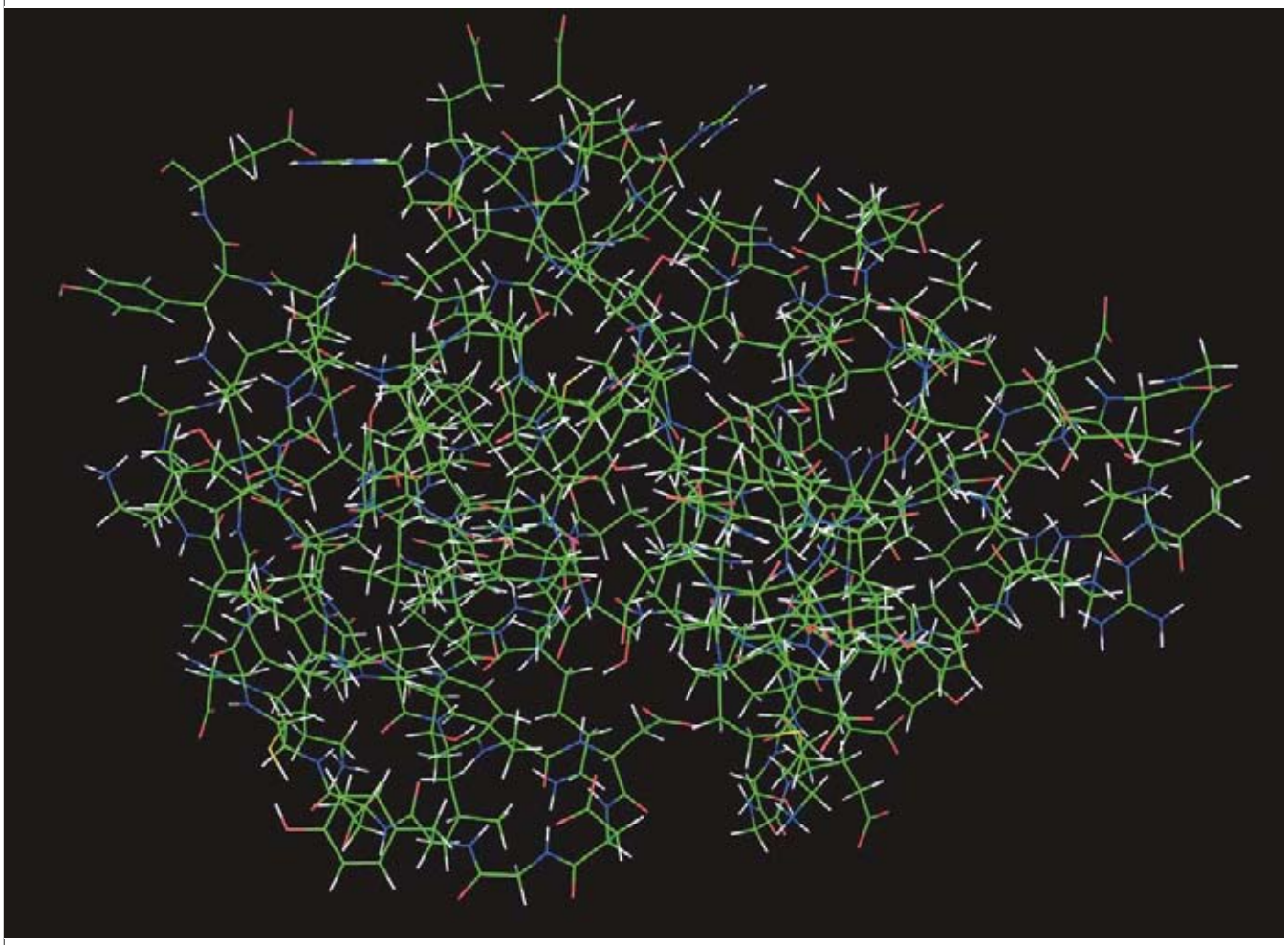
NMR for protein analysis

Proteins have a stunning array of functions closely tied in to their molecular structure. Nuclear magnetic resonance spectroscopy offers vital input to protein science, not only for determining 3D protein structure with atomic resolution but also for elucidating the mechanisms governing dynamic protein assembly.



