of underlying causal pathways should impact on the disease. With the initial set of conditions representing the initial disease state as a baseline, the goal is to simulate the subsequent behaviour of the system in a process termed biosimulation to allow researchers to ask "what if" questions of the model. An analogy has been made to the complex process of airplane design: "If Boeing made airplanes the way the pharmaceutical industry makes drugs, you'd see them making many different designs, flying them, and then mass-producing those models that didn't crash." Such a model would not work economically for Boeing and it is not working optimally for pharma.

Better use of the ability to assist the effectiveness of human intellect by empowering and embodying it with explicit models is one underpinning of the future of pharma. In the Biodesign paradigm, the disease model is the intellectual core used to integrate all data collected from *in vitro*, animal, and human experimentation. The predictions obtained from disease modeling are only as good as the model but the explicit nature of the model facilitates correction when erroneous predictions are made.

Doubtless development of an effective disease model for any disorder, from epilepsy or Parkinson's disease to panic disorder or attention deficit disorder, would be a financially and intellectually intense and expensive investment, but the power of the model to suggest increasingly more powerful interventions would well cover its costs.

One important implication of the Biodesign strategy is that companies would focus on disease areas rather than on molecular platforms and would invest heavily in developing

increasingly sophisticated disease models in their areas of interest. In fact, a highly sophisticated and developed disease model would become the core intellectual property of the pharmaceutical company (much as the patent estate is now) and the centre of deliberations internally and with academic collaborators.

The possession of a sophisticated model would allow synergistic interaction between discovery and clinical colleagues because it would constitute an integrative mechanism, combining the results of animal data collected with the result of assessment of various probes. Moreover, such a model would suggest interventions to be tried in humans and could be modified as the results of those interventions are captured from patients with different stages and severities of the target illness.

Such models would have multiple applications. One of the most promising applications of this approach is to examine the logic of a disease and to predict which interventions at which part of the disease would be most likely to produce desired downstream effects on disease manifestations, facilitating the otherwise lengthy target validation and selection aspect of pharmaceutical development.

Another intriguing aspect of this approach is that it potentially offers an opportunity to assess the impact of agents that affect multiple pathways to see if they are synergistic, thereby presenting a possible solution to the specificity challenge noted previously. An underlying advantage of the disease model is that it forces a movement away from simplistic and reductionistic models of function and disease and toward an integrative model.

Disease is a reflection of complex processes and the more explicit the processes can be made, the better the chance for a rational intervention. One might picture the truth as an intricate stained glass window that has been shattered. A person focused on a particular mechanism or process can look through a fragment and claim, "I see the truth and the truth is green," but another can disagree, saying "I see the truth and the truth is red."

Both are right and neither is right. The truth is a complex composite that contains the information of both but exceeds it by the interaction of the pieces. Although singular geniuses may be able to attain a scholarly comprehension exceeding a model, the model has the advantage of allowing collaborative progressive hypothesis generation, testing, and growth, and does not preclude enrichment with the insight from singular geniuses.

EXPERIMENTAL MEDICINE

Experimental medicine groups within pharma perform a variety of roles. In the context of the Biodesign strategy, the role of experimental medicine is to develop the methodologies and tools needed to safely answer key questions about disease processes in humans, to strengthen the disease model, and to undertake studies in humans intended to assess the biological impact of therapeutic interventions.

On one level, many experimental medicine questions fall into the realm typically associated with clinical pharmacology departments. Is the compound absorbed and what is its concentration time course? Does the compound penetrate to the desired site of action for the appropriate period of time and at the desired concentration? For example, does an

antagonist penetrate to the CNS compartment at concentrations sufficient to produce the blockade desired?

Does the compound mechanistically do what it is intended to do? For example, if the compound is a cyclooxygenase (COX)-2 inhibitor, does it in fact show evidence of inhibiting COX-2 in the target organ? Does the compound exert a biological effect? In the best of all possible worlds, one would want to know if it produces a biological effect that is specifically causally relevant to the disease of interest. Because the full mechanistic causal chain is not always known in CNS diseases, it may not be possible to specify a key causal element.

If the compound does not produce any biological effect on systems it should affect and does not mimic biological effects that were important causally in the animal models used for earlier assessment; however, the probability of producing a profound effect on disease seems less than optimal. On another level, however, the role of experimental medicine goes beyond that of a typical clinical pharmacology assessment by assessing risk factors in an intervention. Many therapies are known to have class-associated limiting toxicities.

For example, some thiazolidinediones are known to be associated with varying degrees of hepatotoxicity. If a compound of this class is being developed, it would make sense to perform the key experiments needed to rule in or rule out the possibility that a compound has an unacceptable level of class-associated limiting toxicities very early in development. This would better protect larger numbers of future subjects from exposure to toxicity and reduce the probability of expensive late-stage failure.

In the Biodesign model, experimental medicine is actively involved in further developing the disease model. It would collaborate closely with discovery, analytical chemistry, a biomarker laboratory, and clinical pharmacology functions in developing sophisticated technologies to allow collection of information about the compound and its effects on human physiology and disease relevant to the model.

Information from pharmacokinetic and pharmacodynamic relationships in the peripheral and central compartments could be assessed using markers derived from biochemical, neurophysiological, pharmacodynamic, or neuroendocrine biomarkers, or from imaging technologies such as PET, single-photon emission tomography, and magnetic resonance spectroscopy, or on functional brain imaging or assessment technologies such as magnetic resonance imaging, electroencephalography, or magnetoencephalography.

Clinical staff familiar with the human disease, academic advisors familiar with the disease and with the biology underlying the selection of biomarkers, and discovery scientists familiar with the results of animal models all have significant input into the model and into the design of experiments, but it is also true that the detailed design and conduct of these highly specialized trials requires intense expertise in itself, expertise that needs the organizational support and rootedness that an experimental medicine group provides.

BIOLOGICAL TECHNOLOGIES

The third part of the Biodesign strategy centers on the development of protein probes to test normal physiology and disease in humans both to enrich the model and the insight

it yields into the targets for therapy and to potentially become therapies in themselves. The potential for biologics both as drugs and potential probes has been recognized. The major weight of pharma personnel and financial resources have focused primarily on the small molecule side, on producing "white powders" of highly defined molecules rationally designed and created to hit some specific target or targets.

The rise of the biotechnology industry, with its intense focus on protein therapeutics over the past 20 years, has been an important complementary development. In 1999, a total of 59 recombinant proteins and monoclonal antibodies were available. In 2001, more than 35% of the 37 new active substances launched were protein therapeutics. The power of rationally engineered proteins is becoming increasingly evident.

The development of protein therapeutics can begin with a naturally occurring protein acting as a preselected "lead" compound. However, a natural protein evolved by nature for one role and the ideal protein therapeutic may have some additional characteristics. For this reason, the lead protein may be modified to produce variant proteins as a therapeutic with more desirable characteristics that influence key features such as mechanism of action, efficacy, adverse events, stability through purification, formulation, storage and administration, solubility, and production costs.

Routes of modification may include site-directed mutagenesis and phage display. For example, structures can be modified to become substantially more robust to oxidative stress, changes in solution and pH, and temperature, or to replace cysteines with serines to reduce disulfide bond

formation. Similarly attachment to polyethylene glycol, fusion to proteins with long serum half-lives, alteration of oligomerization state, glycosylation, and modulation of receptor-mediated uptake and turnover can markedly alter the pharmacokinetic properties of a molecule.

Many engineered molecules are currently being marketed, e.g., aldesleukin (Proleukin), interferon â-1b (Betaseron), insulin lispro (Humalog), insulin aspart (NovoLog), and peginterferon alfa-2b (PEG-Intron). As technologies have developed, the opportunity to manipulate the protein sequence and/or composition, driven by specific hypotheses, has increased. It is remarkable that rational protein design has progressed to the point where proteins can be produced with novel mechanisms of action.

For example, interleukin-6 and vascular endothelial growth factor, both 4-helix bundle cytokines, have been engineered to function as receptor antagonists rather than agonists. One concern with protein therapeutics, particularly therapeutic proteins derived from non-human sources, has been the development of harmful immune responses. Increasing interest has focused on deriving therapeutics from human proteins that are less likely to stimulate an immune response.

Approaches to date include humanization of murine antibodies, mutagenesis of peptides that bind class II major histocompatability complex alleles, and mutagenesis of epitopes in the protein structure that are particularly important in stimulating the immune response. The rational design of protein therapeutics has the potential to use systematic and quantitative engineering approaches for structure-based sampling followed by high throughput screening of a protein library.

As hypotheses are identified, probes must be generated. In the long run, small molecules may be ideal for purposes of manufacture and marketing. However, the speed of development of protein therapeutics argues for a strong capacity to rapidly generate the molecules as probes and potentially as therapeutics. For example, if an antagonist for a particular molecule is required, monoclonal antibody probes for testing the hypothesis might be more rapidly generated than small molecules could be synthesized and allow hypothesis testing with a greater assurance of safety.

As interest focuses on new molecular targets that are very different from the serine proteases, G-protein-coupled receptors, and similar families of targets on which most pharma libraries were developed, the structural diversity of protein therapeutics might be used to create antagonist and agonist compounds as powerful probes.

The structural diversity of the protein probes and the increasing ability to modify and design their functionalities will facilitate a biological way to design therapeutics. These probes may develop into therapies in their own right but also serve the purpose of validating whether a particular target is worth pursuing with the resource-intensive processes needed to develop small molecule therapeutics. Because failure rates for new targets are high, by eliminating the poor targets more quickly the protein probes would facilitate a more focused and cost-effective small molecule effort.

Protein therapeutics have additional attractions as probes and as therapies. Protein therapeutics can be developed more quickly than small molecules, in the range of 2 years rather

than 4 years. While pharma discovery libraries may have a limited number of potential shape-probes to test against new genomic targets, protein therapeutics are capable of rapidly generating extraordinary shape diversity. The Cambridge Antibody Technology Group has created a library of approximately 10 billion different molecules, a size that dwarfs any other small molecule library.

And immunogenicity aside, protein therapeutics are degraded by normal proteolytic processes, reducing the problem of active metabolites, which complicates the development of small molecule therapeutics. The manufacture and cost of goods of protein therapeutics presents challenges and corresponding economic advantages in terms of not being as susceptible to generic competition, which complicates their economics as therapeutics, but their speed of and safety for development as probes is clear.

The penetration of protein therapeutics across the blood-brain barrier is a significant challenge but one that can be addressed; protein therapeutics have been shown to penetrate to the brain and in an animal stroke model prevented the slow progressive death of neurons.

MEETING THE CHALLENGE WITH BIODESIGN: IMPLICATIONS FOR THE FUTURE

The use of protein-based probes and therapeutics as a basis for the evolution of the pharmaceutical industry has been considered. A target molecule can fail in more ways that one. More than a thousand different problems have been noted in the gene underlying cystic fibrosis. As highly targeted protein therapies are developed, the initial molecules will most likely target the most common defects.

Subsequent lifecycle management will likely consist of developing additional molecules that target successively less-common molecular defects. These new therapies might be independent products or might be added to a cocktail, as study results indicate. Therapies are likely to be highly individualized and may be delivered through specialists rather than generalists.

The corporate structures based on marketing high-volume blockbuster compounds that work overall for a population but fail in many subjects will be replaced by structured delivery of high-utility therapies with markedly superior efficacy to smaller numbers of patients for a given therapy. The structure of marketing, advertising, and relations with healthcare providers and patients will change markedly in an environment of individualized medicine.

Advertising may play a lesser role and sophisticated diagnostics a greater role in determining which treatment a patient receives. An underlying theme of all of these considerations is the need for greater intelligence in the way neurotherapeutics are developed. We need to better understand the target diseases, animal models, the relation between the models and the diseases, and the interaction between therapeutics and the diseases and models, and generally have a more thoughtful, explicit, hypothesis-driven approach to therapeutics development.

While the intellectual assets within pharma are significant, intellectual assets in academia and in the clinic are just as important. With the shift in focus from platform-based intellectual property to disease-model-based intellectual property, pharma needs to engage the intellect of a wide base

of disease experts outside the corporate world. Conflicts between academic responsibilities and corporate confidentiality can and must be addressed in a manner fair to the involved entities and society.

The cost of failure is too high to permit anything less than engaging the best minds. Development of better remote-presence interactions such as internet-assisted meetings, methods for secure sharing of electronic data, and alternative employment relationships such as retainer-based hiring of academicians or investigators seem likely. Companies unable to move beyond the "not invented here" mentality will succumb in the competition with companies that are open to engaging the best and the brightest.

Another key implication for this model is that the relationships between regulators and developers will develop more as an interactive partnership. Why? The nature of the Biodesign model includes a lot of early hypothesis testing and developmental work conducted in healthy subjects and volunteer patients as the model is developed and its predictions are tested.

Investigators and developers are responsible for conducting well-designed, safe, and ethical research, but regulators have a key oversight responsibility for ensuring public safety and accountability in the process. An ongoing interactive, consultative relationship in which the data are fully and transparently available to regulators is the best way to ensure public safety.

Data management will likely rely on a rolling, continuously updated database with transparency of analysis and built-in electronic fidelity measurements such as audit trails and

access controls. The continuous safety evaluation of therapeutics based on some sort of sampling access to electronic medical records with appropriate privacy safeguards is also a likely point of collaborative interaction and dialectic between therapy developers and regulators.

4

Amorphous Pharmaceutical Systems

Interest in amorphous pharmaceutical systems has been steadily growing over the last 10 years. This is mainly because of:

- An increased understanding of amorphous systems in allied disciplines (e.g., food science and materials science).
- Greater prevalence of protein and peptide therapeutic agents that cannot be readily crystallized.
- An increasing need to work with active pharmaceutical ingredients (APIs) that have very low solubility in aqueous biological fluids.
- Greater regulatory scrutiny of the physical form of active and inactive pharmaceutical materials and dosage forms, and the manufacturing processes used to make them.

This article provides an overview of the properties and occurrence of amorphous pharmaceutical materials, and outlines their key applications in dosage form development. It describes their characteristics and the fundamental scientific basis for these characteristics. It also highlights the topical issues of chemical/physical stability and "polyamorphism."

BACKGROUND

Amorphous materials are a distinct class of solids, separate from the more common and well-known crystalline solids. At the molecular level, they lack the three-dimensional long-range order that is characteristic of crystalline solids; instead, their molecules are randomly arranged in space and the interactions between neighboring molecules (e.g., hydrogen bonds and electrostatic repulsion) are not repeated with any regularity throughout the sample.

Amorphous materials may be single chemical entities (e.g., a drug substance or excipient), or molecular-level mixtures of materials (e.g., drug–polymer dispersions). The latter trait is a property of amorphous materials that is rarely shared with crystalline solids.

Lack of molecular order in amorphous pharmaceutical systems may be because of difficulty in crystallizing the material in question, such as for high-molecular-weight proteins and polymers, or may be the result of selecting a processing operation that prevents crystallization from occurring (e.g., quench cooling a molten drug).

It is also possible that a normally crystalline sample can become partially amorphous (and thus partially crystalline) by being handled or processed in a manner that causes the

crystal structure to be damaged. The creation of amorphous character within pharmaceutical materials occurs both intentionally (e.g., to improve handling characteristics) and unintentionally (e.g., by poor control of a manufacturing process); in either case, it will have a significant effect on the physical and chemical properties of the sample because the material will have higher average level of molecular mobility, and a higher entropy and enthalpy than the crystalline form of the same material.

APPLICATIONS

Amorphous materials are used for a wide range of pharmaceutical applications:

- Therapeutic proteins typically exist in a noncrystalline or amorphous form because their macromolecular structures are not readily crystallized.

These materials are commonly prepared in an amorphous dispersion with bulking and stabilizing excipients to ensure an adequate product shelf life and ease of administration. Examples of such therapeutic proteins include insulin and interferon.

- Amorphous materials typically have a higher rate of dissolution and a higher kinetic solubility that their crystalline counterparts. These characteristics can be exploited to enhance the rate and extent of absorption of poorly water-soluble APIs from the gastrointestinal tract. Such formulation approaches have been described for many APIs including indomethacin, griseofulvin, and several barbiturates.
- The ability of amorphous materials to form

molecularlevel mixtures has been used by many workers in an attempt to stabilize, and otherwise modify, the properties of difficult-to-handle pharmaceutical materials. Several reviews of the uses of such solid dispersion systems have been published

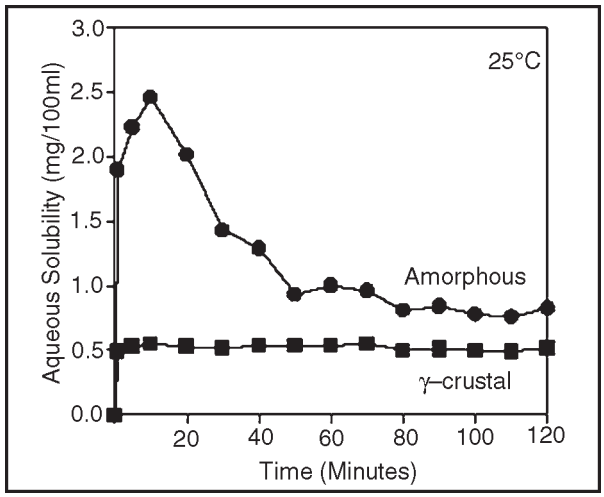


Fig. Aqueous Solubility of the Amorphous and Crystalline Forms of Indomethacin.

- Agreatmany of thematerials that are used aspharmaceutical excipients occur naturally in the amorphous or partially amorphous state (e.g., gelatin and starch).

Many others have been found to possess improved handling and mechanical properties when processed in such a manner as to render them at least partially amorphous. Examples of this include the grades of microcrystalline cellulose and lactose monohydrate used as pharmaceutical tableting diluents.

- Pharmaceutical materials that are processed by "high-energy processes" such as freeze drying, spray

drying, jet milling, and melt extrusion are often rendered at least partially amorphous. This occurs by virtue of the fact that these processes create conditions that can prevent crystallization from occurring as the solid material is formed, or they can mechanically disrupt the structure of an existing crystalline material. In commercially freeze-dried cephalosporin antibiotics, the processing conditions are such that the drug is usually in the amorphous state in the final formulation. Likewise, in jet-milled drug substances used for aerosol drug delivery devices, the API is often in a partially amorphous state after it has been milled.

OCCURRENCE

There seems to be no limit to the types of pharmaceutical systems that can be isolated in the amorphous state. In the literature, samples of sugars, acids, bases, polymers, buffers, inorganics, salts, natural products, proteins, and low-molecular-weight APIs have all been reported to exist in an amorphous form. Likewise, pharmaceutical raw materials, intermediates, and final products that include these amorphous materials are widespread and varied.

The most common methods by which amorphous pharmaceutical materials are intentionally manufactured are freeze drying, spray drying, and melt extrusion; however, many other processes that would create an amorphous sample can be envisaged.

Processes that have been reported to unintentionally induce amorphous character in pharmaceutical samples include milling, blending, wet granulation, and drying,

although one can appreciate that almost any handling or processing operation has the potential to cause the molecular disruption of a sensitive crystalline material.

Different localized levels of molecular order can coexist in some pharmaceutical materials, giving rise to the occurrence of "partially crystalline" (and "partially amorphous") systems. In most cases, the properties of such materials (e.g., density) are intermediate to those of the 100% amorphous and 100% crystalline samples.

By deliberately varying the level of crystallinity in such systems, their properties can be customized for a particular purpose. An example of this is with the tableting excipients microcrystalline cellulose and spray-dried lactose, which have had their compression characteristics optimized by manipulating their amorphous content.

The properties of partially crystalline materials may be approximated in many instances by making physical mixtures of the totally amorphous and crystalline samples. This is known as the "two-state" model for partially crystalline systems. However, such experiments should be undertaken with caution as the mixed "two-state" material can sometimes have significantly different properties from the partially crystalline material that is manufactured directly (the real "one-state" system).

The presence of low levels of amorphous character in predominantly crystalline samples (and vice versa) occurs quite often, and can be the cause of unexpected processing or stability problems in pharmaceutical systems. Much effort has been directed at developing analytical tools both to detect and to quantify small amounts of one phase in

another, primarily through the use of calorimetric and spectroscopic methods.

This is an important area of research and development because the performance and stability of seemingly 100% amorphous (or 100% crystalline) materials can be significantly altered by the presence of very low levels (<1%) of the opposite phase. This includes the phenomenon known as amplification that has been eloquently described by Ahlneck and Zografi.

CHARACTERISTICS

The common physical characteristics of amorphous pharmaceutical materials are quite different from those of their corresponding crystalline materials (Table). Generally:

Table. Examples of Some Amorphous Pharmaceutical Materials.

Material types	Material categories
Organic small molecules (e.g., lactose)	Active pharmaceutical ingredients (API)
Polymers (natural and synthetic) (e.g., nelfinavir mesylate)	Tablet fillers (e.g., microcrystalline cellulose)
Sugars and carbohydrates (e.g., sucrose and dextran)	Glidants (e.g., silicon dioxide)
Peptides and proteins (e.g., insulin)	Suspending agents (e.g., tragacanth and guar gum)
Lipids and oils	
Salts, acids, and bases (e.g., zinc oxide)	
Buffer systems	
Frozen aqueous solutions	
Dosage forms	**Therapeutic areas**
Tablets (e.g., quinapril hydrochloride)	Anti-infectives (e.g., erythromycin ethyl succinate)
Capsules (e.g., pancrease)	Anti-coagulants (e.g., warfarin sodium)

Oral suspensions (e.g., montelukast sodium)	Anti-asthmatics (e.g., cefuroxime axetil)
Injectables (e.g., coumadin)	Anti-psychotics and anxiolytics
Sterile powders (e.g., cefoxitin)	Hypnotics and anticonvulsants (e.g., barbiturates)
Topicals (e.g., zinc oxide powder)	Anti-hypertensives
	Anti-inflammatories (e.g., indomethacin)
	Analgesics (e.g., aspirin)
	Antacids (e.g., aluminum hydroxide)
	Diuretics
	Enzymes (e.g., pancreatin)
	Hormones

- The true or absolute density of the amorphous phase is 5–20% less than that of the crystalline phase (Table 3). This is because of the less efficient packing of molecules that are randomly oriented relative to each other.
- When viewed using polarized light microscopy, particles of an amorphous sample will not exhibit the birefringence that is characteristic of crystalline materials. This provides a very simple qualitative test for amorphous or crystalline character in a pharmaceutical powder.
- Amorphous materials will not diffract X-rays in a coherent manner; thus powder X-ray diffraction patterns are broad halos with no or very few characteristic peaks for these materials.
- The apparent aqueous solubility of amorphous materials is much higher than that of their crystalline

counterparts.This is a kinetic phenomenon and, eventually, the solute in the supersaturated solution that is formed will begin to crystallize and the equilibrium solubility of the crystalline phase will be attained. The transient increase in solubility is often significant (>10_) and can be exploited to give markedly improved biopharmaceutical performance.

- Amorphous materials will absorb significant amounts of water vapour from their surroundings relative to their crystalline counterparts. This is true even for very hydrophobic materials.
- When analyzed using common thermal analytical methods [e.g., differential scanning calorimetry (DSC)], amorphous materials will exhibit an apparent second-order phase transition (the so-called "glass transition temperature" or Tg) in a temperature range that is significantly below the melting point of the crystalline material. The Tg of an amorphous material is one of its characteristic properties and can be used to assess its likely stability and suitability for use in pharmaceutical dosage forms.
- The degree of molecular mobility (assessed as the average molecular relaxation time *t*) of amorphous systems in the region near Tg follows a non- Arrhenius temperature dependence. This so-called "fragility" (d*t*/ dT at Tg) of amorphous materials is a defining characteristic. _ The mechanical properties of amorphous materials are noticeably different from their crystalline counterparts because of the different number and type of intermolecular interactions.

Hancock et al. found that at temperatures more than 50K below Tg, an amorphous drug powder formed compacts that were significantly more brittle than those formed from the crystalline form of the drug although the tensile strengths of the compacts were similar.

FUNDAMENTAL DESCRIPTION AND DEFINITION

Although amorphous pharmaceutical materials can be readily isolated and may persist for many thousands of years, they are in fact a thermodynamically metastable state and will eventually revert to the more stable crystalline form.

Figure shows a "snapshot in time" of the free energy–temperature relationship for a material that can be isolated as both an amorphous form and a crystalline form. This quasi-equilibrium thermodynamic view of the amorphous state shows that the amorphous form has a significantly higher free energy than the crystalline form, and illustrates why it is expected to have a much higher aqueous solubility and significantly different physical properties (e.g., density).

Figure also illustrates one common pathway for forming amorphous materials (i.e., supercooling the molten material to temperatures below the melting point of the crystalline phase). To achieve this, rapid cooling rates are usually required. As the material is cooled, the speed of molecular motions within the sample decreases dramatically and the viscosity of the material increases markedly. In the region known as the glass transition temperature (Tg), the average molecular mobility is sufficiently slow that the system falls out of energetic equilibrium with its surroundings, and the material forms a so-called glassy phase.

Such glasses share many of the properties of crystalline materials (e.g., solid macroscopic appearance, low specific heat capacity, and low thermal expansivity), which makes them attractive for use in pharmaceutical dosage forms. However, all glasses are metastable relative to both the equilibrium supercooled liquid and the crystalline forms of the material, and their longterm physical stability needs to be demonstrated before they can be relied on for use in the manufacture of pharmaceutical products.This has been done for the amorphous drug delivery systems that are currently marketed, and such studies have been the subject of many reports in the literature.

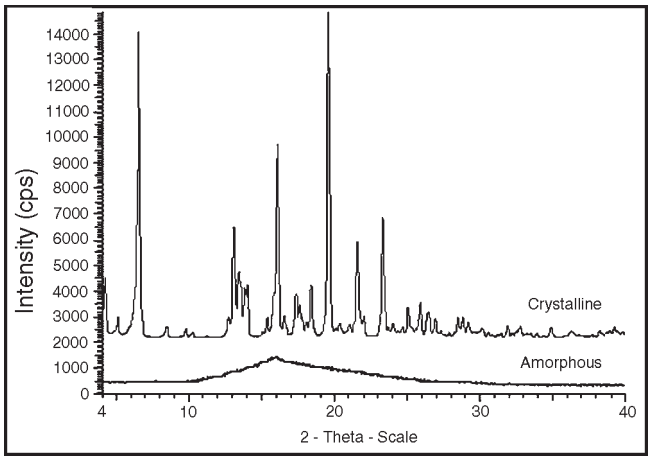


Fig. X-ray Diffraction Patterns for Powdered Amorphous and Crystalline Samples of an Experimental Drug Substance.

STABILITY

The chemical and physical stability of amorphous pharmaceutical materials is controlled by the same basic factors as for crystalline materials [i.e., molecular structure (chemistry), purity (absence of catalysts, chemical reactants, or nucleating agents), molecular orientation (physical form), and molecular mobility (related to temperature)].

For any sample of a given molecular structure and purity, there will be more possible molecular orientations that occur in an amorphous sample than in a crystalline sample.

Table. Mechanical Properties of the Crystalline and Amorphous Forms pf Phenobarbitone.

	Vickers hardness (N/cm^2)	**Bending Strength (N/cm^2)**
Form II	42,000	187
Form III	28,000	567
Amorphous	6,500	1056.24

Thus many more different types of chemical and physical transformations could potentially take place. At a given temperature, the molecular mobility in an amorphous material will also be significantly higher than in any of the corresponding crystalline forms, and this can give rise to a greater chemical and physical reactivity in the amorphous sample.

However, it is important to realise that a very close interdependence (or "coupling") between the mechanism of the chemical or physical instability of interest (e.g., charge transfer, free radical attack) and the molecular orientation and/or mobility of the sample is necessary for an amorphous sample to be significantly less stable than its crystalline counterpart.

In many instances (e.g., free radical-initiated oxidation reactions), the stability of a drug compound is not significantly affected by either its molecular mobility or the orientation of the molecules; thus the amorphous form has comparable stability to the crystalline material.

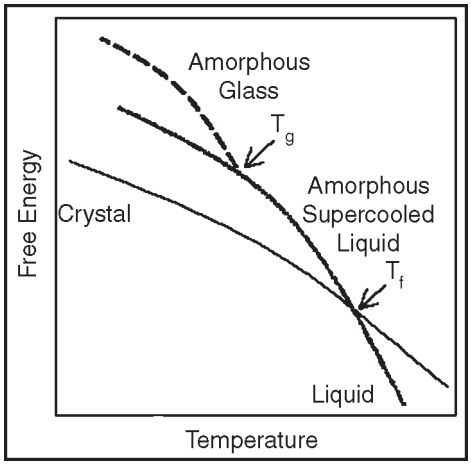


Fig. Free Energy Diagram for a Material in the Amorphous and Crystalline States (Tg Denotes the Glass Transition Temperature; Tf Indicates the Melting/fusion Point).

In some instances (e.g., insulin), the more ordered structure of the crystalline material can actually increase the likelihood of certain intermolecular contacts and cause the crystalline form to have a lower level of stability. From a snapshot of the current literature, it would appear that:

- Physical transformations (usually solid-state crystallization) are more often directly linked to molecular mobility and orientation than the most common chemical reactions (oxidation and hydrolysis); thus the major stability concern for amorphous materials is with their tendency to revert to the crystalline state. As with all crystallization processes, there are the normal nucleation and propagation (crystal growth) stages to consider, and procedures that increase the barrier to nucleation or slow the rate of crystal growth can be used to physically stabilize many amorphous materials.

- One additional factor to keep in mind in mind is the greater purity of most crystalline materials, which can contribute significantly to their enhanced stability, especially during the earliest stages of drug development.
- The tendency for amorphous materials to sorb significant amounts of water vapour from their surroundings can give rise to a markedly reduced chemical and physical stability relative to the crystalline form of the material.

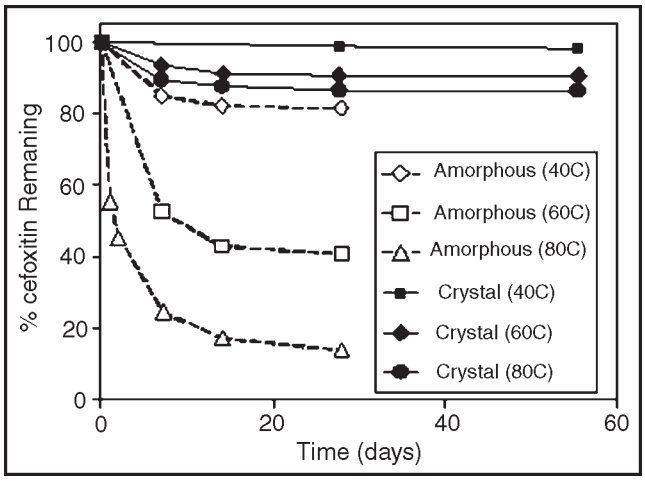


Fig. Chemical Stability of the Crystalline and Amorphous form of Cefoxitin.

The sorbed water may participate in a chemical reaction (e.g., hydrolysis), or may simply act as a catalyst for a chemical reaction.

- Sorbed solvents such as water will also plasticize most amorphous pharmaceutical materials, and this can have a negative impact on both physical and chemical stability by increasing the molecular mobility of the sample at any given temperature. A large number of studies of the water vapour sorption

behaviour of amorphous pharmaceutical materials and its impact on the glass transition temperature have been reported in the literature. Hancock and Zografi summarized the results of many of these studies and provided a simple way in which to predict the magnitude of the plasticizing effect when amorphous materials are exposed to water vapour.

MOLECULAR MOBILITY

To fully understand the performance of amorphous materials, it is necessary to be able to measure the molecular mobility of the samples on interest. This is because at temperatures as far as 50K below the glass transition temperature, pharmaceutical glasses exhibit significant molecular mobility that can contribute to both chemical and physical instability.

The main techniques that have been developed for monitoring molecular motions in amorphous materials are nuclear magnetic resonance (NMR) and calorimetric techniques (e.g., DSC and isothermal microcalorimetry).

Average molecular relaxation times and relaxation time distribution functions obtained from these techniques have been used to predict the relative stability of different materials and the storage conditions required for a normal product shelf life.

This requires an assessment of the speed of the fastest molecular motions relative to the duration of storage that is anticipated for the product and has utilized the well-known concept of the dimensionless Deborah number.

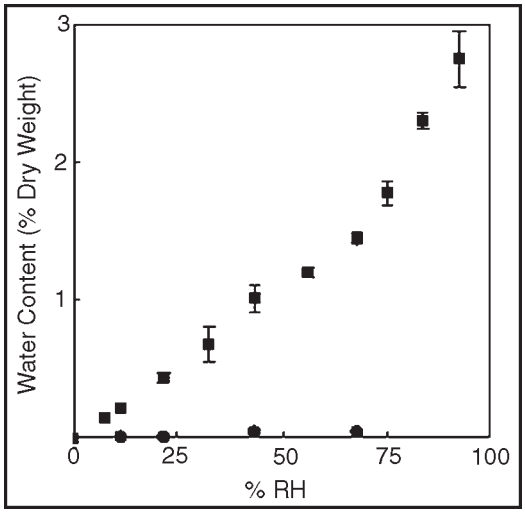


Fig. Water Sorption Isotherm for Crystalline and amorphous Samples of Indomethacin (■) Amorphous;(•) Crystalline.

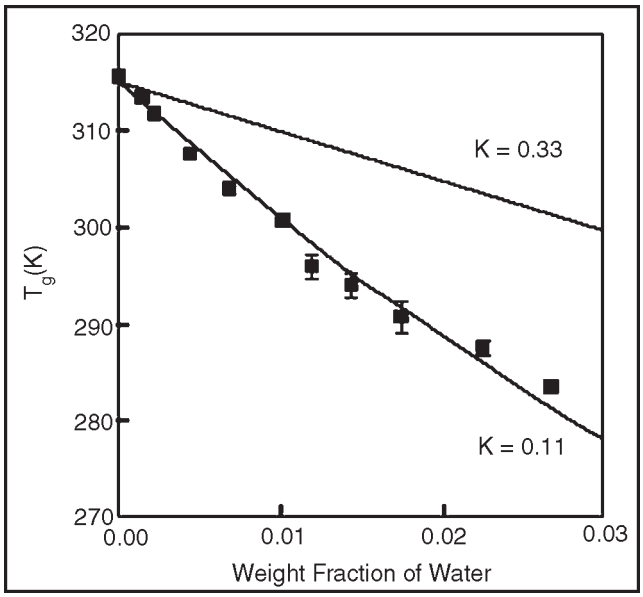


Fig. Glass Transition vs.Water Content Relationship for Amorphous Indomethacin.Lines show fit to the Gordon Taylor Wquation with K Values of 0.33 (Predicted) and 0.11 (best fit).

Such studies have also shown that the temperature dependence of molecular motions in amorphous materials (their fragility) is significantly different from that of crystalline materials (non-Arrhenius), especially at temperatures above

the glass transition temperature (Tg). This can give rise to marked differences in the effects of temperature on the chemical and physical stability of amorphous and crystalline pharmaceutical materials. At temperatures below Tg, amorphous materials may exhibit an Arrhenius-like temperature dependence of molecular mobility, particularly over the relatively narrow temperature ranges where pharmaceutical materials are usually handled.

However, above Tg, the temperature dependence of molecular motions is non-linear and more pronounced, and often needs to be described by the so-called Vogel-Tamman-Fulcher (VTF) or Williams-Landel-Ferry (WLF) relationships.

POLYAMORPHISM

The occurrence of multiple polymorphic forms of crystalline drugs and excipients is well known to pharmaceutical scientists, and the possible occurrence of polymorphic forms of amorphous pharmaceutical materials has recently been reviewed.

"Polyamorphism" is an intriguing concept from both a scientific and commercial perspective because of the significant impact amorphous character can have on the performance of pharmaceutical materials and the potential opportunities that might arise to exploit (and may be patent) new and improved forms of existing pharmaceutical materials.

olyamorphism is strictly defined as the existence of two distinct amorphous states of the same material separated by a clear phase transition. In the most well-known example, it has been noted, based on a thermodynamic analysis of

the heat capacity of water and ice, that there are differences in the properties of amorphous ice samples formed by vapour deposition and by quench cooling from the liquid state.

There have been several anecdotal reports of apparently different forms of amorphous pharmaceutical materials with readily discernable physical and chemical characteristics, and some marked differences in their pharmaceutical performance. Examples include an antibiotic prepared by lyophilization and glasses of an anti-inflammatory agent produced by fast cooling of the molten material.

For these particular materials, even though the amorphous samples had significantly different physical properties, there was no direct evidence of polyamorphism according to the strict thermodynamic definition. Inspection of the relevant literature reveals that most apparently polyamorphic amorphous pharmaceutical materials have been isolated and/or stored below their calorimetric glass transition temperatures.

Such "glassy" amorphous materials are, by definition, not at energetic equilibrium with their surroundings, and their properties reflect the conditions under which they were isolated and subsequently stored. Glass transition vs. water content relationship for amorphous indomethacin.

Lines show fit to the Gordon– Taylor equation with K values of 0.33 (predicted) and 0.11 (best fit). departure from equilibrium and the very long time that it takes glasses to spontaneously relax back to the equilibrium supercooled liquid state, it appears that it is possible to isolate amorphous materials with distinct physical and chemical properties, which are not true polyamorphs.

It has been proposed that the term "pseudo-polyamorph" be used to describe glassy amorphous materials that have different energetic states as a result of their different conditions of production and storage, by analogy to the term "pseudopolymorph," which is used to describe different crystalline forms that do not fit the strict thermodynamic definition for crystalline polymorphs.

Because such pseudo-polyamorphs can exist for significant periods of time (certainly longer than the shelf life of pharmaceutical products), they can present very real and significant challenges for practicing pharmaceutical scientists.

These systems still need to be carefully characterized and understood if they are to be incorporated in pharmaceutical dosage forms and, to this end, techniques need to be developed to monitor and distinguish their characteristics, and to track their very slowly evolving physical and chemical properties (e.g., heat capacity and true density).

Strategic Challenges

INTRODUCTION

The pharmaceutical industry must generate novel, effective, and safe therapeutics that address unmet medical needs at a cost that is palatable to consumers globally, in a time frame that allows effective recovery of the investments in the systems and processes necessary to generate new products, and in a manner compliant with international regulations. In the long run, return on investments in therapeutics development must pay for the full economic costs of development or the process will fail economically

and cease. The seriousness of this challenge can be inferred from the steady attrition of pharmaceutical companies large and small, old and new. The economic component grows steadily each year because of increasing pricing pressures, costs, generic competition, and regulatory hurdles, along with decreasing productivity.

The scientific challenges for therapeutics developers are great in every area but particularly so in neurotherapeutics. On one level, new scientific discoveries and tools create the possibility for a golden age of therapeutics development, in which many long-term scourges of humankind may be addressed in a fundamental and powerful manner. Yet while we move from symptomatic treatments to mechanistically targeted cures, we struggle to base development on fields of knowledge that are rapidly changing and far from complete and in which our understanding of the interactions between factors is fragmentary at best. Neurotherapeutics developers face unusually high failure rates and huge attendant expenses that must be covered by the small number of therapeutics that succeed. The combination of pricing constraints and trends of increasing regulatory hurdles, falling productivity, lengthening cycle times, and falling success rates is not sustainable in the long term. Many developers have focused on strategies to improve scientific and operational performance, such as improved collaboration with academicians and regulatory agencies, better use of technologies to explore study data and design trials, greater operational effectiveness in selecting and training quality sites, monitoring and harvesting data, producing analyses, and amortization of costs over a wider base by global

registration. Although important, these approaches are not enough. Issues of discovery and clinical development strategy must also be addressed. These issues are particularly pertinent to neurotherapeutics development and even more so to the development of therapeutics for neurodegenerative disorders.

A new and better paradigm must be developed for pharma to succeed economically and medically in the long term. While incremental improvement of current approaches can be useful and necessary, a more powerful and deeply penetrating change is needed to meet the challenges the therapeutics development community faces. In the last century, the pharmaceutical industry progressed from a strategy of observation and serendipity to the current strategy of rational design, which argues that mechanistic understanding of disease leads to mechanistically targeted molecules, which works if the theory of disease etiology is correct. The conceptual weakness of this strategy is that the map is not the territory.

The systems biology underlying disease is complex, the relevant mechanisms are often multiple, interactive, and variable across species, the individual variability in an outbred human population is great, and the statistical noise generated by imperfect clinical assessment instruments together are a huge barrier to overcome. Narrowly targeted molecular interventions based on simplistic mental maps of the disorder sometimes work but often run afoul of these complicating factors. The next strategic approach, Biodesign, potentially surpasses the performance of the rational design strategy by embracing complexity by using the capacity of

the immune system to rapidly generate the myriad potential probes needed to assess the therapeutic potential of different interventions in conjunction with *in vivo*assessment technologies targeted to the human disease model and integrated with computerized disease models.

Elements of this proposed paradigm include the use of disease modeling to select better targets and potential therapies, along with information-science-based approaches to enhance small molecule chemistry, exploitation of the potential for biological technologies to rapidly generate mechanistic probes, and development of different strategies for using animal models and enhanced strategies and abilities to study human molecules mechanistically and biologically.

The synergistic use of these approaches will change the overall business model of companies engaged in the development of neurotherapeutics as well as other therapeutic areas. The next-generation pharmaceutical company will be defined by intense, deep, and advanced focus on specific target diseases, working with defined populations and specialists to target disease mechanisms with ever-broadening power. The model will combine expertise in protein therapeutics and small molecules and use them interactively in devising commercially viable therapeutics. This development will both depend on and generate a new set of relationships and interdependencies between academicians, regulators, and industry. In so doing, the Biodesign paradigm may revolutionize medicine and solve the economic challenges facing healthcare systems and therapeutics developers alike.

TERRAIN

In determining the strategy best suited to creating the next and better generation of medical therapies, it is critical begin with a map—a conceptual picture of the terrain, the challenges and obstacles—so that the most effective path forward can be plotted and followed.