CEA

The chemical nature and spatial arrangement of a protein's amino acids make it functionally active and lend it specific biochemical properties. NMR is a first-choice tool for determining the three-dimensional structure of proteins. (The protein shown here with all its atoms is the same protein as depicted in figure 3 and with the same orientation.)

All biological **macromolecules** in general, and **proteins** in particular, are formed by hundreds or even thousands of **atoms**. To be activated, they fold into a specific, atomically-organised spatial conformation called a *three-dimensional protein structure*. There are several methods for determining this structure, differentiated by their **resolution**. Two techniques allow to reach the atomic resolution: one is X-ray crystallography(1), the other is **nuclear magnetic resonance (NMR) spectroscopy** (see *NMR, an ever-advancing spectroscopic technique*, p. 56). Determining the three-dimensional structure of macromolecules provides insight into the molecular mechanisms of cellular machineries and in particular the specificity of the interactions involved in living organisms. Three-dimensional structure is also the key to designing new compounds able to act specifically on certain pharmacological targets. Determining protein structure is therefore the pivotal first step towards rational drug design.

(1) X-ray crystallography: method of determining the crystal structure of compounds via X-ray scattering through **crystals**.

A powerful tool for molecular biology

Protein nuclear magnetic resonance consists in plunging an aqueous sample of protein into an ultra-high **magnetic field** (of between 10 and 20 **tesla**). The **nuclear spins** of the protein's atoms align to the main axis of the magnetic field. It is then possible to perturb the spin energy balances via a series of **electromagnetic** pulses and measure the current induced when the spins relax. The measurements are made on a receiver **coil** located in the

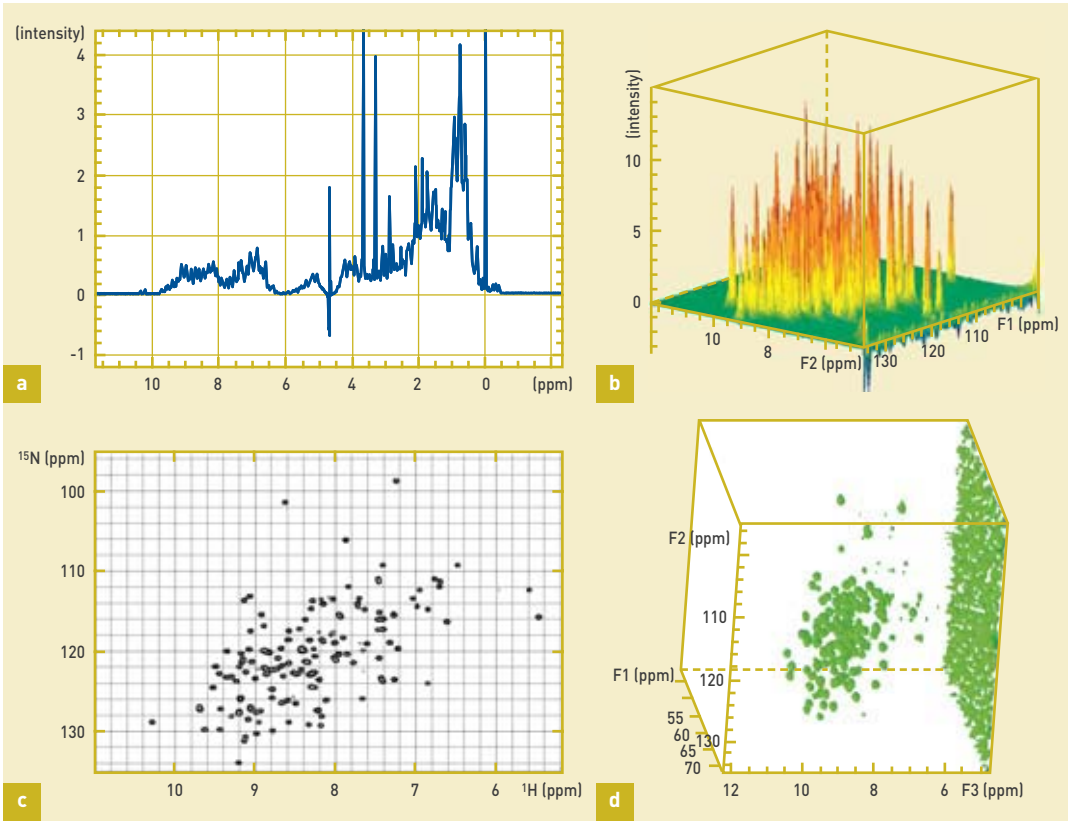


Figure 1.
Examples of NMR spectra, recorded at 14.5 T, for a protein with 150-odd residues. Spectrum *a* is a one-dimensional proton spectrum. The signal corresponding to the protons of a solvent, in this case water at 4.5 ppm, is deleted by a well-targeted sequence. The 1,200 proton signals of the protein form a continuous spectrum. In *b*, a two-dimensional HSQC (Heteronuclear Single Quantum Correlation) spectrum of a protein uniformly labelled with ^{15}N nitrogen, graphed as an oblique view. Each peak matches a magnetic interaction between the amide[6] proton, whose chemical shift is given as the x-coordinate, and the attached nitrogen, whose chemical shift is given in the y-coordinate. In *c*, the same spectrum as in *b* but graphed with contouring. In *d*, a three-dimensional HNCA spectrum of a protein uniformly labelled with ^{15}N nitrogen and ^{13}C carbon. Each peptide[7] N-H group, as characterised by chemical shifts of its nitrogen atom in dimension 2 and its amide proton in dimension 3, leads to two peaks, which are highlighted here in green: the dimension 1 coordinate of one peak is the chemical shift of the preceding Cα carbon, whereas the dimension 1 coordinate of the other peak is the chemical shift of the following Cα carbon, along the polypeptide[7] chain. The signal on the right is the residual signal from the water.

vicinity of the sample. This NMR signal consists of the combined contributions of all the different atoms in solution. A **Fourier transform** of the signal allows to extract all the resonance frequencies of the spins observed. Obviously, only atoms with a non-zero nuclear spin give rise to observable signals. Since proteins are almost exclusively composed of carbon, oxygen, nitrogen and **hydrogen**, the preferred observables in protein NMR were traditionally **protons**. The resonance frequency of protons is modulated by their electronic environment, and it will differ by a few parts per million (ppm) from a reference frequency. The ratio of the resonance frequency to a reference frequency is called *chemical shift*, and is expressed in ppm. Chemical shift is highly sensitive to the type of atom to which the proton is attached (carbon, oxygen, nitrogen), to its involvement in non-**covalent** bonding like hydrogen bonding[2], to the proximity of aromatic nuclei[3], as well as to a **paramagnetic** centre, a metal **ion** or radical[4] containing an unpaired **electron**. Each proton therefore resonates to a distinct chemical shift.

Proteins have so many protons – around 1,200 for a protein with 150 residues[5] – that many signals overlap (Figure 1a). There are relatively few real applications for *one-dimensional proton NMR* spectra. Nevertheless, they are spectra that take just seconds to record, and they can give a quickfire indication as to whether the protein adopts the three-dimensional structure or whether it presents significant structural variations.

Two-dimensional NMR (Figures 1b and 1c) offers greater potential and the spectra are still pretty quick to generate, taking a few minutes or up to a couple of hours. Different types of spectra can be recorded by deploying elaborate sequences of high-field electromagnetic pulses at delays of just microseconds or even milliseconds. The **magnetisation** of the nuclear spin initially disturbed during these sequences is transferred to neighbouring atoms, either through space via what is known as the *Nuclear Overhauser Effect (nOe)*, or through the chemical bonds. The Swiss physical chemist Richard R. Ernst was awarded the 1991 Nobel Prize in Chemistry for these discoveries. When dealing with small proteins containing less than 70 **amino acids**, two-dimensional NMR spectra can be analysed to assign a chemical shift value to each proton and determine the three-dimensional structure of the protein via the strategy detailed below.