This essay will focus on describing key elements that must be overcome by next-generation industry strategy because pharmaceutical development sits at the intersection of two different concerns. First, there must be an unmet medical need of sufficient magnitude as to potentially justify the investments required to generate a novel therapeutic.

Second, the basic science or a key observation must exist to support the hypothesis that some mechanism or molecular target or target process relevant to a disease exists and is sufficiently defined as to permit discovery approaches to operate with some confidence of a potentially successful outcome. When there is an existing and potential market and a clear scientific strategy to produce clinically meaningful benefit, industrial processes come into play.

Neurotherapeutics development exists within the larger context of general pharmaceutical development but has unique scientific and economic challenges and opportunities that make it one of the most demanding and high-risk areas for pharma. (Note that the term pharma is meant in a broad and inclusive sense primarily to define both the pharmaceutical and biotechnology industries, although many issues are equally relevant to device therapeutics.) The economic opportunities may be great, but success rates are

appallingly low, costs of development are high, and timelines for return are unusually long.

OPPORTUNITY

The pharmaceutical industry was born in the 19th century through interactions among analytical chemistry isolating active compounds from natural substances, organic chemistry becoming increasingly sophisticated in modifying those compounds and producing novel therapeutics, and the progress of experimental pharmacology.[1] Historically the industry has evolved with help from serendipity, hard work from many and singular genius from a few. The early successes of the chemotherapeutics era centered on antibiotics. In the mid-20th century, technological and conceptual advances in biochemistry triggered an explosion of therapeutics targeted towards protein receptor and enzyme-target agents. More recently, advances in biotechnology have triggered a creative burst of protein therapeutics.

Technologies for developing therapeutics have improved dramatically. X-ray crystallography and other structural approaches help define target shapes with angstrom accuracy, allowing for more rational design and, at its best, virtual chemical design. Automated combinatorial chemistry vastly speeds the ability to generate new small molecules to test against those targets. Various types of assay chips make it possible to perform tens of thousands of screenings in an hour. The combination of a vastly increased array of targets through genomics, the ability to generate large numbers of potential ligands through combinatorial chemistry, and the ability to conduct automated high-throughput screenings

through *in vitro* or cell-based assays potentially creates a situation in which the magic of large numbers may potentially be more productive of therapies than the older system, emphasizing scientific reasoning and critical discourse between chemists and biologists.

The recent description of the human genome and the rapid developments in understanding how it provides cells with instructions to produce as many as 300,000 different proteins holds the promise of allowing a fundamental understanding of the molecular logic of the life process, thereby creating myriad openings for highly targeted and possibly personalized therapeutics. The development of brain imaging technologies such as positron emission tomography (PET) and functional magnetic resonance imaging opens the door to more direct and quantitative understanding of drug effects in humans. Looking at what has been accomplished and at the ongoing stream of progressive improvements across the board, the historical vector points to a future with remarkable possibilities for addressing major human health problems in all therapeutic areas, including neurotherapeutics.

ECONOMIC CHALLENGES

Economic success

Human behaviour is typically driven by a complex combination of competing and complementary motives. Such motives, which may include interest in science, personal experiences or concerns, economic drives for reward, and humanitarian motives, drive resource allocation and

aggressiveness in pursuing goals for any given therapeutics developer. The magnitude of investment required for success is so huge that it is understood that economic motives must and—from a societal resource allocation perspective—should play a strong role.

The very fact of having sufficient resources to pursue large-scale development implies economically prudent decisions in the past to generate those resources. Rational economic investment presupposes some hope of reward that must be balanced against risks. If there is huge risk in pharmaceutical development, the real impact of successful therapeutics on human lives and national economies is also huge and therefore pharmaceutical development is potentially filled with rewards that drive the economics of investment.

Unfortunately, in therapeutics development, it is impossible to predict with any certainty whether a given effort will succeed or fail, therefore costs will be incurred for both successful and failing programmes. For the overall pharmaceutical development enterprise to continue in the long term, the income that ultimately results from the successful development efforts considered in aggregate must compensate for the costs of and capital for both the successful and unsuccessful efforts.

The same is true for individual companies. If the revenues produced by successful products do not compensate for the fully loaded costs of development, the costs of capital, and some profit incentives, the failing enterprise will eventually become extinct. The difficulty of meeting this basic condition can be inferred from the large number of pharmaceutical

companies that have failed and disappeared or been acquired. Companies may merge for reasons of synergy and/or cost reduction, but the reality is that many mergers are driven by the need of endangered or nonviable entities to find a stronger partner to secure whatever residual value their assets contain, values that would be completely lost if the company went under. Conversely, the more profitable the successful efforts are, the more resources available to explore a wider range of potential projects, including inherently riskier novel approaches. Where successful efforts are less profitable, fewer resources are available to explore novel approaches and attention tends to focus on activities that are perceived as lower risk, such as the development of therapeutics based on more proven mechanisms (often derided as "me too," but conferring some valuable marketplace choices), life-cycle management of existing products, and acquisition of products for which better marketing or combination with the existing business holds the promise of commercial reward.

Economic reward

The human and economic burden from CNS disease is great and drives the need for significant investments in therapeutics development. Size estimates of the CNS market depend in part on the entities considered as neurotherapeutics, *e.g.*, neurologic, psychiatric, pain, anesthetics, or drug addiction and abuse agents. In general, sales of CNS therapeutics comprise approximately 15% of total pharmaceutical sales, approximately 30 billion worldwide. About two-thirds of these sales are for psychiatric

treatments; historically, more mechanisms leading to effective products have been discovered relevant to psychiatric illness than to neurologic illnesses. Effective CNS pharmaceuticals have the potential to provide a huge benefit to patients and economies. For example, the estimated annual economic costs of anxiety disorders, depression, and schizophrenia are 47 billion, 44 billion, and 33 billion per year, respectively.

These numbers reflect the current market but may vastly understate the potential market. First, many important CNS disorders have no curative treatment at all; thus if a treatment is developed, an entirely new market comes into play. Second, other disorders have only ameliorative therapies that either have limited efficacy or are associated with significant side-effects that strongly restrict their use. More effective or better-tolerated safer therapies have the potential to dramatically increase utilization and therefore market size.

Third, markets currently served by workable but suboptimal generic therapies have the potential to grow dramatically with the introduction of more effective or safer proprietary therapeutics. In assessing the economic potential of any prospective therapeutic, the unmet medical need, the number of potential patients who could be served by a safe and effective therapy, the seriousness of the illness, and the disability entailed must be considered (and all affect the potential benefit and hence the price of a therapy). An important concern arises from these assessments. While very well-developed markets such as the market for antidepressants, anti-epilepsy agents, or anxiolytics are well-

understood and valued, the market worth for diseases in which current therapeutics have had less impact, such as stroke or dystonia, is less clearly defined. In this situation there is the potential for underestimation of market size and consequent underinvestment.

Economic risk

A few general factors control whether development of a particular therapeutic is economically feasible. The key economic determinants of reward are the size of the market, the real competitive advantages conferred by the potential new therapeutic, its anticipated patent life ("marketing life") and potential price, and the number of competitors already occupying the market. The key economic determinants of risk are success rate, development time, and development cost.

In general therapeutics development, for every 5,000 to 10,000 compounds screened, typically about 250 will enter preclinical testing; of those, five will enter clinical testing and one will win approval by the countries Food and Drug Administration (FDA). Overall, only about 11% of new active substances entering clinical development are predicted to reach the market. The success rates for neurotherapeutics, however, are far lower than average.

The relative difficulty of neurotherapeutics development is illustrated by a comparison of the chance of compounds initiated into human testing to progress to eventual marketing across therapeutic areas: anti-infectives, 33%; cardiovascular, 6%; anti-cancer, 6%; and nervous system, 1%. Of equal concern is the fact that neurotherapeutic

compounds fail late in development (during phase 3 pivotal testing) far more often than other compound categories. The chance that compounds initiated into pivotal trials will subsequently progress to eventual marketing across therapeutic areas is higher, but nervous system compounds are still dramatically riskier: anti-infectives, 75%; cardiovascular, 43%; anti-cancer; 32%; and nervous system, 14%.

As a rule, pivotal programmes cost about three times as much as the combined cost of phase 1 and 2 trials. Low success rates late in development hugely accelerate the financial risk of neurotherapeutics development. Of pharmaceuticals that do win approval, only one in three will produce revenues that match or exceed development costs. The revenue stream and profits produced by one pharmaceutical will basically have to pay for the tens of thousands of antecedent compounds produced, screened, and rejected along the way.

The length of development time is a key parameter governing economic risk for the developer. Because the ability to recoup the massive investments required to develop a pharmaceutical usually drops dramatically once it goes off patent, the costs of development must be recouped while the product is on patent. Longer development times directly reduce the number of profitable years remaining for any therapeutic. In the Countrys, it takes 10 to 15 years to move a new therapeutic from discovery through regulatory approval: around 4 to 6 years for the discovery/preclinical phase, around 4 to 6 years for the clinical phase, and 1 to 2 years for the initial regulatory approvals. In a patent life of

20 years, that leaves only 5 to 10 years to recoup the fully-loaded costs of development and capital. Median development times vary more on the basis of indication rather than target organ, *i.e.*, analgesics development is comparatively rapid but progressively longer times are required for neuroleptics, Alzheimer's disease therapies, antidepressants, and stroke treatments. The two biggest factors leading to prolonged development times are the time required to enroll enough patients for studies to have statistical significance and the treatment period needed to detect an effect.

Both factors are particularly acute in neurotherapeutics development; most therapeutics for neurodegenerative diseases tend to require relatively prolonged periods of treatment and observation to detect an effect. Additionally, the real cost of a development programme includes the costs of capital, that is, the income that could have been derived had the funds been invested in a different, lower-risk investment. Long development times raise capital costs. Analyses by the Tufts Center for the Study of Drug Development have indicated that reducing the total development time by half will reduce total costs by 29%.

Development costs are a combination of fixed and variable costs. Extensive and expensive discovery resources are needed to explore mechanisms, probe interventions molecularly, develop therapeutic molecules and delivery mechanisms, and assess toxicology and other parameters in detail. Developmental testing of potential therapies in humans, as controlled by international regulations, principles of good clinical practices, and other constraints, involves extensive efforts to conceive development plans and

protocols, conduct detailed testing in thousands of patients, capture and analyse all data in robust and reliable systems, and write extensive reports and submissions.

Cost estimates for pharmaceutical development vary widely but were estimated by the Tufts Center to be approximately 802 million in 2000. This estimate included an average out-of-pocket cost per new drug of 403 million plus costs of capital; it was further noted that the capitalized post-approval development costs raise the overall pre- and post-approval cost to 877 million. This number includes out-of-pocket preclinical and clinical expenses and costs of capital for preclinical and clinical expenditures for the expenses of both project failures and successes. These costs have consistently been driven upward by the progressive increase in the number of clinical trials: from 30 in the period from 1977 to 1980, to 68 in the period from 1994 to 1995. Similarly, the number of patients per New Drug Application (NDA) has increased from 1,576 in the period from 1977 to 1980 to 4,237 in the period from 1994 to 1995.

The 802 million figure is likely conservative. As a cross-check, one could examine the research and development (R&D) budgets of most major pharmaceutical companies, divide by the portion of the 403 million related to direct expenses rather than costs of capitalization, and get a number of predicted compounds far higher than the average yearly number of new chemical entities registered by that company. Even if the $802 million was considered as fully-loaded costs (including costs of capital) and divided into the R&D budgets of the larger companies, the resulting number would be higher than the average number of new chemical

entities registered by those companies per year. In 2000, for example, 11 major pharmaceutical companies had R&D budgets greater than $2 billion per year, yet based on current pipelines and success-rate estimates, predictions suggest launch rates averaging 1.3 new active substances per year over the past 6 years. As already noted, these numbers are of all the more concern when coupled with the realization that the R&D costs are such that only three in 10 marketed drugs produce revenues that exceed or match their development costs.

Investment and productivity

Pharmaceutical R&D investments are high and growing geometrically. In 2002, members of the Pharmaceutical Research and Manufacturers of America spent approximately 32 billion on pharmaceutical R&D, which represents about a 15-fold rise over the past 20 years and exceeds the National Institutes of Health (NIH) budget of 24 billion. NIH funding is critically important for the general advance of health sciences and should not be underestimated, but a 2001 report by the NIH indicated that when specific links to pharmaceutical developments were assessed for 47 drugs with US sales of $500 million or more per year, only four drugs had been developed in part with NIH-funded technologies. Domestic countries pharmaceutical R&D expenditures exceed those of any other major industrial sector, even high investment sectors such as computer software and services and the electrical, electronics, and aerospace industries.

Company-financed research in products affecting the CNS and sense organs was estimated at 7.3 billion in 2001,

significantly exceeding the $3.9 billion expenditure for agents acting on the cardiovascular system and roughly equal to the combined $7.4 billion expenditure on products affecting neoplasms, the endocrine system, and metabolic diseases. More than 80% of larger pharmaceutical companies developing agents to treat CNS disorders. They focus primarily on larger, more well-defined indications such as depression, schizophrenia, and multiple sclerosis, or on underserved indications in which medical need is high, such as dementias, brain tumors, or substance use disorders.

Geometric increases in R&D expenditures notwithstanding, the overall productivity of pharmaceutical research in producing new chemical entities has not increased in proportion to the investment. Over the past 20 years, pharmaceutical research has increased about 1,500% but the number of approvals of new therapeutic agents has been relatively small, rising from approximately 20 per year in the 1980s to approximately 30 per year in the 1990s and currently. This failure of productivity has been partially ameliorated by the wider markets opened by globalization and regulatory developments that facilitate more global registrations of treatments. But the pressures of productivity challenges are exacerbated by both the competition from "fast followers" and the challenge to generate new therapeutics fast enough to replace those that go off-patent. The ability of new technologies to rapidly close discovery gaps once a promising new target for development is proven greatly facilitates the capacity of fast-follower companies to exploit the discoveries of innovator companies.

This process is reflected in the progressively reduced period of marketing exclusivity and market share enjoyed by an innovator company before competitors match the initial breakthrough process. The huge economic importance of patent expiration is driven by the fact that generic production has taken an increasing share of the countries prescription pharmaceutical market. Once the patents constraining the generic use of a compound expire, the compounds are typically produced by a manufacturer with small costs to recover ($1 million to 2 million for bioequivalence studies) and no requirement to meet the significant costs required for the development of next-generation pharmaceuticals.

In a sense, the compound passes into the patrimony of humankind, where it is sold at markedly reduced prices that do not cover the costs of further research and development. Patent expiration is a huge challenge to innovator companies, in some cases engendering crisis and the risk of economic failure.

Many major pharmaceutical companies have recognized the need to double or triple their discovery output to maintain current profitability and growth in the face of generic competition. High throughput screening has been portrayed as a major and massive source of new compounds, yet analyses suggest that actual utilization is at only 2% to 7% of installed capacity and is not likely the rate-limiting step. Later issues such as biometabolism and compound toxicology are more important limitations to discovery output and are being managed by industrializing the screening process. Price increases are unlikely to play a significant role in compensating for the productivity challenge to the pharmaceutical industry.

Because most of the expenses associated with pharmaceuticals derive from R&D and marketing costs rather than the specific cost of goods production, prices tend to be driven by the relative benefit conferred, not the cost of unit production (although in general the cost of goods is much higher for protein therapeutics than for small molecule therapeutics and can constitute a significant portion of the total price). Price pressures are high.

Healthcare expenditures are a large budget item worldwide. In 1997, healthcare costs as a percentage of gross domestic product were higher in the Countrys than in other major industrialized nations. Pharmaceutical costs are a small percentage of overall healthcare costs, about 8% in the Countrys. On average, pharmaceutical costs are similar to the average telephone bill, but are of concern because they disproportionately affect vulnerable segments of the population such as the elderly.

In most of the world, pricing is tightly controlled by governments at levels that are, in aggregate, not compatible with sustaining current worldwide pharmaceutical R&D expenditures. These prices tend to cause pharmaceutical research to shift outside the borders of the countries with lower pricing and to decrease availability of therapeutics to patients by slowing introduction of new therapies.

It can be argued that pharmaceuticals may actually decrease overall healthcare costs by reducing larger expenditure items such as hospitalization. But the key strategic implication of this pressure for developers of neurotherapeutics and other drugs is the fact that increased costs of research and development are unlikely to be covered

by price increases, and therefore the need to dramatically improve productivity is inescapable.

ANTIFUNGAL DRUG THERAPY

Fungi, like mammalian cells but unlike bacteria, are eukaryotic and possess nuclei, mitochondria and cell membranes. Their membranes are unusual in containing distinctive sterols. The similarity between fungal and mammalian cells mitigates against selective toxicity and antifungal drugs are in general more toxic than antibacterial agents. The very success of antibacterial therapy has created ecological situations in which opportunistic fungal infections can flourish. In addition, potent immunosuppressive and cytotoxic therapies have produced patients with seriously impaired immune Defences, in whom fungi that are non-pathogenic to healthy individuals become pathogenic and cause disease.

AMPHOTERICIN B

Amphotericin is uniquely valuable in treating life-threatening systemic fungal infectins, but has considerable toxicity. It is given as an intravenous infusion freshly prepared in 5% dextrose over 4-6 hours or topically as lozenges (10 mg 3 hourly) or suspension for oral or esophageal and gastrointestinal moniliasis, respectively. There is some evidence that effective therapy may be achieved by reduced doses and therefore lower toxicity if amphotericin is combined with 5-fluorocytosine A test dose of 1 mg is given at least 6 hours before starting treatment. The initial dose is 5 mg daily in 500 ml of 5% dextrose which is increased daily by 5 mg to a dose of 0,5-1 mg/kg daily and continued for 6-12 weeks.

Treatment on alternate days at 1-1,5 mg/kg may reduce toxicity. The antifungal spectrum amphotericin B is broad.

Mechanism of Action

Amphotericin binds to a sterol in fungal cell membranes and increases their permeability, allowing leakage and loss of small molecules such as glucose and potassium ions.

Adverse Effects

- Fever, chills, headache, nausea and vomiting, and hypotension during intravenous infusion. Pulse and temperature should be monitored every 30 min and the infusion can be halted if necessary.
- Nephrotoxicity is almost invariable and results from vasoconstriction, tubular damage resulting in renal tubular acidosis, and acute renal failure. Fortunately most of these effects are reversible if detected early and the drug discontinued or the dose reduced.
- Hypokalemia.
- Normochromic normocytic anemia due to temporary marrow suppression is common.

Amphotericin B should not be withheld in serious progressive infection caused by a sensitive fungus despite toxic effects. Dose reduction may be appropriate.

Pharmacokinetics

Amphotericin *is poorly absorbed* following oral administration, therefore for systemic mycoses it must be given *by intravenous infusion. Liposomal delivery systems* deliver adequate plasma concentrations, with a lower incidence of toxicity. Given intravenously it distributes very

unevenly throughout the body: concentrations in the CSF are only one-fortieth of the plasma concentration. It is concentrated in the reticuloendothelial system.

NYSTATIN

Nystatin is another polyene antifungal antibiotic isolated from Streptomyces with an identifical mode of action to amphotericin B, but its *greater toxicity precludes systemic use*. Nystatin has *a broad antifungal spectrum*. Its indications are limited to *cutaneous and mucocutaneous* infections, especially those caused by *Candida* spp. which do not gain resistance to nystatin during therapy. Preparations of nystatin include tablets, pastilles, lozenges or suspension, given in doses of 100 000-500 000 units three times daily for oral or intestinal Candida infections. Patients often prefer amphotericin B because nystatin has an intensely bitter taste. Cutaneous infections are treated with ointment, vaginitis by suppositories and aerosol has been used for bronchopulmonary fungal colonization.

Adverse Effects

- Adverse effects seldom result from the topical use of nystatin.
- Large oral doses cause *nausea and diarrhea.*

Very little nystatin is absorbed from the gastrointestinal tract.

Griseofulvin

Griseofulvin was isolated from Penicillium griseofulvium. It is *systemically active*, but unlike amphotericin is useful only for mild infections since its spectrum is limited to *dermatophytes* (ringworm fungi).

Distance is not a problem. Treatment is given *orally* (0,5-1 g daily in two divided doses) with *meals.* Treatment should be for 6 weeks in skin infections and up to 12 months for nail infections.

Mechanism of Action

Griseofulvin is *actively taken up by fungi*. Its mode of action is obscure, but it binds to the *microtubules* that form the mitotic spindle and *blocks polymerization* of the microtubule. It also interferes with fungal DNA replication.

Adverse Effects

- Headaches and mental dullness or inattention (uncommon).
- Diarrhea or nausea (uncommon).
- Rashes

Pharmacokinetics

Griseofulvin is *nearly insoluble in water* and is formulated as *micronized particles*. Its absorption is *facilitated by a fatty meal*. It has a slow onset of action because it must first be taken up into slow *growing keratinized structures to reach the site of infection*. Cell turnover time thus determines efficacy of treatment so that palmar and *plantar skin requires at least 8 weeks´ treatment, fingernails 6 months and toenails up to 1 year for eradicative treatment.*

Drug Interactions

Griseofulvin induces hepatic cytochrome P_{450} enzyme activity and consequently interacts with warfarin reducing its anticoagulant effect.