Advances made in molecular biology in the 1990s made it possible to produce proteins uniformly labelled with **stable isotopes** of nitrogen and carbon, ^{15}N nitrogen and ^{13}C carbon, both of which

(2) A hydrogen bond is formed between a proton-donor atom and a proton acceptor. The strength of the hydrogen bond will depend on the angle and distance between proton donor and proton acceptor atoms (less than 3.5 ångströms, 1 Å = 10^{-10} m). It is twenty times weaker than a covalent bond formed between two atoms that saturate their final atomic orbit by sharing one or more pairs of electrons.

(3) Aromatic (nucleus): a compound presenting delocalised electrons cycling around generally carbon rings.

(4) Radical: a chemical species (widely known as a free radical) that possesses one or more unpaired electrons on its outer shell. This means the atom has a non-zero electron spin.

(5) Residue: a coverall term to describe the building blocks in a **polymer** chain, which in the case of proteins are amino acids.

(6) Amide: molecule containing a carbonyl group (C=O) bound to an amine (NH_2). Adjacent amino acids in a protein molecule are linked together by amide groups.

(7) Peptide: a molecule formed of a short chain of a few amino acids. Polypeptide: molecule built from a long chain of several dozen peptide units.

share the same half-integer spin as the proton. This means that an NMR signal can now be recorded for virtually all the atoms of a protein except oxygen atoms. This drove the development and widespread use of *three-dimensional spectra* (Figure 1d), which makes it possible to correlate a proton to the carbon or nitrogen atom to which it is attached and to adjacent atoms in space or through their chemical bonds. Furthermore, the continual improvement in pulse sequences – work for which Swiss chemist Kurt Wüthrich was recompensed with the 2002 Nobel Prize in Chemistry – has lifted many of the barriers to larger protein sizes, and NMR can now study proteins of up to 50 **kDa** or more.

CEA

16.4-T NMR spectrometer **magnet** used to study three-dimensional protein structure. Inside this magnet, the hydrogen nuclei, i.e. protons, have a resonance frequency of 700 **MHz**.

Mapping the regions of protein-protein interaction

NMR has proven particularly well-geared to efficiently mapping the regions of interaction between macromolecules. The quickest way to achieve this is to record a two-dimensional ^{15}N HSQC spectrum for a protein uniformly labelled with ^{15}N nitrogen with and without an unlabelled biological partner like a **DNA**, a **RNA**, another protein, or even a small chemical **ligand**. This type of spectrum makes it possible to edit the chemical shift of all the protons attached to nitrogen atoms (Figure 2). It is almost like the protein's fingerprint. Chemical shifts of N-H protons are particularly sensitive to their environment (pH, temperature, distance in space to other atoms). It is therefore possible to image the surfaces involved in the interaction by exploiting the signal modifications that occur when the partner is added. One of the limitations to this method is the protein concentration needed to make NMR detection workable (from 100 **µM** to 1 mM). However, recently developed cryoprobes and nanoprobes can cut the quantity of product needed to just a few dozen **micromoles**. This approach has been extended to screening new ligands or new inhibitors of pharmacological targets, by coupling spectrometers with sample-changers.

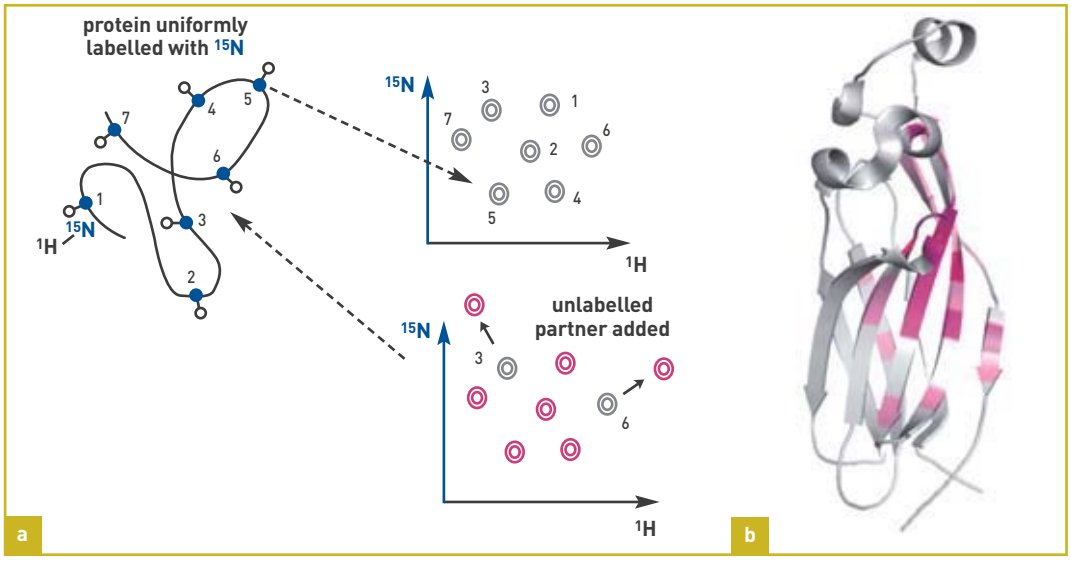


Figure 2.
Mapping variations in chemical shift to determine the regions of protein-protein interaction. Part *a* illustrates the underlying principle: the protein of interest is uniformly labelled with ^{15}N nitrogen. Then, the HSQC spectrum generated is used to correlate the proton and nitrogen chemical shifts of each N-H peptide[7] bond. Adding the unlabelled partner induces variations in the chemical shift of certain signals corresponding to amino acids interacting with it. These residues are integrated into the protein structure to define a region of interaction. Part *b* is an example of an application employed to determine the region of interaction of a histone chaperone with the histones[8].

CEA

Figure 3.
NMR determination of the three-dimensional structure of proteins. In *a*, a set of restraints establishes a matrix of distances defining the protein structure. The protein is shown as a ribbon, with the distance restraints in white. In *b*, when the structure resolution procedure has been completed, the result is a set of very similar structures, all of which are compatible with the experimental data. In *c*, a ribbon diagram of the protein structure. In this example, the protein contains two mainly β-sheet domains (in yellow and green), one helix (in red), and flexible links (in blue) (G. Charier, J. Couprie, B. Alpha-Bazin, V. Meyer, E. Quemeneur, R. Guerois, I. Callebaut, B. Gilquin and S. Zinn-Justin, *Structure*, 12, p. 1551-1562, 2004). This is the same protein as on p. 52 and with the same orientation.

Determining protein structure

Determining protein structure from NMR data is a two-step process. Step 1 is to inventory the chemical shifts of all the observable nuclei in the protein, ^{1}H, ^{15}N, ^{13}C, which is done using three-dimensional spectra correlating the atoms through their chemical bonds. Step 2 involves analysing the nOe experiments to edit the correlations resulting from the spatial proximity of the different nuclei. Correlations are assigned to coupled nuclei and then introduced into a molecular **modelling** procedure as distance restraints. Taken together, the full dataset of distance restraints will produce a matrix with which the structural model of the protein has to converge (Figure 3a). Other NMR-generated experimental data (like chemical shifts and J-coupling constants) yielding angular data characterising the protein structure are also used as inputs to the model. The nOe correlations can be assigned manually if the proteins are small (less than 70 residues) but larger proteins can feature many atoms giving overlapping chemical shifts. If so, any ambiguity in the nOe correlations assigned has to be eliminated. Automatic correlation assignment tools have been developed to speed up this otherwise time-consuming task. These computer programmes are based on an iterative procedure that progressively removes the ambiguities by generating molecular models. The quality and convergence of the selected structures are key criteria for the quality of the protein structure. The end-product of the procedure is a family of very similar structures, all of which are compatible with the experimental data (Figure 3b). The polypeptide(7) backbone includes canonical elements of secondary structure (α helix and β sheets) that can be pictured schematically (Figure 3c). A detailed analysis of side-chain position according to side-chain type (hydrophobic or hydrophilic) will indicate the kind of interactions holding the structure together and highlight those that are important to interactional specificity.

The method has recently been extended to solving the structures of macromolecular complexes(9) and protein-small molecule complexes, including pharmaceutical compounds. To date, NMR input has produced 16% of all 44,000 known biological macromolecule structures.