Flucytosine (5-fluorocytosine)

Flucytosine is used for systemic candidiasis and cryptococcosis, providing the strain is sensitive. The optimal oral dose is 200 mg/kg/day in 6-hourly divided doses. For very ill patients an intravenous preparation is available. *It is used in combination therapy with amphotericin.*

Topical use is unacceptable because of the danger of widespread emergence of resistant Candida spp. Its spectrum is relatively restricted to Cryptococcus neoformans, Candida albicans and some other Candida spp., Torulopsis spp. and Cladosporum spp. There are big differences in sensitivity between strains: 5-15% have innate resistance, and resistance is relatively easily acquired during therapy. Filamentous fungi, especially Aspergillus, are resistant.

Mechanism of Action

Flucytosine enters the fungus by active transport. It is deaminated to *5-fluorouracil*, a known antimetabolite that inhibits thymidylate synthetase thereby depressing DNA synthesis. Its relative specificity is due to the presence of cytosine deaminase in fungi but not in mammalian cells.

Adverse Effects

Flucytosine is less toxic than amphotericin B.

- Gastrointestinal upsets.
- Leukopenia.
- Hepatitis may occur so liver function tests should be monitored.

At plasma concentrations below 100 mg/ml there is little danger of toxicity. Depression of bone marrow and

hepatotoxicity are associated with higher concentrations. Plasma drug assays are useful in patients with impaired renal function.

Pharmacokinetics

Flucytosine is well absorbed from the gut, peak concentrations of 75-90 mg/ml being attained on a dose of 50 mg/kg 6 hourly. It penetrates adequately into CSF, in contrast to amphotericin B, and is consequently particularly useful when combined with amphotericin in treating cryptococcal meningitis.

Drug Interactions

Amphotericin acts additively or synergistically with flucytosine, and flucytosine should always be used with amphotericin.

This combination is useful because amphotericin is more effective but more toxic than *flucytosine which also penetrates the blood-brain barrier* better than amphotericin. *Concurrent use of amphotericin reduces the likelihood of resistance to flucytosine emerging during therapy*.

Imidazoles

Imidazole antifungal drugs are *fungistatic* at low concentrations and *fungicidal* at higher concentrations.

They are *used topically* and are active against both *dermatophytes* and *yeasts* such as Candida. Some imidazoles are also used systemically although they have limited efficacy and significant toxicity, limiting systemic use. They act similarly to one another by inhibiting *fungal ergosterol synthesis*. Ergosterol is an important constituent of fungal

membranes. Imidazoles inhibit lanosterol 14a-demethylase (which is a fungal cytochrome P_{450} enzyme) and have considerable specificity for fungal cytochromes. Membrane leakage and dysfunction of membrane adenosine triphosphatase (ATPase) ensue.

KETOCONAZOLE

Ketoconazole is used *as oral or topical* therapy for dermatophytic infections and some phycomycetes. It is active against systemic infection with Candida,

Blastomyces, Histoplasma capsulatum and Cryptococcus neoformans. Aspergillus and Mucor spp. are resistant. It is given *orally* (200-400 mg once daily). Its systemic use has waned because of the high incidence of hepatic and endocrine side effects.

Adverse effects

- *Nausea and vomiting*, reduced by giving the drug with food.
- *Transient liver function* abnormalities in 5-10% of patients, fulminant hepatic damage, jaundice and fever are rare.
- *Gynecomastia* (by blocking testosterone synthesis).
- Impotence and azoospermia.
- *Adrenal insufficency* (by inhibiting cortisol biosynthesis).

Pharmacokinetics

Ketoconazole is given *orally* and achieves maximum plasma concentrations in 1-2 hours.

Drug interactions

- Antacids and H_2 -blockers reduce ketoconazole absorption
- Ketoconazole should not be used with amphotericin B because it reduces its effectiveness.

CLOTRIMAZOLE

This imidazole is only used for topical treatment of Candida or dermatophyte infections as a 1% cream, solution of powder. It is *poorly absorbed* from the gastroinstestinal tract and *induces its own metabolism, therefore it is not used systemically.*

MICONAZOLE

Miconazole (2% cream or powder applied twice daily) is usually used topically to treat cutaneous Candida, ringworm and pityriasis rosea.

Triazoles

This group of drugs (e.g. fluconazole) is derived from the imidazoles. They are nitroimidazoles and have a *wider antifungal spectrum*. Their mechanism of action *is idential* to that of other imidazoles (e.g. ketoconazole), by inhibition of lanosterol a-demethylase.

FLUCONAZOLE

Fluconazole is a *potent and broad-spectrum* antifungal agent. It may be *given orally or intravenously as a once daily dose*. For superficial infections it is given as 50-100 mg/ day. In systemic or meningitic infections 200-400 mg *intravenously daily* is required for 4-6 weeks, followed by a

daily maintenance dose. For prophylaxis in cytotoxic immunosuppressed patients 50-100 mg is adequate.

Adverse Effects

- Gastroinstestinal upsets with nausea, abdominal distension, diarrhea and flatulence.
- Skin rashes - erythema multiforme.
- Hepatitis - raised liver enzymes.

Contraindications

Fluconazole is *contraindicated in pregnancy* because of fetal defects in rodents. Breast milk concentrations are similar to those in plasma, and fluconazole should not be used in *nursing* mothers.

Pharmacokinetics

Fluconazole is absorbed rapidly after oral administration with maximum plasma concentrations achieved within 1-2 hours.

Absorption is virtually *complete* and is unaffected by food or gastric pH. Fluconazole *is widely distributed throughout the body penetrating the CSF well.*

ITRACONAZOLE

Itraconazole *is similar to fluconazole in its antifungal spectrum and mechanism of action. It is* given orally once daily as a 100 mg dose to treat *dermatophyte infections, superficial oropharyngeal candidiasis and pityriasis versiColour.* Its major side effects are gastrointestinal disturbances. It causes several adverse drug interactions, increasing cyclosporin and warfarin plasma concentrations.

TIACONAZOLE

Tiaconazole is a *broad-spectrum antifungal triazole*, marketed only as a solution for topical administration *to treat nail infections with dermatophytes and yeasts*. No other topical imidazole preparations are effective in fungal nail infections. The recommended duration of therapy is 6-12 months.

Terbinafine

Terbinafine is fungicidal. It can be given orally and is used to treat ringworm (tinea pedis, cruris or corporis) or dermatophyte infections of the nails if oral therapy is considered appropriate. It is given 250 mg once daily for 2-6 weeks or longer in infections of the nailbed as an alternative to griseofulvin. It acts by inhibiting sterol synthesis by the fungal enzyme squalene epoxidase. It *is well absorbed*, strongly bound to plasma proteins and *concentrated* in *the stratum corneum*. Its major side effects *are nausea, abdominal discomfort, anorexia, diarrhea and rashes (including urticaria).*

5

Calorimetry in Pharmaceutical Research

Calorimetry is the measurement of energy changes within a material that are either manifested as exothermic (heat liberating) or endothermic (heat consuming) events. Changes in energy (not absolute energies) are conventionally determined, and quantitative measurements may be made if the mass of the sample(s) is accurately known.

Recently, nanocalorimetry or calorimetry microarrays are expanding the application of calorimeters to high throughput screening (HTS), providing high throughput thermodynamic measurements at the microgram scale.

Preliminary studies using these microchip calorimeters have shown interesting potential for their applications in studying biological macromolecules in solution, such as protein ligand binding and measuring heat capacities on samples as small as 10 *mg*. Additional applications could

include the study of the thermal properties of drug molecules in HTS. The most common applications of calorimetry in the pharmaceutical sciences are found in the "subfields" of differential scanning calorimetry (DSC) and microcalorimetry. State-of-the-art DSC instruments and microcalorimeters are extremely sensitive and are powerful analytical tools for the pharmaceutical scientist.

Differential scanning calorimetry usually involves heating and/or cooling samples in a controlled manner, whereas microcalorimetry maintains a constant sample temperature. The DSC instruments are considered to be part of the "Thermal Analysis" armamentarium; for additional information, the reader should refer to the Thermal Analysis section of this Encyclopedia.

BACKGROUND

The beginning of this article gives a brief introduction to thermodynamics. A description of DSC, which includes instrumentation, calibration, and applications, follows. A section on microcalorimetry is next, with a brief introduction into microcalorimetry, instrumentation, calibration, and applications. The article ends with a general comment on the regulatory aspects of calorimetry. A general description of the underlying physical or chemical transitions/reactions can be found in the section on DSC.

THERMODYNAMICS

The field of calorimetry relies on the principles of thermodynamics; the next section provides a brief overview of the general principles. References are provided for those who are unfamiliar with thermodynamics. A calorimeter

consists of a container that is isolated from its exterior surroundings, where the heat exchange that occurs between the system and the environment can be measured.

The "environment" is defined as the calorimeter and its contents, and the "system" is either a chemical reaction or physical change of state.

The system can either absorb (endothermic) or lose energy (exothermic) to or from the environment. Exothermic changes will require the temperature of the environment to increase because it is the environment that is receiving the energy lost by the system.

The energy of an isolated system remains constant and the energy exchange of the system must be equal but opposite in sign to the energy of the environment (First Law of Thermodynamics, Conservation of Energy).

Endothermic changes in the system will involve a decrease in temperature of the environment because the environment is providing the energy absorbed by the system.

Based on the assumption that the system is closed, which is usually the case in DSC and microcalorimetry, any reaction or change in state is independent of the path and can be subdivided into small reversible steps (Hess' Law of Summation).

The First Law of Thermodynamics states that energy may neither be created nor be destroyed. It defines the internal energy, dU, as the sum of the change in heat that has been transferred to the system, dq, and the work done on the system, dw.

$$dU = dq + dw$$

Table. Common Thermal Events that can be Detected Using Calorimetric Techniques.

Event	Example
Endothermic	
Fusion	Melting of drug substances; purity evaluations
Vaporization	Evaporation of liquid or semisolid excipients
Sublimation	Removal of frozen water during lyophilization
Desorption	Drying of wet granulated formulations
Desolvation	Removal of stoichiometric water from crystalline hydrates
Exothermic	
Crystallization	Solvent vapour induced crystallization of amorphous excipients
Precipitation	Formation of salt forms of drug substances
Solidification	Melt granulation with semisolid excipients
Adsorption	Solvent vapour sorption by drug substances
Chemisorption	
Solvation	Water vapour sorption by excipients
Curing of resins	Curing of polymeric packaging materials
Other	
Glass transition	Variation of glass transition temperature with water content
Relaxation of	Enthalpic recovery of amorphous drug glasses substance upon storage or annealing
Decomposition	Thermal decomposition of drug substance
Dissolution	Dissolving drug substance in dissolution media
Complexation	Complex formation between drug and cyclodextrin

When operated at constant pressure, Equation can be written in terms of the enthalpy, H. The total energy exchange between the system and the environment, the enthalpy change (dH), is the sum of the change in internal energy of the system, dU, and the change in the amount of work, PV $dH = dU + PdV\ (at\ con\tan t\ \Pr essure)$

At zero net work and negligible change in volume (a close approximation for solids and liquids), the equation reduces to

$$(dU)_P = (dH)_P = (dq)_P$$

Thus, the enthalpy is effectively equal to the heat added or lost from the system, and changes in enthalpy can be measured directly in a calorimeter as dq (heat flow). The heat

exchange, dq, entering or exiting the system is equal to the change in enthalpy, dH, which is related to the heat capacity,

$$Cp \quad dq = dH \int_{T_2}^{T_1} Cp dT_1 .$$

The increase in temperature of the system (from T1 to T2) is a function of its heat capacity. If Cp is large, then the transfer of a given amount of heat to a system results in only a small temperature increase. Two principal DSC designs are commercially available— power compensated DSC and heat flux DSC. The two instruments provide the same information but are fundamentally different. Power-compensated DSCs heat the sample and reference material in separate furnaces while their temperatures are kept equal to one another. The difference in power required to "compensate" for equal temperature readings in both sample and reference pans are recorded as a function of sample temperature.

Heat flux DSCs measure the difference in heat flow into the sample and reference, as the temperature is changed. The differential heat flow to the sample and reference is monitored by chromel/constantan area thermocouples.

MODULATED TEMPERATURE DIFFERENTIAL SCANNING CALORIMETRY/DYNAMIC DIFFERENTIAL SCANNING CALORIMETRY

Conventional DSC measures a sample's total heat flow. This total heat flow is comprised of a heat capacity component and a kinetic component in that a sinusoidal modulation is overlaid on the conventional linear heating or cooling rate to produce a continuously changing non-linear sample temperature.

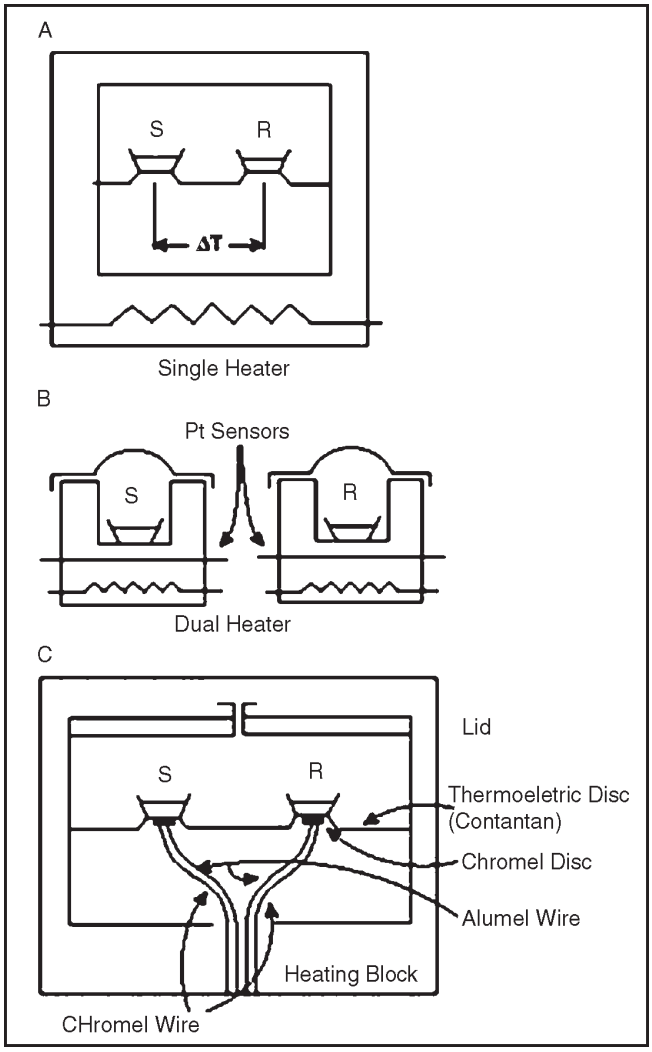


Fig. Schematic Diagrams of the (A) Differential thermal Analysis (DTA); (B) Power-compensated DSC; and (C) Heatflux DSC Cells.

$$\text{total heat flow} = \text{heat capacity Component} + \text{kinetic component}$$

$$dq/dt = CpdT/dt + f(T,t)$$

This can be viewed as running two experiments at once. The first experiment consists of heating the sample at a constant linear rate to obtain the total heat flow much like the conventional DSC.

During the second experiment, the heat capacity component of the heat flow is obtained by continuously varying the temperature sinusoidally with a zero net temperature change during the course of the modulation.

The experimental parameters may be optimized by modifying three variables—the average heating rate, the period of modulation, and the temperature amplitude of modulation. Fourier transformation of the modulated heat flow signal is used to calculate an average heat flow value, which is similar to the total heat flow obtained by conventional DSC.

The heat capacity is determined by the ratio of the heat flow amplitude to the modulated heating rate amplitude. The heat capacity of heat flow is then obtained by multiplying the heat capacity by the average heating rate. The kinetic component heat flow is obtained by the difference between the total heat flow and the heat capacity component.

$$dq/ = Cp\left(dT/dt + A_{tw^{*}COSwt}\right)$$
$$+ f'(t,T) + Ak(\sin wt)$$

where $(dT/dt + A_T w * \cos wt)$ = measured heating rate, f0(t,T)= kinetic response without temperature modulation, and AK = amplitude of kinetic response to temperature modulation.

DDSC provides heat capacity and kinetic component

Information differently from MDSC. The temperature programme consists of an "Iso-Scan" whereby the traditional heating rate programme is combined with several isothermal holds or a "Heat-Cool" programme, which consists of combined heating and cooling temperature programs.

The user selects the appropriate method depending on the type of experiment being performed. From the dynamic

component of the sample response, the complex heat capacity can be calculated. The complex heat capacity, Cp_, is the vector sum of the storage, Cp0, and loss heat capacity, Cp00. It is generally the same as the storage heat capacity except in the melting region where heat losses dominate.

The storage heat capacity is associated with molecular motions within the sample in a manner similar to the storage modulus in dynamic mechanical measurements. The out-of-phase component, the loss heat capacity, Cp00 is associated with the dissipative properties of the material.

The loss heat capacity is out-of-phase with the temperature change because heat flow has resulted in molecular structural changes in the material. The loss tangent is the ratio of the loss heat capacity to the storage capacity and is a measure of the relative importance of each component.

$$Cp^* = Cp' + Cp''$$

where Cp* = complex heat capacity, Cp' =storage heat capacity, and Cp'' = loss heat capacity. Some of the advantages and disadvantages of using MDSC are given in Table 2.

High Speed DSC

High speed DSC or HyperDSC2 is a proprietary technology developed by Perkin Elmer to be used with their power-compensated DSCs. It enables the use of very fast heating and cooling rates (100–500K/min) that provides increased sensitivity with the compromise of reduced resolution.

In the pharmaceutical industry, HyperDSC has been used for the detection of very small signals arising from weak transitions or small sample sizes (*m*g range). For example, some researchers have used HyperDSC to detect 1.5%

amorphous content in amorphous/crystalline lactose blends. Using a fast heating/cooling rate also enables the measurement of samples while minimizing the potential for recrystallization or reorganization. The thermal properties of two polymorphs of the drug carbamazepine, Forms I and III, were studied using HyperDSC. Previously, accurate determination of the heat enthalpy of fusion of Form III had not been possible using conventional heating rates owing to concurrent exothermic recrystallization to the highermelting Form I. The use of HyperDSC enabled the measurement of the heat of fusion by altering the kinetics of melting, where it was inhibited.

Owing to its speed of analysis, it has also been proposed that HyperDSC can be used as a HTS tool for the analysis of well-understood samples.

Sample Preparation and Calibration

DSC samples are generally analyzed in small metal pans that consist of inert or treated metals (aluminum, platinum, silver, stainless steel, etc.). Several pan configurations exist such as open, pinhole, covered, or sealed. Reference pans should be made of the same material as the sample pan and in identical configurations. Typical DSC sample sizes are 3–5 mg for pharmaceutical materials. The material should completely cover the bottom of the pan to ensure good thermal contact. The pan should not be overfilled to prevent thermal lag from the bulk of the material to the sensor.

Physically stable compounds that consist of large granular particles should be ground to reduce unwanted thermal effects. Accurate weights are imperative if quantitative data of the sample's energetic parameters are desired.

A scanning technique that is usually used for relative rather than absolute measurements is DSC. The meaningfulness of the results depends on the care taken in calibrating the instrument as close to the transition temperatures of interest as possible.

The accuracy of any thermoanalytical instrument is strongly dependent on the use of high purity calibration standards. Well-defined standards are especially important when analyses are carried out using different instruments and at different times. In general, metal calibration standards such as indium, tin, bismuth, and lead are utilized owing to ready availability and ease of use.

Table. Advantages and Disadvantages of Modulated DSC

Advantages	**Disadvantages**
Ability to differentiate overlapping transitions	More complex thermal lag effects
Increased resolution without loss of sensitivity	Not as precise linear heating rate
Measurement of heat capacity and heat flow in a single experiment	Many experimental parameters
Measurement of initial	Not recommended for melting crystallinity transitions
Ability to study previous history	Gives sample a complex thermal thermal history
Ability to distinguish between reversible and nonreversible transitions	Sometimes difficult to interpret

Low melting metals such as mercury and gallium, are used to a lesser extent because of toxicity and handling problems. Organic compounds have been recommended as standards when studying organic material to minimize differences in thermal conductivity, heat capacity, and heat of fusion.

It is likely that metals will continue to be popular temperature and enthalpy standards because of availability and ease of use, and organic standards may be used predominantly at temperatures below 300 K.

The results of the DSC are dependent on the calibration of the instrument, sample preparation, and sample configuration. Some researchers argue that power compensated DSCs need to be properly calibrated upon both heating and cooling at the same rates to maintain a high level of accuracy.

Standard procedures can be obtained from the American Society for Testing of Materials (ASTM). In addition, all results obtained by DSC are a function of the scanning rate used and should be reported with the scanning rate.

The shape, the area of the transition or change of baseline, and particularly the temperature of the transition will be dependent on the scanning rate; it will move to higher temperatures with increasing heating rate.

Increasing the scanning rate increases sensitivity, while decreasing the scanning rate increases resolution. To obtain thermal event temperatures close to the true thermodynamic value, slow scanning rates should be used (e.g., 1–5K/min).

Definitions and Applications of DSC

The purpose of this section is to define the various parameters that are measured by DSC. The types of thermal events, exothermic or endothermic, that can be measured by DSC are reported in Table. The following sections will describe some of the more fundamental thermal events.

Examples from the pharmaceutical field will be given to illustrate the techniques. The examples will be based on either

single components such as drug substance and bulk excipients or on a mixture of components such as physical blends of drugs and excipients, solid dispersions, formulated drugs after granulation, and/or compression.

Melting

Melting is a first order endothermic process by which the compound takes in a net quantity of heat (molar heat of fusion). Through DSC, melting can be seen as an endothermic peak.

The broadness of the peak defines the purity of the crystalline compound undergoing melting, with the less pure and less perfect smaller crystals melting first followed by melting of the purer larger crystals.

The melting temperature is the temperature at which the three-dimensionally ordered crystalline state changes to the disordered liquid state. It is defined either as an extrapolated melting temperature onset, Te, obtained at the intersection of the extrapolated baseline prior to the transition with the extrapolated leading edge, or as the peak melting temperature, Tm.

Other temperatures that describe the melting process are the onset of melting, To, and the extrapolated end of the transition. The enthalpy of fusion, *D*Hf, is obtained fromthe area of the endothermic transition.

The area of the transition is affected by the selection of the baseline. The baseline is generally obtained by connecting the point at which the transition deviates from the baseline of the scan to where it rejoins the baseline after melting is completed.

For some materials that undergo a significant change in heat capacity change on melting, other baseline approximations (such as a sigmoidal baseline) are used.

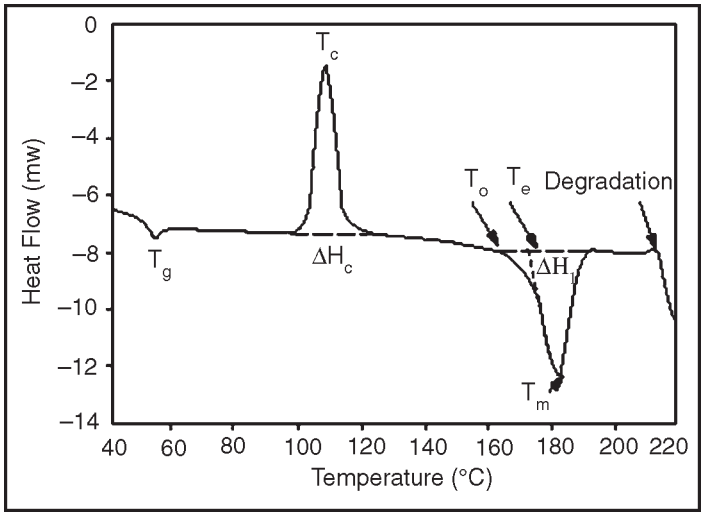


Fig. DSC Scan of Sucrose Showing the Glass Transition Temperature, (Tg), Recrystallization Exotherm Temperature (Tc) and Enthalpy (DHc), Onset of Melting (To), Extrapolated Melting Onset (Te), Peak Melting Temperature (Tm) Enthalpy of Fusion (DHf), and Onset of Degradation at 10 K/min. Endothermic transitions are Shown.

Generally obtained by connecting the point at which the transition deviates from the baseline of the scan to where it rejoins the baseline after melting is completed.

Purity

The purity of crystalline compounds can be calculated using the van't Hoff equation from the enthalpy of fusion and melting temperature obtained by DSC.

$$T_{s(i)} = T_c - RT_e^2 X / (\Delta H_f F_i)$$

where Ts(i) is the sample temperature at equilibrium corrected for thermal lag effects (K), Te is the melting temperature of the pure compound (K), R is the gas constant (8.314 J/mol/K), X is the molar fraction of impurity, DHf is the enthalpy of fusion of the pure compound (J/mol), and F is the fraction of the

sample that is molten at Ts(i). The melted fraction is equal to the area of the section melted (Ai) divided by the total area of the melting endotherm (AT) as shown in Figure. The melting depression, (Te - Ts(i)) is equal to the slope, (RT2 e/DHf) X, of the straight line obtained when Ts(i) is plotted as a function of 1/Fi.

The theoretical melting temperature is obtained on extrapolation to 1/Fi = 0. A straight line may not be obtained owing to thermal lag, sensitivity, lack of a eutectic point detection, and formation of solid solutions. In addition, a significant amount of material may have melted before a measurable heat flow is observed by using DSC.

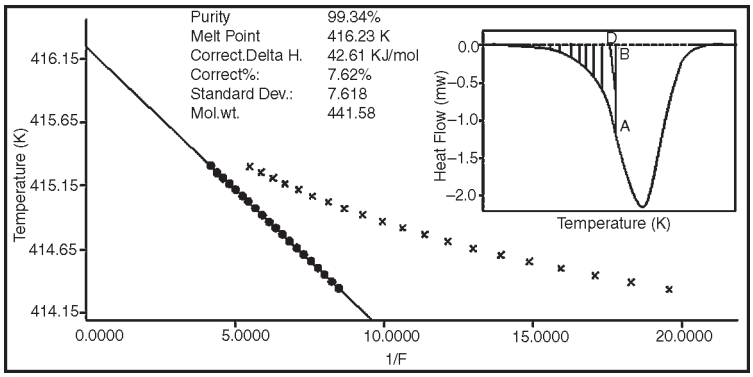


Fig. DSC Scan of Drug Substance Divided Into Segments, Ai, for Purity Calculations of a Compound of Total Enthalpy or Area, AT (Inset, only a few Segments are Shown).

The van't Hoff plot of the temperature of each segment as a function of 1/F = 1/(Ai/AT), and with correction (straight line) for determination of purity of the drug substance is shown. As a result, a correction constant Kcorr is added to the measured areas (each fraction) to correct the curvature of the plot of Ts(i) as a function of 1/Fi . The melting depression (Te - Ts(i)) is then obtained when Fi =1

$$1/f_i = (A_T + K_{corr})/(A_i + K_{corr})$$

It is necessary that the melting curve is obtained with a calibrated DSC using small samples (1–3mg) and slow scanning speeds (<5K/min, preferably 2K/min). The purity of a compound should be determined at several scanning speeds to ensure that the compound does not undergo any solid–solid transitions, such as polymorphic conversion or degradation.

The three advantages of obtaining purity by DSC are:

- Its speed of measurement;
- The type of impurity does not have to be known;
- A minimal amount of sample is required.

However, in the case of salts, excess base or acid is counted as an impurity. Calculations of drug purity using DSC were used to assess the quality of progesterone and lipoic acid, and to define specifications for the drugs. The enantiomeric purity of (-)ephedrinum 2-naphthalenesulfonate has also been determined using DSC.

Crystallization

Crystallization can occur on cooling from the melt and/or heating above the glass transition temperature of amorphous materials. The temperature at which this occurs is the crystallization temperature, Tc. Through DSC, crystallization is observed as an exothermic transition with an enthalpy of crystallization, *D*Hc.

The energy released when the molecules, atoms, or ions organize into a 3-D solid state is related to the crystal lattice energy. Some compounds can crystallize into different molecular arrangements called polymorphs, discussed later.

Quantification of Crystallinity

The crystallinity of drugs and excipients before and after formulation processing can be determined using calorimetry. In some cases, crystalline compounds can be converted during pharmaceutical processing to the amorphous form, which is a thermodynamically less stable form.

Amorphous compounds consist of non-ordered molecules (see the section on Glass Transition). This can have important implications for the chemical and physical stability of the formulations. The effect of grinding on the crystallinity of different crystal forms of indomethacin was evaluated using DSC and other techniques.

An exothermic transition can sometimes be observed by DSC on crystallization of the amorphous form. This can be used to quantify the amorphous content of crystalline drugs.

A calibration curve that consists of a plot of the enthalpy of crystallization as a function of crystalline content was used to determine if the lyophilized MK-0591 drug substance was completely amorphous or contained some crystalline compound.

Polymorphism

Polymorphs are crystalline compounds of the same molecular structure that have a different arrangement of molecules in the unit cell. Polymorphs have the same chemical composition but have unique cell parameters.

Therefore, polymorphs can have very different melting temperatures, densities, solubilities, chemical and physical stabilities, dissolution rates, and bioavailabilities. Polymorphs are either enantiotropic or monotropic.

Enantiotropic polymorphs have a thermodynamic conversion temperature where one form is more stable above this temperature while the other is more stable below this temperature.

Processing the least stable form, dissolution/ recrystallization, and certain storage conditions might cause enantiotropic polymorphs to later convert.

If there is no conversion temperature below the melting temperatures of the polymorphic pair, then the different crystal forms are monotropic. That is, there is only one crystal form that is thermodynamically stable at all temperatures and pressures.

Calorimetry can be used to determine which polymorph is the more stable form. The DSC can provide accurate unambiguous melting temperatures and enthalpies of fusion.

Based on the melting temperature and the enthalpy of fusion, the relative thermodynamic stability of the polymorphic pair can be determined (e.g., using the Heat of Fusion Rule).

DSC and complimentary thermal techniques such as temperature X-ray powder diffraction were used to determine the thermodynamic relationship of the six anhydrous polymorphs of tetracaine hydrochloride. The phase diagram of the polymorphic conversion of diflunisal in polyethylene glycol 4000 solid dispersions was obtained as a function of polymer content.

Heat Capacity

Accurate heat capacity, Cp, measurements may be obtained by DSC under strict experimental conditions, which

include the use of calibration standards of known heat capacity, such as sapphire, slow accurate heating rates (0.5–2.0 K/min), and similar sample and reference pan weights. MDSC or DDSC also have been used to determine the heat capacity of several pharmaceutical materials.[8,24]

Glass Transition

By the use of various pharmaceutical manufacturing processes, (e.g., lyophilization or comminution techniques), drugs or excipients may be made amorphous. Amorphous compounds are defined by their lack of long-range molecular order and structural periodicity.

Their high-energy state is of great interest to the pharmaceutical industry as it can lead to fast dissolution rates and increased bioavailabilities.

However, amorphous compounds are thermodynamically unstable, although depending on their glass transition temperature, they may be kinetically stable for extended times. Amorphous compounds are characterized by a glass transition, which by DSC is seen as an increase in heat capacity change.

$$\Delta Cp = Cp_{liq} - Cp_{glass}$$

where Cpliq is the heat capacity of the liquid, and Cpglass is the heat capacity of the glassy phase. The glass transition temperature is measured either at its onset or at its midpoint as shown in Figure. Structural relaxation can occur owing to the restricted but finite mobility of the molecules below the glass transition.

This gradual volume or enthalpy change is observed by DSC as an endothermic peak superimposed on the glass

transition, and this may lead to difficulties in interpretation of the transition. Modulated temperature DSC can sometimes be used to separate the enthalpic overshoot from the glass transition temperature.

Defining the glass transition temperature is important to the development of stable amorphous pharmaceutical materials.

A leukotriene biosynthesis inhibitor, MK-0591, has been shown to be kinetically stable in the amorphous phase at normal storage temperatures if protected from moisture, because of its elevated glass transition temperature of 125°C.

In lyophilized systems, a high Tg0, defined as the apparent glass transition temperature and observed as the change of the heat capacity of the lyophilized formulations, is important to define the stability of such formulations. MDSC has been used to select optimal freeze-drying conditions to avoid cake collapse.

The glass transition temperature of amorphous multicomponent mixtures can be used to determine the miscibility of the components. If the mixture is miscible, then a single glass transition temperature is usually obtained. Various equations can be used to predict the glass transition temperature of miscible mixtures. Examples include the Gordon–Taylor equation or the Fox–Flory equation.

$$T_{gmix} = [w_1 T_{gl} + K w_2 T_{g2}] / (w_1 + K w_2);$$

$$K = \rho_1 Tgl / \rho_2 T_{g2}$$

$$1/T_{g\,mix} = 1/T_{gl} + 1/T_{g2}$$

where ρ1 and ρ2 are the densities of the two components and Tg1 and Tg2 are their respective glass transition temperatures.

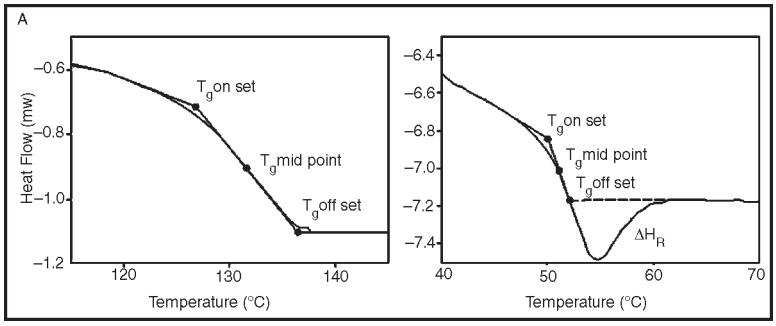


Fig. (A) DSC Scan of the Glass Transition Temperature of a Miscible Blend of a MK-0591 with 10% PVP, Showing the Onset, Midpoint, and Offset Glass Transition Temperatures. (B) DSC Scan of the Glass Transition of Sucrose with the Enthalpic Relaxation Endotherm and Enthalpy (ΔHR).

Temperature Dependence of Molecular Motions in Amorphous Materials

A critical attribute that dictates the stability and performance of any amorphous material is the manner in which its rate of molecular motions (*t*) varies with changing temperature (T) (i.e., d*t*/dT).

At temperatures that are approximately 0–100K above the calorimetric glass transition temperature (Tg), this property is known as the fragility of the material.

Several workers have suggested that dt/dT below Tg (in the non-equilibrium glassy state) is the most appropriate descriptor of amorphous pharmaceutical materials as this is the normal state for the storage and processing of such systems.

A simple graphical plot of t vs. T can be constructed at temperatures below Tg from the results of enthalpy relaxation experiments. These measurements can be performed using either a conventional DSC or a microcalorimeter. Alternate calorimetric methods of estimating dt/dT at Tg have been

described in the literature, but the applicability of these methods to pharmaceutical materials has not yet been clearly demonstrated.

Degradation, Decomposition, Stability Determinations, and Drug-Excipient Compatibility

The degradation, decomposition, and stability of drugs, or formulations can be determined by DSC or microcalorimetry. The advantages of the techniques are their speed of measurement and the small amounts of sample required.

At times, interpretation of the results can be difficult, particularly when simultaneous reactions occur. Decomposition kinetics is generally determined using the Arrhenius equation. The samples are stored at elevated temperatures for known periods of time and analyzed by DSC. Alternatively, they can be held isothermally in the DSC at different temperatures, followed by scanning at heating rates sufficiently fast to avoid additional decomposition.

A rate constant is calculated for each storage condition by plotting the logarithms of the areas of the transitions (e.g., the decomposition endotherm, etc.) as a function of time. The natural logarithm of the reaction rates, k, are then plotted as a function of 1/T as per the Arrhenius equation.

$$k = Ze^{-(Ea/RT)}; \ln k = \ln k - E_a / RT$$

where Z is the Arrhenius frequency or preexponential factor, Ea is the Arrhenius activation energy (J/mol) for the reaction, and R is the gas constant. The activation energy and preexponential factor are assumed to be constant and independent of temperature.

Alternatively, the reaction peak maxima may be determined at different heating rates (*j*) and used to calculate the activation energy, assuming first order kinetics.

$$E_a = -2.19\,T\,\frac{d\log\varphi}{d(1/T)}$$

The energy of activation is obtained from the slope of the log of the heating rate (φ) as a function of 1/T. It is assumed that only one reaction occurs during the transition and that the peak maximum represents a point of constant conversion for each heating rate.

The method cannot be used with compounds that decompose on melting or undergo isomerizations at the reaction temperature or any other simultaneous reaction. Some modification of DSC may be needed to determine the degradation kinetics of compounds under different environmental conditions.

DSC is often used for the rapid screening of excipients for drug-excipient compatibility studies. Certain assumptions have to be made, which include that the thermal properties of these mixtures are the sum of the individual components when there are no interactions between the components.

The method does not take into consideration:

- Effects owing to thermal conductivity (thermal lag effects);
- Mixing effects that can lower the purity of each component resulting in slightly broader, lower melting temperatures;
- Sample geometry effects that result in variations in peak shapes and peak temperatures.

In addition, reduction in enthalpies of fusion can occur as a result of the solubilization of the drug in molten excipients.

This latter phenomenon can be used in part to determine the solubility in different molten excipients.

Interactions with Water/Solvents, Hydrates

Water can have a significant impact on the physical and chemical stability of drugs. Water may be present as part of the crystalline lattice (hydrate), or it may be on the surface ("free") or more tightly incorporated ("bound").

The evaluation of the type of water present in a pharmaceutical material has been determined using subambient DSC (thermoporosimetry), such as in the case of magnesium stearate hydrates, as well as thermogravimetric techniques.

Free or surface water can crystallize and the melting enthalpy of this free water can be used to calculate the surface water content of compounds from the melting enthalpy of pure water. The state of water in hydroxypropyl methylcellulose gels with and without drugs such as propranolol hydrochloride or diclofenac was determined in this way by DSC.

MICROCALORIMETRY

Microcalorimetry is used to monitor thermal changes associated with physical and/or chemical events that do not require heating or cooling for their initiation. Such events include dissolution, precipitation, reaction, and crystallization.

In a typical microcalorimetry experiment, these events are ''triggered'' in a controlled manner by mixing two preequilibrated and separate phases (e.g., water vapour and amorphous drug, solvent and crystalline drug, or protein and carbohydrate solutions). Solid-state processes may also

be measured such as in the case of drug degradation or during drug-excipient screening studies.

Microcalorimetry techniques are sometimes referred to by the processes that are monitored (e.g., immersion calorimetry, solution calorimetry, titration calorimetry, etc.). Differential scanning calorimeters may be operated in isothermal mode; however, for highly accurate and reliable isothermal measurements, specially designed microcalorimeters are required.

Thermodynamics

Microcalorimeters have the ability of directly measuring the order of the reaction (n), the rate constant (k), the reaction enthalpy (ΔRH), and the equilibrium constant (Keq).

For example, solution microcalorimetry may be used to determine the free energy of dissolution of a solid compound, which is particularly important in pharmaceutical research for dissolution studies and in the determination of the relative thermodynamic stability of polymorphs. The change in the Gibbs–Helmholtz free energy, ΔGsol, on dissolution is

$$\Delta G_{sol} = RT \ln k_{eq}$$

where T is the temperature (Kelvin, K), R is the gas constant (8.314 J/mol/K), and Keq is the equilibrium constant for the change of the compound from the solid state to the dissolved liquid state.

The equilibrium constant can be determined, at low concentrations, from the ratio of the concentration of the compound in the solution, or its solubility, to that in the solid state

$$K_{eq} = [C]_{soln} / [C]_{solid} = [C]_{soln}$$

The ΔHsol is the enthalpy change that occurs on dissolution of one mole of compound in a solvent. The solution microcalorimeter may be used to obtain the enthalpy of solution directly.

The change of free energy can be calculated from the concentration using the enthalpy obtained. The change in the entropy of solution ΔSoln can then be determined from the Gibbs–Helmholtz equation. Alternatively, the change in free energy of solution *D*Gsoln can be calculated from the van't Hoff equation

$$\delta\left(\Delta G_{soln}/T\right)/\delta T = -\Delta H/T^2$$

Additionally, in cases where the reaction of interest is monitored at at least three different temperatures, the activation energy (Ea) of that reaction may be determined. Microcalorimeters can offer scientists a large range of important information provided that their systems are sufficiently well understood.

Signals arising from more than one process may compromise quantitative results if they are not sufficiently separated. There are cases where concomitant processes, if occurring at sufficiently different reaction rates (at least 2X), may be quantitatively analyzed using an iterative procedure that is described in more detail by Skaria et al.

Instrumentation

The simplest type of non-scanning calorimeter is the isoperibol instrument. In this type of calorimeter, a constant environment is maintained inside an insulated reaction vessel. Typically, a silvered dewar is used, and the interacting components (e.g., solvent and solute) are held in

subcontainers. Liquid phases are usually stirred and the temperature is accurately recorded using a thermometer or thermocouple.

At the start of the experiment, the reaction vessel is allowed to reach a steady state, and a baseline temperature or temperature drift is recorded. The interaction of interest is then initiated by permitting the two components to mix, and the resulting temperature increase from baseline is recorded.

The system is calibrated by monitoring a standard reaction (e.g., neutralization of hydrochloric acid) or by applying a controlled amount of electrical energy via a heating coil. From this, the heat capacity of the system is determined, and the enthalpy change for any monitored process can be calculated from the observed temperature change. Isoperibol calorimeters can be easily constructed from their individual components, and several different instruments are commercially available at a modest cost. Owing to the high degree of accuracy and sensitivity that is required for pharmaceutical analysis, more sophisticated microcalorimeters are frequently used for studying pharmaceutical systems. An example of such an instrument is the TAMmanufactured by Thermometrics (Sweden).