Combining NMR with other techniques

NMR applied to biological macromolecules has several differentiating advantages over crystallography: the structure experiments do not damage the protein samples and are conducted in solution state, whereas protein X-ray crystallography needs to start with a very-high-quality protein crystal. NMR can also highlight protein conformational fluctuations, which often play a major role in intermolecular recognition or in the internal dynamics of **enzymes**. It is possible to accurately characterise fluctuating or partially folded protein states, which is a key step towards understanding the processes underpinning protein folding. Furthermore, although NMR remains relatively limited in terms of its ability to generate high-resolution information at broad cellular machinery scale, it does yield information on the dynamic properties and transient bonding of very large protein complexes, such as the proteasome(10, 11) (670 kDa) or the GroEL chaperone(12) complex (800 kDa). In all likelihood, it will be NMR combined with other techniques that will provide the key to understanding the molecular mechanisms involved in the biological functions of large molecular assemblies. This technical coupling could bring together low-resolution methods like electron microscopy, homology modelling, or possibly even *ab initio* structure prediction which has made huge progress in recent years. Preliminary results from new approaches strongly suggest that we are on the verge of characterising the structure of proteins directly in their physiological setting, in cell cultures.

> **Françoise Ochsenbein** and **Bernard Gilquin**
Institute of Biology and Technologies - Saclay
Life Sciences Division
CEA Saclay Centre

(8) Histone: the basic building block-protein associated with chromosomal DNA. A chaperone is a protein that helps other proteins to avoid the misfolding that produces inactive or aggregate states.

(9) M. Agez, J. Chen, R. Guerois, C. van Heijenoort, J.-Y. Thuret, C. Mann and F. Ochsenbein, *Structure*, 15, p. 191-199, 2007.

(10) Proteasome: large protein complex that regulates the degradation of proteins marked for destruction.

(11) R. Sprangers and L. E. Kay, *Nature*, 445, p. 618-622, 2007.

(12) Chaperone: the special feature of chaperone proteins is that they assist with protein folding and the assembly of complexes.

The different types of magnetism

The origins of magnetism lie in the properties of electrons as explained by the laws of quantum physics. Part of an electron's magnetic properties (*spin magnetism*) results from its quantum-mechanical spin state, while another part results from the orbital motion of electrons around an atom's nucleus (*orbital magnetism*) and from the magnetism of the nucleus itself (nuclear magnetism). This is put to use, in particular, for nuclear magnetic resonance imaging in the medical field. Magnetism is therefore produced by electric charges in motion. The force acting on these charges, called the Lorentz force, demonstrates the presence of a magnetic field.

Electrons have an intrinsic magnetic dipole moment (the magnetic quantum state being the Bohr *magneton*), which can be pictured as an electron's rotational motion of spin around itself in one direction or another, oriented either upwards or downwards. The *spin quantum number* (one of the four numbers that 'quantifies' the properties of an electron) equals 1/2 (+ 1/2 or - 1/2). A pair of electrons can only occupy the same *orbital* if they have opposite magnetic dipole moments.

Each atom acts like a tiny magnet carrying an intrinsic magnetic dipole moment. A nucleus (the neutron and proton individually have a half-integer spin) will have a half-integer spin if it has an odd atomic mass number; zero spin if the atomic mass number and charge are even, and an integer spin if the atomic mass number is even and the charge odd.

On a larger scale, several magnetic moments can together form magnetic domains in which all these moments are aligned in the same direction. These spatial regions are separated by domain walls. When grouped together, these domains can themselves form a macroscopic-scale magnet (Figure E1).

The type of magnetism that comes into play is determined by how these elementary constituents are ordered, and is generally associated with three main categories of material: *ferromagnetic*, *paramagnetic* and *diamagnetic*.

Any material that is not diamagnetic is by definition paramagnetic provided that its magnetic susceptibility is positive. However, ferromagnetic materials have particularly high magnetic susceptibility and therefore form a separate category.