In this instrument, twin sample cells are used (one for the sample and one for a reference) to achieve greater signal stability and to minimize the effects of spurious thermal fluctuations. Comparably, a similar instrument is available from Setaram known as a microDSC (France).

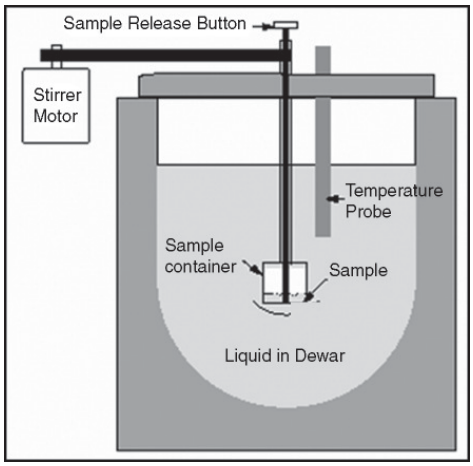
The cells are housed in a constant temperature environment maintained via a sophisticated heater and water jacket system. Minor changes in sample heat flow are detected with relative ease using this arrangement. The TAM and microDSC are

calibrated electrically, and the commercially available sample configurations allow the mixing of solids, liquids, and gases in various proportions.

Controlled gas and liquid flow rates, and changing sample environments (e.g., relative humidities) can also be achieved with appropriate accessories.

A major practical advantage of this type of calorimeter when compared with less sophisticated instruments is the small sample size requirement of only a few tens or hundreds of milligrams per determination.

This type of microcalorimeter has been used for detection and monitoring of crystallization events, sorption and desorption of organic and inorganic vapors, chemical reactions (drug degradation and drug interactions with excipients), molecular motions in amorphous pharmaceutical materials, ligand binding phenomena, microbiological growth, and the determination of solid heat capacities.



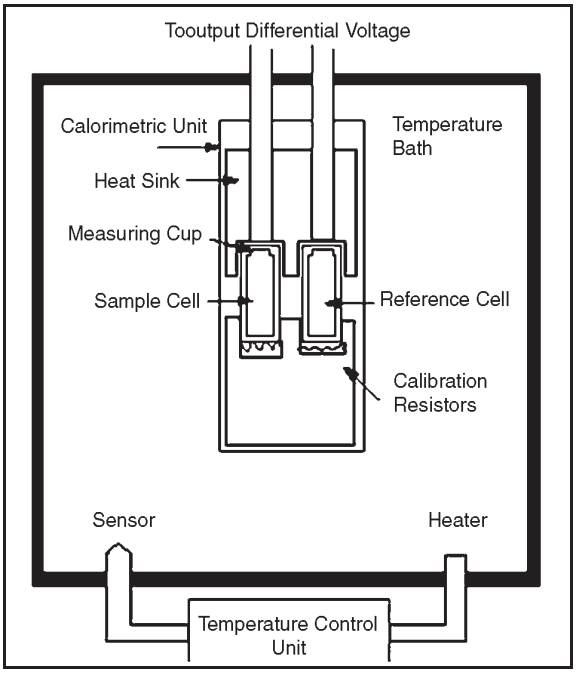


Fig. Schematic Diagram of an Isoperibol Calorimeter (top) and a Thermal Activity Monitor (TAM) (bottom).

Sample Preparation and Calibration

Microcalorimeters are usually operated in a similar way irrespective of the source of the energy change that is being monitored. Specimens are preequilibrated at the desired measuring temperature for several hours and then introduced into the calorimeter chamber.

After a short delay, the external stimulus is applied to trigger the event of interest, and then the energy that is liberated or consumed is measured.

The reaction is isolated from the environment by a jacket, which serves as a thermal shield to minimize the absorption and emission of radiant heat. Calibration of microcalorimeters is usually achieved by direct heating using an electric heating element (as mentioned previously).

Extreme care is required to achieve consistent sample preparation and to maintain constant experimental procedures because the interpretation of results can be confused easily by experimental artifacts.

Simultaneous thermal events of opposite sign (exothermic and endothermic) are quite common and may often confound the interpretation of data. In all experiments, an appropriate thermal reference is required as the energy changes that are measured are simply energy changes relative to the reference specimen.

Common references include an empty sample container, or a sample container filled with an inert material, which has a similar heat capacity and mass to the sample.

Definitions and Applications of Microcalorimetry

Microcalorimeters have found widespread use in the pharmaceutical sciences in recent years for applications as diverse as determining the degradation rate of drugs, estimating the strength of binding between proteins and receptor sites, and monitoring metabolic processes in microorganisms.

Interactions between water vapour and amorphous pharmaceutical solids were evaluated using isothermal microcalorimetry.

The desorption of water from theophylline monohydrate has been investigated using microcalorimetric approaches. The properties of surfactants and surface-active drugs in solution were studied by Attwood et al. using calorimetry, while titration microcalorimetry has been utilized to elucidate the nature of specific interactions in several pharmaceutical polymer-surfactants systems.

Drug decomposition was evaluated as a function of different excipient blends in compressed tablets using isothermal heat conduction microcalorimetry.

A more unusual pharmaceutical use of microcalorimetry is to study energetic changes that occur during tablet compaction. The compression calorimeter used is a custom-made research instrument that appears to have many potential applications for the pharmaceutical scientist.

Microcalorimetry has proven to be a particularly useful tool to detect different levels of disorder in pharmaceutical materials. Gao and Rytting demonstrated the validity of solution microcalorimetry to measure changes in the crystallinity during processing of both the drug compound and the excipients.

Other workers have used elevated vapour pressures to trigger crystallization of disordered materials in the calorimeter and have been able to use the measured energy output to directly quantify the levels of disorder crystallinity in their samples.

Stability studies using microcalorimetry are widely reported in the literature. Some authors have monitored exothermic degradation reactions over several days or weeks and have projected the degradation extent and rate over the shelf life of the drug or drug product.

The sensitivity of modern instruments is capable of measuring reaction rates of 1 _ 10_11 sec_1 directly at 25°C, corresponding to 0.03% degradation per year. Some researchers have also described qualitative screens for drug-excipient compatibility studies. Other authors have used microcalorimetry to monitor relaxation of amorphous

pharmaceutical materials and have then calculated relaxation time constants from these data for use in shelf life predictions. The use of microcalorimetry for preformulation stability screening of a drug with potentially reactive excipients has also been described.

REGULATORY CONSIDERATIONS

Calorimetric methods are infrequently used for routine quality control purposes because of their non-specific nature and relatively slow speed. However, data from calorimetry experiments are commonly presented in applications for new product licenses and in support of patent applications.

To ensure the integrity of all calorimetry data, normal procedures for good laboratory practices, standard operating procedures, appropriate calibration methods, and regular instrument servicing are necessary.

The use of DSC for the measurement of transition temperatures and sample purity is described in the United States Pharmacopoeia, and standard procedures for DSC analyses are also suggested by the ASTM.

6

Medical Imaging

Medical imaging refers to the techniques and processes used to create images of the human body (or parts thereof) for clinical purposes (medical procedures seeking to reveal, diagnose or examine disease) or medical science (including the study of normal anatomy and function).

As a discipline and in its widest sense, it is part of biological imaging and incorporates radiology (in the wider sense), radiological sciences, endoscopy, (medical) thermography, medical photography and microscopy (e.g. for human pathological investigations).

Measurement and recording techniques which are not primarily designed to produce images, such as electroencephalography (EEG) and magnetoencephalography (MEG) and others, but which produce data susceptible to be represented as maps (i.e. containing positional information), can be seen as forms of medical imaging.

In the clinical context, medical imaging is generally equated to Radiology or "clinical imaging" and the medical practitioner responsible for interpreting (and sometimes acquiring) the images is a radiologist. Diagnostic radiography designates the technical aspects of medical imaging and in particular the acquisition of medical images. The radiographer or radiologic technologist is usually responsible for acquiring medical images of diagnostic quality, although some radiological interventions are performed by radiologists. As a field of scientific investigation, medical imaging constitutes a sub-discipline of biomedical engineering, medical physics or medicine depending on the context: Research and development in the area of instrumentation, image acquisition (e.g. radiography), modelling and quantification are usually the preserve of biomedical engineering, medical physics and computer science; Research into the application and interpretation of medical images is usually the preserve of radiology and the medical sub-discipline relevant to medical condition or area of medical science (neuroscience, cardiology, psychiatry, psychology, etc) under investigation.

Many of the techniques developed for medical imaging also have scientific and industrial applications.Medical imaging is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. In this restricted sense, medical imaging can be seen as the solution of mathematical inverse problems. This means that cause (the properties of living tissue) is inferred from effect (the observed signal).

In the case of ultrasonography the probe consists of ultrasonic pressure waves and echoes inside the tissue show

the internal structure. In the case of projection radiography, the probe is X-ray radiation which is absorbed at different rates in different tissue types such as bone, muscle and fat.

Radio frequency system

The radio frequency (RF) transmission system consists of a RF synthesizer, power amplifier and transmitting coil. This is usually built into the body of the scanner. The power of the transmitter is variable, but high-end scanners may have a peak output power of up to 35 kW, and be capable of sustaining average power of 1 kW.

The receiver consists of the coil, pre-amplifier and signal processing system. While it is possible to scan using the integrated coil for transmitting and receiving, if a small region is being imaged then better image quality is obtained by using a close-fitting smaller coil. A variety of coils are available which fit around parts of the body, e.g., the head, knee, wrist, or internally, e.g., the rectum.

A recent development in MRI technology has been the development of sophisticated multi-element phased array coils which are capable of acquiring multiple channels of data in parallel. This 'parallel imaging' technique uses unique acquisition schemes that allow for accelerated imaging, by replacing some of the spatial coding originating from the magnetic gradients with the spatial sensitivity of the different coil elements.

However the increased acceleration also reduces the signal-to-noise ratio and can create residual artifacts in the image reconstruction. Two frequently used parallel acquisition and reconstruction schemes are SENSE and GRAPPA.

Gradients

Magnetic gradients are generated by three orthogonal coils, oriented in the *x*, *y* and *z* directions of the scanner. These are usually resistive electromagnets powered by sophisticated amplifiers which permit rapid and precise adjustments to their field strength and direction.

Typical gradient systems are capable of producing gradients from 20 mT/m to 100 mT/m (i.e. in a 1.5 T magnet, when a maximal *z*-axis gradient is applied the field strength may be 1.45 T at one end of a 1 m long bore, and 1.55 T at the other). It is the magnetic gradients that determine the plane of imaging - because the orthogonal gradients can be combined freely, any plane can be selected for imaging.Scan speed is dependent on performance of the gradient system.

Stronger gradients allow for faster imaging, or for higher resolution; similarly, gradients systems capable of faster switching can also permit faster scanning. However, gradient performance is limited by safety concerns over nerve stimulation.

In order to understand MRI contrast, it is important to have some understanding of the time constants involved in relaxation processes that establish equilibrium following RF excitation. As the high-energy nuclei relax and realign they emit energy at rates which are recorded to provide information about the material they are in.

The realignment of nuclear spins with the magnetic field is termed *longitudinal relaxation* and the time required for a certain percentage of the tissue's nuclei to realign is termed "Time 1" or T1, which is typically about 1 second at 1.5 Tesla main field strength. T2-weighted imaging relies upon local

dephasing of spins following the application of the transverse energy pulse; the *transverse* relaxation time is termed "Time 2" or T2, typically < 100 ms for tissue at 1.5 Tesla main field strength. A subtle but important variant of the T2 technique is called T2* imaging. T2 imaging employs a spin echo technique, in which spins are refocused to compensate for local magnetic field inhomogeneities.

T2* imaging is performed without refocusing. This sacrifices some image integrity (resolution) but provides additional sensitivity to relaxation processes that cause incoherence of transverse magnetization.

Applications of T2* imaging include functional MRI (fMRI) or evaluation of baseline vascular perfusion (e.g. cerebral blood flow (CBF)) and cerebral blood volume (CBV) using injected agents; in these cases, there is an inherent trade-off between image quality and detection sensitivity.

Because T2*-weighted sequences are sensitive to magnetic inhomogeneity (as can be caused by deposition of iron-containing blood-degradation products), such sequences are utilized to detect subtle areas of recent or chronic intra cranial hemorrhage ("Heme sequence").

Image contrast is created by using a selection of image acquisition parameters that weights signal by T1, T2 or T2, or no relaxation time ("proton-density images").

In the brain, T1-weighting causes the nerve connections of white matter to appear white, and the congregations of neurons of gray matter to appear gray, while cerebrospinal fluid appears dark.

The contrast of "white matter," "gray matter'" and "cerebrospinal fluid" is reversed using T2 or T2* imaging,

whereas proton-weighted imaging provides little contrast in normal subjects. Additionally, functional information (CBF, CBV, blood oxygenation) can be encoded within T1, T2, or T2.Diffusion weighted imaging (DWI) uses very fast scans with an additional series of gradients (diffusion gradients) rapidly turned on and off.

Protons from water diffusing randomly within the brain, via Brownian motion, lose phase coherence and, thus signal during application of diffusion gradients. In a brain with an acute infarction water diffusion is impaired, and signal loss on DWI sequences is less than in normal brain. DWI is the most sensitive method of detecting cerebral infarction (stroke) and works within 30 minutes of the ictus.

Contrast Enhancement

Both T1-weighted and T2-weighted images are acquired for most medical examinations; However they do not always adequately show the anatomy or pathology. The first option is to use a more sophisticated image acquisition technique such as fat suppression or chemical-shift imaging.

The other is to administer a contrast agent to delineate areas of interest. A contrast agent may be as simple as water, taken orally, for imaging the stomach and small bowel although substances with specific magnetic properties may be used. Most commonly, a paramagnetic contrast agent (usually a gadolinium compound) is given.

Gadolinium-enhanced tissues and fluids appear extremely bright on T1-weighted images. This provides high sensitivity for detection of vascular tissues (e.g. tumors) and permits assessment of brain perfusion (e.g. in stroke).

There have been concerns raised recently regarding the toxicity of gadolinium-based contrast agents and their impact on persons with impaired kidney function. Special actions may be taken, such as hemodialysis following a contrast MRI scan for renally-impaired patients.

More recently, superparamagnetic contrast agents (e.g. iron oxide nanoparticles) have become available. These agents appear very dark on T2-weighted images and may be used for liver imaging - normal liver tissue retains the agent, but abnormal areas (e.g. scars, tumors) do not.

They can also be taken orally, to improve visualization of the gastrointestinal tract, and to prevent water in the gastrointestinal tract from obscuring other organs (e.g. pancreas).

Diamagnetic agents such as barium sulfate have been studied for potential use in the gastrointestinal tract, but are less frequently used.

MRI vs CT

A computed tomography (CT) scanner uses X-rays, a type of ionizing radiation, to acquire its images, making it a good tool for examining tissue composed of elements of a relatively higher atomic number than the tissue surrounding them, such as bone and calcifications (calcium based) within the body (carbon based flesh), or of structures (vessels, bowel).

MRI, on the other hand, uses non-ionizing radio frequency (RF) signals to acquire its images and is best suited for non-calcified tissue.

CT may be enhanced by use of contrast agents containing elements of a higher atomic number than the surrounding flesh such as iodine or barium. Contrast agents for MRI are

those which have paramagnetic properties. One example is gadolinium. Both CT and MRI scanners can generate multiple two-dimensional cross-sections (slices) of tissue and three-dimensional reconstructions. Unlike CT, which uses only X-ray attenuation to generate image contrast, MRI has a long list of properties that may be used to generate image contrast. By variation of scanning parameters, tissue contrast can be altered and enhanced in various ways to detect different features.

MRI can generate cross-sectional images in any plane (including oblique planes). CT was limited to acquiring images in the axial (or near axial) plane in the past. The scans used to be called Computed *Axial* Tomography scans (CAT scans). However, the development of multi-detector CT scanners with near-isotropic resolution, allows the CT scanner to produce data that can be retrospectively reconstructed in any plane with minimal loss of image quality.

For purposes of tumor detection and identification, MRI is generally superior. However, CT usually is more widely available, faster, much less expensive, and may be less likely to require the person to be sedated or anesthetized.

MRI is also best suited for cases when a patient is to undergo the exam several times successively in the short term, because, unlike CT, it does not expose the patient to the hazards of ionizing radiation.

Economics of MRI

MRI equipment is expensive. New 1.5 Tesla scanners often cost between $1,000,000 USD and $1,500,000 USD. New 3.0 Tesla scanners often cost between $2,000,000 and

$2,300,000 USD. Construction of MRI suites can cost $500,000 USD. For over a dozen years, MRI scanners have been significant sources of revenue for healthcare providers in the US. This is because of favorable reimbursement rates from insurers, both private and federal government programs. Insurance reimbursement has historically been provided in two components, technical for the actual performance of the MRI scan and professional for the radiologist's review of the images and/or data.

In the US, the 2007 Deficit Reduction Act (DRA) significantly reduced reimbursement rates paid by federal insurance programs for the technical component of many scans, shifting the economic landscape. Many private insurers have followed suit.

Currently, in the US, there is increasing interest in reducing the costs associated with MRI services and simultaneously improving the ability to effectively and efficiently provide MRI examination services to larger numbers of patients with the same equipment.

The k-space Formalism

In 1983 Ljunggren and Tweig independently introduced the k-space formalism, a technique that proved invaluable in unifying different MR imaging techniques.

They showed that the demodulated MR signal *S*(*t*) generated by freely precessing nuclear spins in the presence of a linear magnetic field gradient *G* equals the Fourier transform of the effective spin density i.e.

$$S(t) = Peffective\left(\vec{k}(t)\right) \equiv \int d^3x\, p\left(\vec{x}\right) . e^{2\pi 2\vec{k}(t).\vec{x}}$$

where: $\vec{k}(t) \equiv \int_0^t \vec{G}(t')dt'$

In other words, as time progresses the signal traces out a trajectory in k-space with the velocity vector of the trajectory proportional to the vector of the applied magnetic field gradient. By the term *effective spin density* we mean the true spin density $P(\vec{x})$ corrected for the effects of T_1 preparation, T_2 decay, dephasing due to field inhomogeneity, flow, diffusion, etc. and any other phenomena that affect that amount of transverse magnetization available to induce signal in the RF probe.

From the basic k-space formula, it follows immediately that we reconstruct an image $I(\vec{x})$ simply by taking the inverse Fourier transform of the sampled data viz.

$$I(\vec{x}) = \int d^3k S(\vec{k}(t)).e^{-2m\vec{k}(t).\vec{x}}$$

Using the k-space formalism, a number of seemingly complex ideas become simple. For example, it becomes very easy to understand the role of phase encoding (the so-called spin-warp method).

In a standard spin echo or gradient echo scan, where the readout (or view) gradient is constant (e.g. G_x), a single line of k-space is scanned per RF excitation. When the phase encoding gradient is zero, the line scanned is the k_x axis. When a non-zero phase-encoding pulse is added in between the RF excitation and the commencement of the readout gradient, this line moves up or down in k-space i.e. we scan the line k_y=constant.

The k-space formalism also makes it very easy to compare different scanning techniques. In single-shot EPI, all of k-space is scanned in a single shot, following either a sinusoidal or zig-zag trajectory. Since alternating lines of k-space are scanned in opposite directions, this must be taken into

account in the reconstruction. Multi-shot EPI and fast spin echo techniques acquire only part of k-space per excitation. In each shot, a different interleaved segment is acquired, and the shots are repeated until k-space is sufficiently well-covered. Since the data at the centre of k-space represent lower spatial frequencies than the data at the edges of k-space, the T_E value for the centre of k-space determines the image's T_2 contrast.

The importance of the centre of k-space in determining image contrast can be exploited in more advanced imaging techniques. One such technique is spiral acquisition - a rotating magnetic field gradient is applied, causing the trajectory in k-space to trace out spiral out from the centre to the edge.

Due to T_2 and T_2 * decay the signal is greatest at the start of the acquisition, hence acquiring the centre of k-space first improves contrast to noise ratio (CNR) when compared to conventional zig-zag acquisitions, especially in the presence of rapid movement.

Since $\vec{x}$ and $\vec{k}$ are conjugate variables (with respect to the Fourier transform) we can use the Nyquist theorem to show that the step in k-space determines the field of view of the image (maximum frequency that is correctly sampled) and the maximum value of k sampled determines the resolution i.e.

$$FOV \propto \frac{1}{\Delta k} \text{ Resolution} \propto |k_{\max}|$$

(these relationships apply to each axis [X, Y, and Z] independently).

Application

In clinical practice, MRI is used to distinguish pathologic tissue (such as a brain tumor) from normal tissue. One

advantage of an MRI scan is that it is thought to be harmless to the patient. It uses strong magnetic fields and non-ionizing radiation in the radio frequency range. Compare this to CT scans and traditional X-rays which involve doses of ionizing radiation and may increase the risk of malignancy, especially in a fetus.

While CT provides good spatial resolution (the ability to distinguish two structures an arbitrarily small distance from each other as separate), MRI provides comparable resolution with far better contrast resolution (the ability to distinguish the differences between two arbitrarily similar but not identical tissues). The basis of this ability is the complex library of *pulse sequences* that the modern medical MRI scanner includes, each of which is optimized to provide *image contrast* based on the chemical sensitivity of MRI.

For example, with particular values of the *echo time* (TE) and the *repetition time* (TR), which are basic parameters of image acquisition, a sequence will take on the property of T2-weighting.

On a T2-weighted scan, fat-, water- and fluid-containing tissues are bright (most modern T2 sequences are actually *fast T2* sequences).

Damaged tissue tends to develop edema, which makes a T2-weighted sequence sensitive for pathology, and generally able to distinguish pathologic tissue from normal tissue.

With the addition of an additional radio frequency pulse and additional manipulation of the magnetic gradients, a T2-weighted sequence can be converted to a FLAIR sequence, in which free water is now dark, but edematous tissues remain bright.

This sequence in particular is currently the most sensitive way to evaluate the brain for demyelinating diseases, such as multiple sclerosis.

The typical MRI examination consists of 5-20 sequences, each of which are chosen to provide a particular type of information about the subject tissues. This information is then synthesized by the interpreting physician.

Specialized MRI Scans

Diffusion MRI

Diffusion MRI measures the diffusion of water molecules in biological tissues. In an isotropic medium (inside a glass of water for example) water molecules naturally move randomly according to Brownian motion.

In biological tissues however, the diffusion may be anisotropic. For example a molecule inside the axon of a neuron has a low probability of crossing the myelin membrane.

Therefore the molecule will move principally along the axis of the neural Fibre. If we know that molecules in a particular voxel diffuse principally in one direction we can make the assumption that the majority of the fibers in this area are going parallel to that direction.

The recent development of diffusion tensor imaging (DTI) enables diffusion to be measured in multiple directions and the fractional anisotropy in each direction to be calculated for each voxel. This enables researchers to make brain maps of Fibre directions to examine the connectivity of different regions in the brain (using tractography) or to examine areas of neural degeneration and demyelination in diseases like Multiple

Sclerosis. Another application of diffusion MRI is diffusion-weighted imaging (DWI). Following an ischemic stroke, DWI is highly sensitive to the changes occurring in the lesion.

It is speculated that increases in restriction (barriers) to water diffusion, as a result of cytotoxic edema (cellular swelling), is responsible for the increase in signal on a DWI scan. Other theories, including acute changes in cellular permeability and loss of energy-dependent (ATP) cytoplasmic streaming, have been proposed to explain the phenomena.

The DWI enhancement appears within 5-10 minutes of the onset of stroke symptoms (as compared with computed tomography, which often does not detect changes of acute infarct for up to 4-6 hours) and remains for up to two weeks.

CT, due to its insensitivity to acute ischemia, is typically employed to rule out hemorrhagic stroke, which would entirely prevent the use of tissue plasminogen activator (tPA). Further, coupled with scans sensitized to cerebral perfusion, researchers can highlight regions of "perfusion/diffusion mismatch" that may indicate regions capable of salvage by reperfusion therapy.

Finally, it has been proposed that diffusion MRI may be able to detect minute changes in extracellular water diffusion and therefore could be used as a tool for fMRI. The nerve cell body enlarges when it conducts an action potential, hence restricting extracellular water molecules from diffusing naturally. Although this process works in theory, evidence is only moderately convincing.

Like many other specialized applications, this technique is usually coupled with a fast image acquisition sequence, such as echo planar imaging sequence.

Magnetic Resonance Angiography

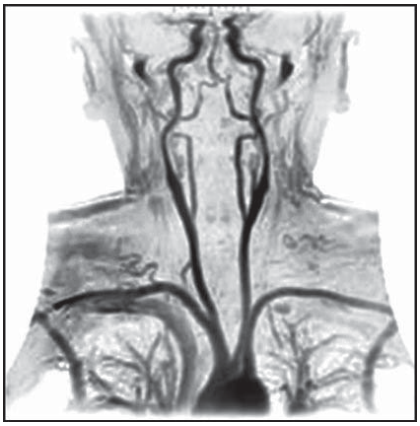


Fig. Magnetic Resonance Angiography.

Magnetic resonance angiography (MRA) is used to generate pictures of the arteries in order to evaluate them for stenosis (abnormal narrowing) or aneurysms (vessel wall dilatations, at risk of rupture).

MRA is often used to evaluate the arteries of the neck and brain, the thoracic and abdominal aorta, the renal arteries, and the legs (called a "run-off").

A variety of techniques can be used to generate the pictures, such as administration of a paramagnetic contrast agent (gadolinium) or using a technique known as "flow-related enhancement" (e.g. 2D and 3D time-of-flight sequences), where most of the signal on an image is due to blood which has recently moved into that plane.

Magnetic resonance venography (MRV) is a similar procedure that is used to image veins.

In this method the tissue is now excited inferiorly while signal is gathered in the plane immediately superior to the excitation plane, and thus imaging the venous blood which has recently moved from the excited plane.

Magnetic Resonance Spectroscopy

In vivo ('in the living organism') magnetic resonance spectroscopy (MRS), also known as MRSI (MRS imaging) and volume selective NMR spectroscopy, is a technique which combines the spatially-addressable nature of MRI with the spectroscopically-rich information obtainable from NMR.

That is to say, MRI allows one to study a particular region within an organism or sample, but gives relatively little information about the chemical or physical nature of that region (its chief value is in being able to distinguish the properties of that region, how much fat or water is present, relative to those of surrounding regions).

MR spectroscopy, however, provides a wealth of information about other biological chemicals ('metabolites') within that region, as would an NMR spectrum of that region.

Functional MRI

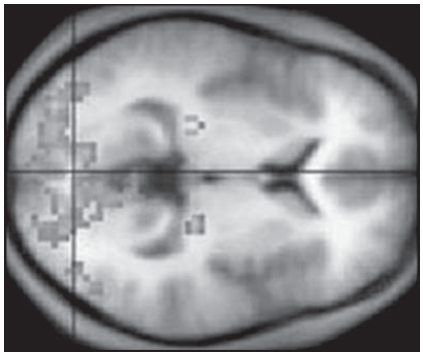


Fig. A fMRI Scan Showing Regions of Activation, Including the Primary Visual Cortex.

Functional MRI (fMRI) measures signal changes in the brain that are due to changing neural activity. The brain is scanned at low resolution but at a rapid rate (typically once

every 2-3 seconds). Increases in neural activity cause changes in the MR signal via T2* changes; this mechanism is referred to as the BOLD (blood-oxygen-level dependent) effect.

Increased neural activity causes an increased demand for oxygen, and the vascular system actually overcompensates for this, increasing the amount of oxygenated hemoglobin relative to deoxygenated hemoglobin.

Because deoxygenated hemoglobin attenuates the MR signal, the vascular response leads to a signal increase that is related to the neural activity. The precise nature of the relationship between neural activity and the BOLD signal is a subject of current research.

The BOLD effect also allows for the generation of high resolution 3D maps of the venous vasculature within neural tissue.While BOLD signal is the most common method employed for neuroscience studies in human subjects, the flexible nature of MR imaging provides means to sensitize the signal to other aspects of the blood supply.

Alternative techniques employ arterial spin labeling (ASL) or weight the MRI signal by cerebral blood flow (CBF) and cerebral blood volume (CBV). The CBV method requires injection of a class of MRI contrast agents that are now in human clinical trials.

Because this method has been shown to be far more sensitive than the BOLD technique in preclinical studies, it may potentially expand the role of fMRI in clinical applications. The CBF method provides more quantitative information than the BOLD signal, albeit at a significant loss of detection sensitivity.

Interventional MRI

The lack of harmful effects on the patient and the operator make MRI well-suited for "interventional radiology", where the images produced by a MRI scanner are used to guide minimally-invasive procedures.

Radiation Therapy Simulation

Because of MRI's superior imaging of soft tissues, it is now being utilized to specifically locate tumors within the body in preparation for radiation therapy treatments. For therapy simulation, a patient is placed in specific, reproducible, body position and scanned.

The MRI system then computes the precise location, shape and orientation of the tumor mass, correcting for any spatial distortion inherent in the system. The patient is then marked or tattooed with points which, when combined with the specific body position, will permit precise triangulation for radiation therapy.

Current Density Imaging

Current density imaging (CDI) endeavors to use the phase information from images to reconstruct current densities within a subject. Current density imaging works because electrical currents generate magnetic fields, which in turn affect the phase of the magnetic dipoles during an imaging sequence.

To date no successful CDI has been performed using biological currents, but several studies have been published which involve applied currents through a pair of electrodes.

Magnetic Resonance Guided Focused Ultrasound

In MRgFUS therapy, ultrasound beams are focused on a tissue - guided and controlled using MR thermal imaging - and due to the significant energy deposition at the focus, temperature within the tissue rises to more than 65°C, completely destroying it. This technology can achieve precise "ablation" of diseased tissue.

MR imaging provides a three-dimensional view of the target tissue, allowing for precise focusing of ultrasound energy. The MR imaging provides quantitative, real-time, thermal images of the treated area.

This allows the physician to ensure that the temperature generated during each cycle of ultrasound energy is sufficient to cause thermal ablation within the desired tissue and if not, to adapt the parameters to ensure effective treatment.

Multinuclear Imaging

Hydrogen is the most frequently imaged nucleus in MRI because it is present in biological tissues in great abundance. However, any nucleus which has a net nuclear spin could potentially be imaged with MRI. Such nuclei include helium-3, carbon-13, fluorine-19, oxygen-17, sodium-23, phosphorus-31 and xenon-129. ^{23}Na and ^{31}P are naturally abundant in the body, so can be imaged directly.

Gaseous isotopes such as ^{3}He or ^{129}Xe must be hyperpolarized and then inhaled as their nuclear density is too low to yield a useful signal under normal conditions. ^{17}O, ^{13}C and ^{19}F can be administered in sufficient quantities in liquid form (e.g. ^{17}O-water, ^{13}C-glucose solutions or perfluorocarbons) that hyperpolarization is not a necessity.

Multinuclear imaging is primarily a research technique at present. However, potential applications include functional imaging and imaging of organs poorly seen on ^{1}H MRI (e.g. lungs and bones) or as alternative contrast agents. Inhaled hyperpolarized ^{3}He can be used to image the distribution of air spaces within the lungs.

Injectable solutions containing ^{13}C or stabilized bubbles of hyperpolarized ^{129}Xe have been studied as contrast agents for angiography and perfusion imaging. ^{31}P can potentially provide information on bone density and structure, as well as functional imaging of the brain.

Experimental MRI Techniques

Currently there is active research in several new MRI technologies like magnetization transfer MRI (MT-MRI), diffusion tensor MRI (DT-MRI), and proton MR spectroscopy, plus recent research in to Dendrimer-enhanced MRI as a diagnostic and prognostic biomarker of sepsis-induced acute renal failure.

Safety

Implants and foreign bodies: Pacemakers are generally considered an absolute contraindication towards MRI scanning, though highly specialized protocols have been developed to permit scanning of select pacing devices. Several cases of arrhythmia or death have been reported in patients with pacemakers who have undergone MRI scanning without appropriate precautions.

Other electronic implants have varying contraindications, depending upon scanner technology, implant properties, scanning protocols and anatomy being imaged. Though

pacemakers receive significant attention, it should also be noted that many other forms of medical or biostimulation implants may be contraindicated for MRI scans.

These may include Vagus nerve stimulators, implantable cardioverter-defibrillators (ICD), loop recorders, insulin pumps, cochlear implants, deep brain stimulators and many others. Medical device patients should always present complete information (manufacturer, model, serial number and date of implantation) about all implants to both the referring physician and to the radiologist or technologist before entering the room for the MRI scan.

While these implants pose a current problem, scientist and manufacturers are working on improved designs which will further minimize the risks that MRI scans pose to medical device operations. One such development in the works is a nano coating for implants intended to screen them from the radio frequency waves, helping to make MRI exams available to patients currently prohibited from receiving them. The current article for this is from New Scientist.

Ferromagnetic foreign bodies (e.g. shell fragments), or metallic implants (e.g. surgical prostheses, aneurysm clips) are also potential risks, and safety aspects need to be considered on an individual basis. Interaction of the magnetic and radio frequency fields with such objects can lead to: trauma due to movement of the object in the magnetic field, thermal injury from radio-frequency induction heating of the object, or failure of an implanted device.

These issues are especially problematic when dealing with the eye. Most MRI centers require an orbital x-ray be performed on anyone who suspects they may have small

metal fragments in their eyes, perhaps from a previous accident, something not uncommon in metalworking.

Because of its non-ferromagnetic nature and poor electrical conductivity, titanium and its alloys are useful for long term implants and surgical instruments intended for use in image-guided surgery.

In particular, not only is titanium safe from movement from the magnetic field, but artifacts around the implant are less frequent and less severe than with more ferromagnetic materials e.g. stainless steel.

Artifacts from metal frequently appear as regions of empty space around the implant - frequently called 'black-hole artifact' e.g. a 3mm titanium alloy coronary stent may appear as a 5mm diameter region of empty space on MRI, whereas around a stainless steel stent, the artifact may extend for 10-20 mm or more.

In 2006, a new classification system for implants and ancillary clinical devices has been developed by ASTM International and is now the standard supported by the US Food and Drug Administration:

Fig. MR Safe Sign.

MR-Safe: The device or implant is completely non-magnetic, non-electrically conductive, and non-RF reactive, eliminating all of the primary potential threats during an MRI procedure.

Fig. MR Conditional Sign.

MR-Conditional: A device or implant that may contain magnetic, electrically conductive or RF-reactive components that is safe for operations in proximity to the MRI, provided the conditions for safe operation are defined and observed (such as 'tested safe to 1.5 teslas' or 'safe in magnetic fields below 500 gauss in strength').

Fig. MR Unsafe Sign.

MR-Unsafe: Nearly self-explanatory, this category is reserved for objects that are significantly ferromagnetic and pose a clear and direct threat to persons and equipment within the magnet room. In the case of pacemakers, the risk is thought to be primarily RF induction in the pacing electrodes/wires causing inappropriate pacing of the heart, rather than the magnetic field affecting the pacemaker itself. Much research and development is being undertaken, and many tools are being developed in order to predict the effects of the RF fields inside the body.

Projectile or Missile Effect

As a result of the very high strength of the magnetic field needed to produce scans (frequently up to 60,000 times the earth's own magnetic field effects), there are several incidental safety issues addressed in MRI facilities.

Missile-effect accidents, where ferromagnetic objects are attracted to the centre of the magnet, have resulted in injury and death. A video simulation of a fatal projectile effect accident illustrates the extreme power that contemporary MRI equipment can exert on ferromagnetic objects.

In order to help reduce the risks of projectile accidents, ferrous objects and devices are typically prohibited in proximity to the MRI scanner, with non ferro-magnetic versions of many tools and devices typically retained by the scanning facility.

Patients undergoing MRI examinations are required to remove all metallic objects, often by changing into a gown or 'scrubs'. The magnetic field and the associated risk of missile-effect accidents remains a permanent hazard — as superconductive MRI magnets retain their magnetic field, even in the event of a power outage.

Radio Frequency Energy

A powerful radio transmitter is needed for excitation of proton spins.

This can heat the body significantly, with the risk of hyperthermia in patients, particularly the obese or patients with thermoregulation disorders. Several countries have issued restrictions on the maximum specific absorption rate that a scanner may produce.

Peripheral Nerve Stimulation

The rapid switching (on and off) of the magnetic field gradients needed for imaging is capable of causing nerve stimulation. Volunteers report a twitching sensation when exposed to rapidly switched fields, particularly in their extremities.

The reason the peripheral nerves are stimulated is that the changing field increases with distance from the centre of the gradient coils (which more or less coincides with the centre of the magnet). Note however that when imaging the head, the heart is far off-centre and induction of even a tiny current into the heart must be avoided at all costs.

Although PNR was not a problem for the slow, weak gradients used in the early days of MRI, the strong, rapidly-switched gradients used in techniques such as EPI, fMRI, diffusion MRI, etc. are indeed capable of inducing PNR.

American and European regulatory agencies insist that manufacturers stay below specified dB/dt limits (dB/dt is the change in field per unit time) or else prove (via clinical studies) that no PNR is induced for any imaging sequence. As a result of dB/dt limitation software and/or hardware, commercial MRI systems cannot use the full rated power of their gradient amplifiers.

Acoustic Noise

Loud noises and vibrations are produced by forces resulting from rapidly switched magnetic gradients interacting with the main magnetic field, in turn causing minute expansions and contractions of the coil itself.

This is most marked with high-field machines and rapid-imaging techniques in which sound intensity can reach 130 dB (equivalent to a jet engine at take-off).Appropriate use of ear protection is essential for anyone inside the MRI scanner room during the examination.

Cryogens

As described above in 'Scanner Construction And Operation', many MRI scanners rely on cryogenic liquids to enable superconducting capabilities of the electromagnetic coils within. Though the cryogenic liquids most frequently used are non-toxic, their physical properties present specific hazards.

An emergency shut-down of a superconducting electromagnet, an operation known as "quenching", involves the rapid boiling of liquid helium from the device. If the rapidly expanding helium cannot be dissipated though an external vent, sometimes referred to as 'quench pipe', it may be released into the scanner room where it may cause displacement of the oxygen and present a risk of asphyxiation.

Liquid helium, the most commonly used cryogen in MRI, undergoes near explosive expansion as it changes from liquid to a gaseous state. Rooms built in support of superconducting MRI equipment should be equipped with pressure relief mechanisms and an exhaust fan, in addition to the required quench pipe. Since a quench results in rapid loss of all cryogens in the magnet, recommissioning the magnet is extremely expensive and time-consuming.

Spontaneous quenches are uncommon, but can occur at any time. Quenches may also be triggered by equipment malfunction, improper cryogen fill technique, contaminates inside the cryostat, or extreme magnetic or vibrational disturbances.

Contrast Agents

The most frequently used intravenous contrast agents are based on chelates of gadolinium. In general, these agents have proved safer than the iodinated contrast agents used in X-ray radiography or CT. Anaphylactoid reactions are rare occurring in approx 0.03-0.1 per cent.

Of particular interest is the lower incidence of nephrotoxicity, compared with iodinated agents, when given at usual doses—this has made contrast-enhanced MRI scanning an option for patients with renal impairment, who would otherwise not be able to undergo contrast-enhanced CT.

Although gadolinium agents have proved useful for patients with renal impairment, in patients with severe renal failure requiring dialysis there is a risk of a rare but serious illness, nephrogenic systemic fibrosis, that may be linked to the use of certain gadolinium-containing agents: the most frequently linked is gadodiamide, but other agents have been linked too. Although a causal link has not been definitively established, current guidelines in the United States are that dialysis patients should only receive gadolinium agents where essential, and that dialysis should be performed as soon as possible after the scan is complete, in order to remove the agent from the body promptly. In Europe where more gadolinium-containing agents are available, a classification of agents according to potential risks has been released.

Pregnancy

No harmful effects of MRI on the fetus have been demonstrated. In particular, MRI avoids the use of ionizing radiation, to which the fetus is particularly sensitive.

However, as a precaution, current guidelines recommend that pregnant women undergo MRI only when essential. This is particularly the case during the first trimester of pregnancy, as organogenesis takes place during this period.

The concerns in pregnancy are the same as for MRI in general, but the fetus may be more sensitive to the effects—particularly to heating and to noise. However, one additional concern is the use of contrast agents; gadolinium compounds are known to cross the placenta and enter the fetal bloodstream, and it is recommended that their use be avoided.

Despite these concerns, MRI is rapidly growing in importance as a way of diagnosing and monitoring congenital defects of the fetus because it can provide more diagnostic information than ultrasound and it lacks the ionizing radiation of CT. MRI without contrast is the imaging mode of choice for pre-surgical, in-utero diagnosis and evaluation of fetal tumors, primarily teratomas, facilitating open fetal surgery, other fetal interventions, and planning for procedures (such as the EXIT procedure) to safely deliver and treat babies whose defects would otherwise be fatal.

Claustrophobia and Discomfort

Due to the construction of MRI scanners, they are potentially unpleasant in which to lie. The part of the body being imaged needs to lie at the centre of the magnet (which is often a long, narrow tube). Because scan times may be long (perhaps one hour), people with even mild claustrophobia are often unable to tolerate an MRI scan without management.

Management may include:

- Advance preparation
  - Visiting the scanner to see the room and practice lying on the table
  - Visualization techniques
  - Chemical sedation
  - General anesthesia
- Coping while inside the scanner
  - Holding a "panic button"
  - Listening to music on headphones or watching a movie with a Head-mounted display while in the machine
- Modified scanner designs
  - Upright MRIs (made exclusively by FONAR)
  - Open-bore design scanners.

Though open MRIs have increased in popularity as of late, they produce inferior scan quality because they operate at lower magnetic fields than closed MRIs.

For babies and children, chemical sedation or general anesthesia are the norm. These MRI subjects are too young to be instructed to hold still during the scanning session. Obese patients and pregnant women may find the MRI machine to be a tight fit, and even some claustrophobics may find the experience intolerable without sedation. Pregnant women may also have difficulty lying on their backs for an hour or more without moving.