1. Ferromagnetic materials are formed of tiny domains inside which atoms exhibiting parallel magnetisation tend to align themselves in the direction of an external magnetic field like elementary dipoles. In fact, the magnetic moments of each atom can align themselves spontaneously within these domains, even in the absence of an external magnetic field. Applying an external field triggers domain wall movement that tends to strengthen the applied field. If this field exceeds a certain value, the domain most closely oriented with the direction of the applied field will tend to grow at the expense of the other domains, eventually occupying the material's whole volume. If the field diminishes, the domain walls will move, but not symmetrically as the walls cannot fully reverse back to their original positions. This results in remanent magnetisation, which is an important feature of naturally occurring magnetite, or of magnets themselves.

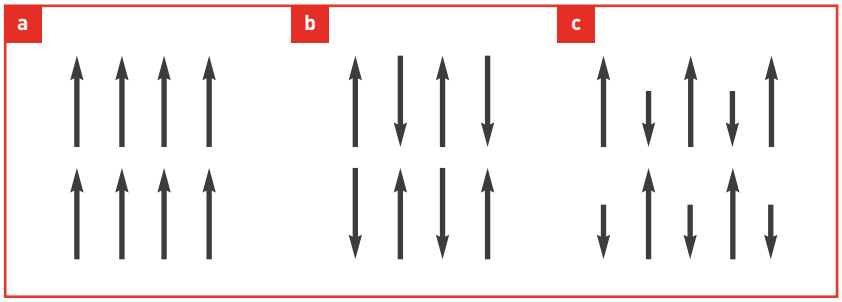


Figure E1.
Intrinsic magnetic dipole moments have parallel alignment in ferromagnetic materials (a), anti-parallel alignment but zero magnetisation in antiferromagnetic materials (b), and anti-parallel alignment with unequal moments in ferrimagnetic materials (c).

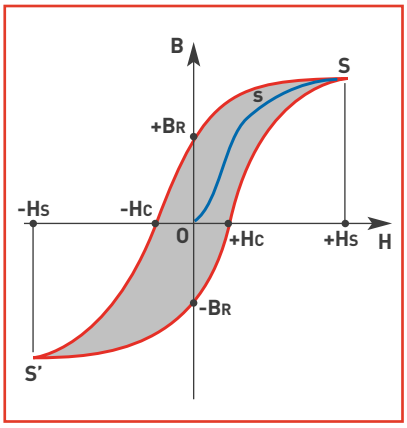


Figure E2.
The induction B of a magnetic material by a coil is not proportional to its magnetic excitation (*field H*). While the initial magnetisation forms an OsS-type curve, shown in blue in the figure, it reaches saturation at point s. Only a partial induction is retained if the field approaches zero; this remanent induction can only be cancelled out by reversing the magnetic field to a "coercive" field value. This hysteresis loop illustrates the losses due to "friction" between the magnetic domains shown on the area bounded by the magnetisation and demagnetisation curves.

The whole process forms a hysteresis loop, i.e. when the induced field is plotted against the applied field it traces out a *hysteresis curve* or *loop* where the surface area represents the amount of energy lost during the irreversible part of the process (Figure E2). In order to cancel out the induced field, a coercive field has to be applied: the materials used to make artificial permanent magnets have a high coercivity.

Ferromagnetic materials generally have a zero total magnetic moment as the domains are all oriented in different directions. This ferromagnetism disappears above a certain temperature, which is known as the Curie Temperature or Curie point.

The magnetic properties of a given material stem from the way the electrons in the metallic cores of a material or of a transition metal complex collectively couple their spins as this results in all their spin moments being aligned in the same direction.

Materials whose atoms are widely distributed throughout their crystal structure tend to better align these elementary magnets via a coupling effect. This category of materials, which is characterised by a very high positive magnetic

Stoiber Productions, München

A Transrapid train using magnetic levitation arriving at the Long Yang bus station in Shanghai (China). This German-built high-speed, monorail train was commissioned in 2004 to service the rail link to Pudong international airport.

susceptibility, includes iron, cobalt and nickel and their **alloys**, steels in particular, and some of their compounds, and, to a lesser extent, some **rare earth** metals and alloys with large crystal lattices, and certain combinations of elements that do not themselves belong to this category. In **ferrimagnetic** materials, the magnetic domains group into an anti-parallel alignment but retain a non-zero magnetic moment even in the absence of an external field. Examples include magnetite, ilmenite and iron oxides. Ferrimagnetism is a feature of materials containing two types of atoms that behave as tiny magnets with magnetic moments of unequal magnitude and anti-parallel alignment. **Antiferromagnetism** occurs when the sum of a material's parallel and anti-parallel moments is zero (e.g. chromium or haematite). In fact, when atoms are in a close configuration, the most stable magnetic arrangement is an anti-parallel alignment as each magnet balances out its neighbour so to speak (Figure E1).

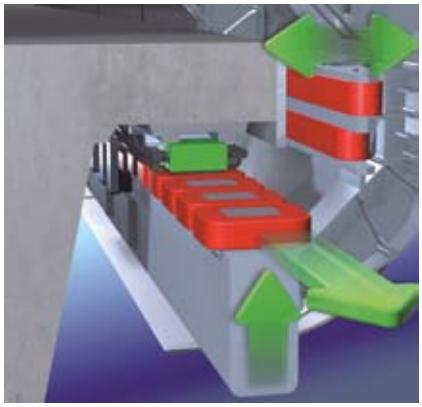
2. **Paramagnetic** materials behave in a similar way to ferromagnetic materials, although to a far lesser degree (they have a positive but very weak magnetic susceptibility of around 10^{-3}). Each atom in a paramagnetic material has a non-zero magnetic moment. In the presence of an external magnetic field, the magnetic moments align up, thus amplifying this field. However, this effect decreases as temperature rises since the thermal agitation disrupts the alignment of the elementary dipoles. Paramagnetic materials lose their magnetisation as soon as they are released from the magnetic field. Most metals, including alloys comprising ferromagnetic elements are paramagnetic, as are certain minerals such as pegmatite.

3. **Diamagnetic** materials exhibit a negative and an extremely weak magnetic susceptibility of around 10^{-5}. The magnetisation induced by a magnetic field acts in the opposite direction to this field and tends to head away from **field lines** towards areas of lower field strengths. A perfect diamagnetic material would offer maximum resistance to an external magnetic field and exhibit zero **permeability**. Metals such as silver, gold, copper, mercury or lead, plus quartz, graphite, the noble gases and the majority of organic compounds are all diamagnetic materials.

In fact, all materials exhibit diamagnetic properties to a greater or lesser extent, resulting from changes in the orbital motion of electrons around atoms in response to an external magnetic field, an effect that disappears once the external field is removed. As Michael Faraday showed all that time ago, all substances can be "magnetised" to a greater or lesser degree provided that they are placed within a sufficiently intense magnetic field.

Electromagnetism

It was the Danish physicist Hans Christian Ørsted, professor at the University of Copenhagen, who, in 1820, was first to discover the relationship between the hitherto separate fields of **electricity** and **magnetism**. Ørsted showed that a compass needle was deflected when an electric current passed through a wire, before Faraday had formulated the physical law that carries his name: the magnetic field produced is proportional to the intensity of the current. **Magnetostatics** is the study of static magnetic fields, i.e. fields which do not vary with time.



Close-up of the magnets used to guide and power the train.

Magnetic and **electric fields** together form the two components of **electromagnetism**. **Electromagnetic waves** can move freely through space, and also through most materials at pretty much every frequency band (radio waves, microwaves, infrared, visible light, ultraviolet light, X-rays and gamma rays). **Electromagnetic fields** therefore combine electric and magnetic **force fields** that may be natural (the Earth's magnetic field) or man-made (low frequencies such as electric power transmission lines and cables, or higher frequencies such as radio waves (including cell phones) or television.

Mathematically speaking, the basic laws of electromagnetism can be summarised in the four **Maxwell equations** (or **Maxwell-Lorentz equations**) which can be used to provide a coherent description of all electromagnetic phenomena from electrostatics and magnetostatics to electromagnetic wave propagation. James Clerk Maxwell set out these laws in 1873, thirty-two years before Albert Einstein incorporated the theory of electromagnetism in his *special theory of relativity*, which explained the incompatibilities with the laws of classical physics.