The noise associated with the operation of an MRI scanner (especially the audible noise associated with the gradient pulses applied to the subject) can also exacerbate the discomfort associated with the procedure.

The European Physical Agents Directive

The European Physical Agents (Electromagnetic Fields) Directive is European legislation that has been adopted in European legislature. By 2008 each individual state within the European Union must include this directive in its own law.

The directive applies to occupational exposure to electromagnetic fields (not medical exposure) and was intended to limit workers' acute exposure to strong electromagnetic fields, as may be found near electricity substations, radio or television transmitters or industrial equipment. However, the regulations impact significantly on MRI, with separate sections of the regulations limiting exposure to static magnetic fields, changing magnetic fields and radio frequency energy. Field strength limits are given which may not be exceeded for any period of time.

An employer may commit a criminal offense by allowing a worker to exceed an exposure limit if that is how the Directive is implemented in a particular Member State.

The Directive is based on the international consensus of established effects of exposure to electromagnetic fields, and in particular the advice of the European Commissions's advisor, the International Commission on Non-Ionizing Radiation Protection (ICNIRP). The aims of the Directive, and the ICNIRP guidelines upon which it is based, are to prevent exposure to potentially harmful fields. The actual limits in the Directive are very similar to the limits advised by the Institute of Electrical and Electronics Engineers, with the exception of the frequencies produced by the gradient coils, where the IEEE limits are significantly higher.

Many Member States of the EU already have either specific EMF regulations or (as in the UK) a general requirement under workplace health and safety legislation to protect workers against electromagnetic fields. In almost all cases the existing regulations are aligned with the ICNIRP limits so that the Directive should, in theory, have little impact on any employer already meeting their legal responsibilities.

The introduction of the Directive has brought to light an existing potential issue with occupational exposures to MRI fields. There are at present very few data on the number or types of MRI practice that might lead to exposures in excess of the levels of the Directive.

There is a justifiable concern amongst MRI practitioners that if the Directive were to be enforced more vigorously than existing legislation, the use of MRI might be restricted, or working practices of MRI personnel might have to change.

In the initial draft a limit of static field strength to 2 T was given. This has since been removed from the regulations, and whilst it is unlikely to be restored as it was without a strong justification, some restriction on static fields may be reintroduced after the matter has been considered more fully by ICNIRP. The effect of such a limit might be to restrict the installation, operation and maintenance of MRI scanners with magnets of 2 T and stronger. As the increase in field strength has been instrumental in developing higher resolution and higher performance scanners, this would be a significant step back. This is why it is unlikely to happen without strong justification. Individual government agencies and the

European Commission have now formed a working group to examine the implications on MRI and to try to address the issue of occupational exposures to electromagnetic fields from MRI.

Modern imaging technology

Fluoroscopy

Fluoroscopy produces real-time images of internal structures of the body in a similar fashion to radiography, but employs a constant input of x rays. Contrast media, such as barium, iodine, and air are used to visualize internal organs as they work. Fluoroscopy is also used in image-guided procedures when constant feedback during a procedure is required.

Magnetic Resonance Imaging (MRI)

A Magnetic Resonance Imaging instrument (MRI scanner) uses powerful magnets to polarise and excite hydrogen nuclei (single proton) in water molecules in human tissue, producing a detectable signal which is spatially encoded resulting in images of the body.

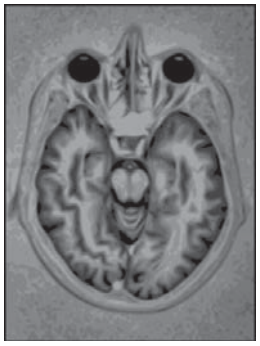


Fig. Human Brain MRI

In brief, MRI involves the use of three kinds of electromagnetic field: a very strong (of the order of units of teslas) static magnetic field to polarize the hydrogen nuclei, called the static field; a weaker time-varying (of the order of 1 kHz) for spatial encoding, called the gradient field(s); and a weak radio-frequency (RF) field for manipulation of the hydrogen nuclei to produce measurable signals, collected through an RF antenna.

Like CT, MRI traditionally creates a 2D image of a thin "slice" of the body and is therefore considered a tomographic imaging technique.

Modern MRI instruments are capable of producing images in the form of 3D blocks, which may be considered a generalisation of the single-slice, tomographic, concept. Unlike CT, MRI does not involve the use of ionizing radiation and is therefore not associated with the same health hazards; for example there are no known long term effects of exposure to strong static fields and therefore there is no limit on the number of scans to which an individual can be subjected, in contrast with X-ray and CT.

However, there are well identified health risks associated with tissue heating from exposure to the RF field and the presence of implanted devices in the body, such as pace makers. These risks are strictly controlled as part of the design of the instrument and the scanning protocols used.

CT and MRI being sensitive to different properties of the tissue, the appearance of the images obtained with the two techniques differ markedly. In CT, X-rays must be blocked by some form of dense tissue to create an image, therefore the image quality when looking at soft tissues will be poor.

While any nucleus with a net nuclear spin can be used, the proton of the hydrogen atom remains the most widely used, especially in the clinical setting, since it is so ubiquitous and returns much signal. This nucleus, present in water molecules, allows excellent soft-tissue contrast.MRI, or "NMR imaging" as it was originally known, has only been in use since the early 1980s. Effects from long term, or repeated exposure, to the intense static magnetic field are not known.

Nuclear Medicine

Images from gamma cameras are used in Nuclear Medicine to detect regions of biological activity that are often associated with diseases. A short lived isotope, such as ^{131}I is administered to the patient. These isotopes are more readily absorbed by biologically active regions of the body, such as tumors or fracture points in bones.

Positron Emission Tomography (PET)

Positron emission tomography is primarily used to detect diseases of the brain and heart. Similarly to nuclear medicine, a short-lived isotope, such as ^{18}F, is incorporated into a substance used by the body such as glucose which is absorbed by the tumor of interest. PET scans are often viewed along side computed tomography scans, which can be performed on the same equipment without moving the patient. This allows the tumors detected by the PET scan to be viewed next to the rest of the patient's anatomy detected by the CT scan.

Projection Radiography

Radiographs, more commonly known as x-rays, are often used to determine the type and extent of a fracture as well

as for detecting pathological changes in the lungs. With the use of radio-opaque contrast media, such as barium, they can also be used to visualize the structure of the stomach and intestines - this can help diagnose ulcers or certain types of colon cancer.

Photoacoustic Imaging

Photoacoustic imaging is a recently developed hybrid biomedical imaging modality based on the photoacoustic effect. It combines the advantages of optical absorption contrast with ultrasonic spatial resolution for deep imaging in (optical) diffusive or quasi-diffusive regime. Recent studies have shown that photoacoustic imaging can be used in vivo for tumor angiogenesis monitoring, blood oxygenation mapping, functional brain imaging, and skin melanoma detection etc.

Tomography

Tomography is the method of imaging a single plane, or slice, of an object resulting in a tomogram. There are several forms of tomography:

Linear Tomography

This is the most basic form of tomography. The X-ray tube moved from point "A" to point "B" above the patient, while the cassette holder (or "bucky") moves simultaneously under the patient from point "B" to point "A."

The fulcrum, or pivot point, is set to the area of interest. In this manner, the points above and below the focal plane are blurred out, just as the background is blurred when panning a camera during exposure. No longer carried out and replaced by computed tomography.

Poly Tomography

This was a complex form of tomography. With this technique, a number of geometrical movements were programmed, such as hypocycloidic, circular, figure, and elliptical. Philips Medical Systems produced one such device called the 'Polytome.' No longer carried out, replaced by computed tomography.

Zonography

This is a variant of linear tomography, where a limited arc of movement is used. It is still used in some centres for visualising the kidney during an intravenous urogram (IVU)

Orthopantomography

The only common tomographic examination in use. This makes use of a complex movement to allow the radiographic examination of the mandible, as if it were a flat bone. It is often referred to as a "Panaray", but this is incorrect, as it is a trademark of a specific company's equipment

Computed Tomography (CT)

A CT scan, also known as a CAT scan, is a helical tomography (latest generation), which traditionally produces a 2D image of the structures in a thin section of the body. It uses X-rays. It has a greater ionizing radiation dose burden than projection radiography; repeated scans should be limited.

Clinical imaging or biological imaging techniques

Electron Microscopy

The electron microscope is a microscope that can magnify very small details with high resolving power due to the use

of electrons as the source of illumination, magnifying at levels up to 2,000,000 times.Electron microscopy is employed in anatomic pathology to identify organelles within the cells. Its usefulness has been greatly reduced by immunhistochemistry but it is still irreplaceable for the diagnosis of kidney disease, identification of immotile cilia syndrome and many other tasks

Creation of Three-Dimensional Images

Recently, techniques have been developed to enable CT, MRI and ultrasound scanning software to produce 3D images for the physician. Traditionally CT and MRI scans produced 2D static output on film. To produce 3D images, many scans are made, then combined by computers to produce a 3D model, which can then be manipulated by the physician. 3D ultrasounds are produced using a somewhat similar technique.

With the ability to visualize important structures in great detail, 3D visualization methods are a valuable resource for the diagnosis and surgical treatment of many pathologies. It was a key resource for the famous, but ultimately unsuccessful attempt by Singaporean surgeons to separate Iranian twins Ladan and Laleh Bijani in 2003. The 3D equipment was used previously for similar operations with great success.

Other proposed or developed techniques include:

- Diffuse optical tomography
- Elastography
- Electrical impedance tomography
- Optoacoustic imaging
- Ophthalmology

- A-scan
- B-scan
- Corneal topography
- Heidelberg retinal tomography
- Optical coherence tomography
- Scanning laser ophthalmoscopy

Some of these techniques are still at a research stage and not yet used in clinical routines.

Non-Diagnostic Imaging

Neuroimaging has also been used in experimental circumstances to allow people (especially disabled persons) to control outside devices, acting as a brain computer interface.

Medical Imaging Service

This is a specialized area of medical equipment service and repair, which is separate from the biomedical field, although a hospital with their own service group may include them in the biomed department. At one time, there were only two ways to receive training for this field. One was to learn it in the military, and the other was on-the-job training (OJT) from the manufacturer. But since the 1980s several independent training centers have been started. One such school is RSTI.

There are different means of employment in this occupation. Working for the manufacturer's field service department (OEM), working for a hospital (in-house), and working for an independent (outside, or independent provider). The most stable positions are with the OEM or hospital, as you can remain current through on-going training, and the two have good working relationships.

The OEM service engineer can expect to spend a lot of time driving from one site to another during the work day, and working non-standard hours. They will install, remove, diagnose, repair, calibrate, perform preventive maintenance, and interface equipment, all while ensuring good customer relations. You may also be required to do yearly testing of the radiation sources for Federal and State compliance.

The in-house person is employed by the hospital. With larger medical facilities, travel between the hospitals' other locations may be required to perform the required services. You may also be required to do yearly testing of the radiation sources for compliance. The OEM or independent will provide installation of purchased equipment, and can be used for back-up service.

An independent is typically someone who has left an OEM, and started their own service business. Staying up-to-date as an independent can be difficult and expensive, as the OEM is usually reluctant to provide training. However, non-OEM training facilities are available, such as the aforementioned RSTI.

Competition for service can be aggressive, with OEM's giving hospitals or clinics a reduction in equipment purchase price if they retain some form of OEM service. The independent may also sell and install refurbished equipment, or de-install equipment.

They will repair, calibrate, and perform preventative maintenance. Because many of the tasks associated with imaging service require expensive, specialized equipment, there may be a financial limit to the independent.

Typical equipment used routinely are a Storage Oscilloscope and multimeter (if servicing old vacuum-state equipment, a VOM would be helpful). Additional equipment: Keithley dosimeter, mAs meter, Biddle contact tachometer, light to radiation template, etc.

Magnetic resonance imaging

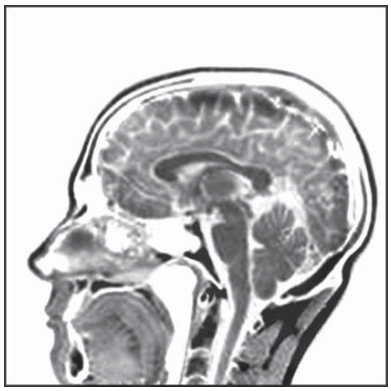


Fig. Magnetic Resonance Image Showing a Median Sagittal Cross Section Through a Human Head.

Magnetic resonance imaging (MRI) is primarily used in medical imaging to visualize the structure and function of the body. It provides detailed images of the body in any plane. MR has much greater soft tissue contrast than Computed tomography (CT) making it especially useful in neurological, musculoskeletal, cardiovascular and oncolological diseases. Unlike CT it uses no ionizing radiation.

The scanner creates a powerful magnetic field which aligns the magnetization of hydrogen atoms in the body. Radio waves are used to alter the alignment of this magnetization. This causes the hydrogen atoms to emit a weak radio signal which is amplified by the scanner.

This signal can be manipulated by additional magnetic fields to build up enough information to reconstruct an image of the body.

Magnetic resonance spectroscopy is used to measure the levels of different metabolites in body tissues. The MR signal produces spectrum of difference resonances that correspond to different molecular arrangements of the isotope being "excited". This signature is used to diagnose certain metabolic disorders, especially those affecting the brain, as well as to provide information on tumor metabolism.

The scanners used in medicine have a typical magnetic field strength of 0.2 to 3 teslas. Construction costs approximately US$ 1 million per tesla and maintenance an additional several hundred thousand dollars per year. Research using MRI scanners operating at ultra high field strength (up to 21.1 tesla) can produce images of the mouse brain with a resolution of 18 micrometres.

Background

Nomenclature

Magnetic resonance imaging was developed from knowledge gained in the study of nuclear magnetic resonance. In its early years MRI was referred to as nuclear magnetic resonance imaging (NMRI), but the word nuclear has been associated with ionizing radiation exposure, which is not used in an MRI, so to prevent patients from making a negative association between MRI and ionizing radiation, the word has been almost universally removed. Scientists still use the term *NMRI* when discussing non-medical devices operating on the same principles.

One of the inventors of MRI, Paul Lauterbur, originally named the technique *zeugmatography*, a Greek term meaning "that which is used for joining". The term referred to the

interaction between the static and the gradient magnetic fields necessary to create an image, but the nomenclature never caught on.

Brief Explanation For the Layperson

MRI works by making certain atoms in the body emit radiowaves by using a powerful magnet.

When a person is in the scanner some of the hydrogen atoms, which are mainly found in water, align with this magnetic field. A radiowave at just the right frequency makes these atoms resonate and absorb energy.

The atoms then release this energy in the form of a very weak radiowave that can be measured by the scanner and amplified.

Extra magnetic fields are used to constantly change the magnetic field to allow images of the body to be reconstructed. These fields are created by gradient coils which make the familiar banging sounds of an MRI scan. Contrast agents may be injected to demonstrate blood vessels or inflammation in the tissues.

Unlike CT scanning MRI uses no ionizing radiation and is generally a very safe procedure.

Patients with some metal implants and cardiac pacemakers are prevented from having an MRI due to effect of the powerful magnetic field.

MRI is used to image every part of the body, but is particularly useful in neurological conditions, disorders of the muscles and joints, for evaluating tumors and showing abnormalities in the heart and blood vessels.

Principle

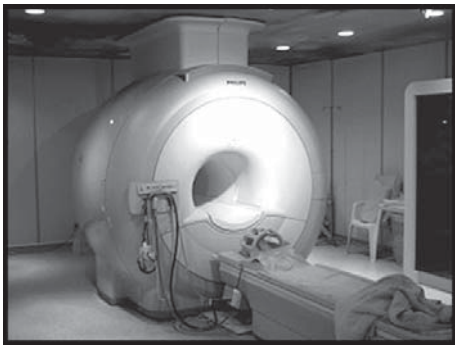


Fig. Modern 3 Tesla Clinical MRI Scanner.

Magnetism

Elementary subatomic particles such as protons have the quantum mechanical property of spin. Nuclei such as ^{1}H or ^{31}P, with an odd number of nucleons, always have a non–zero spin and therefore a magnetic moment. Some other isotopes such as ^{12}C have no unpaired neutrons or protons, and no net spin. When these spins are placed in an external magnetic field they start to precess around the direction of that field.

The magnetic field also creates two energy states that the protons can occupy which are separated by a quantum of energy. The thermal energy of the sample causes the molecules to tumble leaving only a very small excess of protons to cause magnetic polarization.

Resonance

The energy difference between the proton energy states corresponds to electromagnetic radiation at radio frequency wavelengths. Resonant absorption of energy by the protons due to an external oscillating magnetic field (radio wave) will occur at the Larmor frequency.

The net magnetization vector has two components. The longitudinal magnetization is due to an excess of protons in the lower energy state. This gives a net polarization parallel to the external field.

The transverse magnetization is due to coherences forming between the two proton energy states. This gives a net polarization perpendicular to external field in the transverse plane. The recovery of longitudinal magnetization is called T1 relaxation and the loss of phase coherence in the transverse plane is called T2 relaxation.

When the radio frequency pulse is turned off, the transverse vector component produces an oscillating magnetic field which induces a small current in the receiver coil. This free induction decay (FID) lasts only a few milliseconds before the thermal equilibrium of the spins is restored. The actual signal that is measured by the scanner is formed by a refocusing gradient or radio wave to create a gradient or spin-echo.

Imaging

Slice selection is achieved by applying a magnetic gradient in addition to the external magnetic field during the radio frequency pulse. Only one plane within the object will have protons that are on–resonance and contribute to the signal.

A real image can be considered as being composed of a number of spatial frequencies at different orientations. A two–dimensional Fourier transformation of a real image will express these waves as a matrix of spatial frequencies known as k–space.

Low spatial frequencies are represented at the centre of k–space and high spatial frequencies at the periphery.

Frequency and phase encoding are used to measure the amplitudes of a range of spatial frequencies within the object being imaged.

The frequency encoding gradient is applied during readout of the signal and is orthogonal to the slice selection gradient. During application of the gradient the frequency differences in the readout direction progressively change.

At the midpoint of the readout these differences are small and the low spatial frequencies in the image are sampled filling the centre of k-space. Higher spatial frequencies will be sampled towards the beginning and end of the readout filling the periphery of k-space.

Phase encoding is applied in the remaining orthogonal plane and uses the same principle of sampling the object for different spatial frequencies. However, it is applied for a brief period before the readout and the strength of the gradient is changed incrementally between each radio frequency pulse. For each phase encoding step a line of k–space is filled.

Magnet

The magnet is the largest and most expensive component of the scanner, and the remainder of the scanner is built around it. Just as important as the strength of the main magnet is its precision. The straightness of magnet lines within the centre or, as it is known as, the iso-centre of the magnet, needs to be nearly perfect.

This is known as homogeneity. Fluctuations (non-homogeneities in the field strength) within the scan region should be less than three parts per million (3 ppm). *Three types of magnet have been used:*

- *Permanent magnet:* Conventional magnets made from ferromagnetic materials (e.g., steel) can be used to provide the static magnetic field. These are extremely bulky (the magnet can weigh in excess of 100 tonnes), but once installed require little costly maintenance. Permanent magnets can only achieve limited field strength (usually < 0.4 T) and have limited stability and precision. There are also potential safety issues, as the magnetic field cannot be removed in case of entrapment.
- *Resistive electromagnet:* A solenoid wound from copper wire is an alternative to a permanent magnet. The advantages are low cost, but field strength is limited, and stability is poor. The electromagnet requires considerable electrical energy during operation which can make it expensive to operate. This design is essentially obsolete.
- *Superconducting electromagnet:* When a niobium-titanium alloy is cooled by liquid helium at 4 K (“269 °C, “452 °F) it becomes superconducting where it loses all resistance to flow of electrical current. By building an electromagnet from superconducting wire, it is possible to develop extremely high field strengths, with very high stability. The construction of such magnets is extremely costly, and the cryogenic helium is expensive and difficult to handle. However, despite its cost, helium cooled superconducting magnets are the most common type found in MRI scanners today.

Most superconducting magnets have their coils of superconductive wire immersed in liquid helium, inside a vessel called a cryostat. Despite thermal insulation, ambient heat causes the helium to slowly boil off. Such magnets, therefore, require regular topping-up with helium.

Generally a cryocooler, also known as a coldhead is used to recondense some helium vapour back into the liquid helium bath. Several manufacturers now offer 'cryogenless' scanners, where instead of being immersed in liquid helium the magnet wire is cooled directly by a cryocooler.

Magnets are available in a variety of shapes. However, permanent magnets are most frequently 'C' shaped, and superconducting magnets most frequently cylindrical. However, C-shaped superconducting magnets and box-shaped permanent magnets have also been used.

Magnetic field strength is an important factor determining image quality. Higher magnetic fields increase signal-to-noise ratio, permitting higher resolution or faster scanning.

However, higher field strengths require more costly magnets with higher maintenance costs, and have increased safety concerns. 1.0 - 1.5 T field strengths are a good compromise between cost and performance for general medical use. However, for certain specialist uses (e.g., brain imaging), field strengths up to 3.0 T may be desirable.