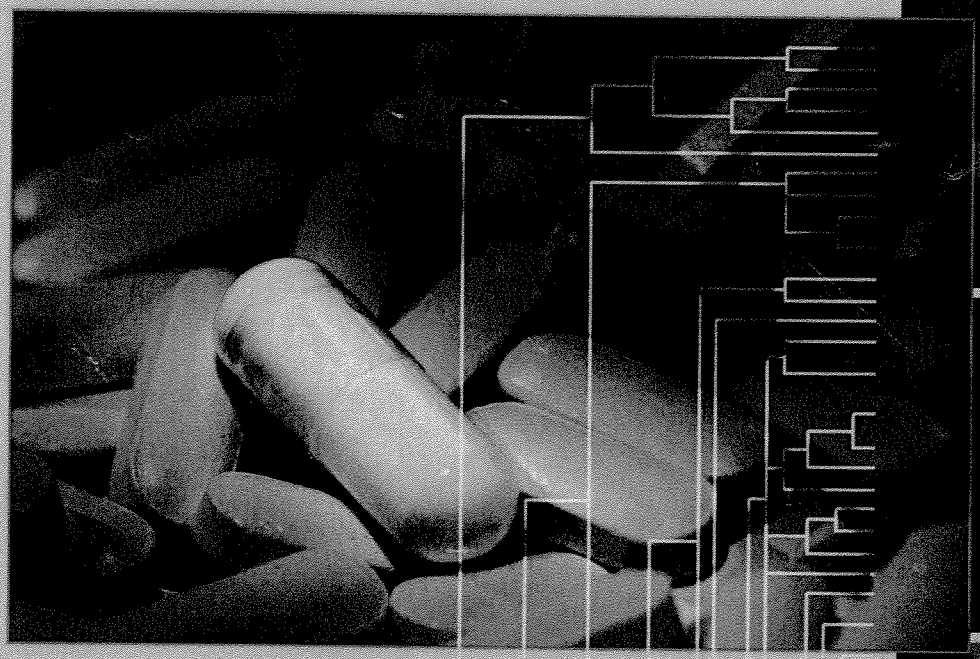
DRUG DISCOVERY AND DEVELOPMENT

TECHNOLOGY IN TRANSITION



Edited by H P RANG

Foreword by DANIEL VASELLA

CHURCHILL
LIVINGSTONE
ELSEVIER

1 The development of the pharmaceutical industry

H P Rang

Antecedents and origins

Our task in this book is to give an account of the principles underlying drug discovery as it happens today, and to provide pointers to the future. The present situation, of course, represents merely the current frame of a long-running movie. To understand the significance of the different elements that appear in the frame, and to predict what is likely to change in the next few frames, we need to know something about what has gone before. In this chapter we give a brief and selective account of some of the events and trends that have shaped the pharmaceutical industry. Most of the action in our metaphorical movie happened in the last century, despite the film having started at the birth of civilization, some 10 000 years ago. The next decade or two will certainly see at least as much change as the past century.

Many excellent and extensive histories of medicine and the pharmaceutical industry have been published, to which readers seeking more detailed information are referred (Mann, 1984; Sneader, 1985; Weatherall, 1990; Porter, 1997; see also Drews, 2000, 2003).

Disease has been recognized as an enemy of humankind since civilization began, and plagues of infectious diseases arrived as soon as humans began to congregate in settlements about 5000 years ago. Early writings on papyrus and clay tablets describe many kinds of disease, and list a wide variety of herbal and other remedies used to treat them. The earliest such document, the famous Ebers papyrus, dating from around 1550BC, describes more than 800 such remedies. Disease was in those times regarded as an affliction sent by the gods; consequently, the remedies were aimed partly at neutralizing or purging the affliction, and partly at appeasing the deities. Despite its essentially theistic basis, early medicine nevertheless discovered, through empiricism and common sense, many plant extracts whose pharmacological properties we recognize and still use today; their active principles include opium alkaloids, ephedrine, emetine, cannabis, senna and many others[1].

In contrast to the ancient Egyptians, who would, one feels, have been completely unsympathetic to medical

[1]There were, it should be added, far more – such as extracts of asses' testicles, bats' eyes and crocodile dung – that never found their way into modern pharmacology.

science had they been time-warped into the 21st century, the ancient Greeks might have felt much more at home in the present era. They sought to understand nature, work out its rules and apply them to alleviate disease, just as we aim to do today. The Hippocratic tradition had little time for theistic explanations. However, the Greeks were not experimenters, and so the basis of Greek medicine remained essentially theoretical. Their theories were philosophical constructs, whose perceived validity rested on their elegance and logical consistency; the idea of testing theory by experiment came much later, and this aspect of present-day science would have found no resonance in ancient Greece. The basic concept of four humours – black bile, yellow bile, blood and phlegm – proved, with the help of Greek reasoning, to be an extremely versatile framework for explaining health and disease. Given the right starting point – cells, molecules and tissues instead of humours – they would quickly have come to terms with modern medicine. From a therapeutic perspective, Greek medicine placed rather little emphasis on herbal remedies; they incorporated earlier teachings on the subject, but made few advances of their own. The Greek traditions formed the basis of the prolific writings of Galen in the 2nd century AD, whose influence dominated the practice of medicine in Europe well into the Renaissance. Other civilizations, notably Indian, Arabic and Chinese, similarly developed their own medical traditions, which – unlike those of the Greeks – still flourish independently of the western ones.

Despite the emphasis on herbal remedies in these early medical concepts, and growing scientific interest in their use as medicines from the 18th century onwards, it was only in the mid-19th century that chemistry and biology advanced sufficiently to give a scientific basis to drug therapy, and it was not until the beginning of the 20th century that this knowledge actually began to be applied to the discovery of new drugs. In the long interim, the apothecaries' trade flourished; closely controlled by guilds and apprenticeship schemes, it formed the supply route for the exotic preparations that were used in treatment. The early development of therapeutics – based, as we have seen, mainly on superstition and on theories that have been swept away by scientific advances – represents prehistory as far as the development of the pharmaceutical industry is concerned, and there are few, if any, traces of it remaining[2].

[2]Plenty of traces remain outside the pharmaceutical industry, in the form of a wide variety of 'alternative' and 'complementary' therapeutic procedures, such as herbalism, moxibustion, reflexology and acupuncture, whose underlying principles originated in the prescientific era and remain largely beyond the boundaries of science. It may not be long, given the growing appeal of such approaches in the public's eye, before the mainstream pharmaceutical industry decides that it must follow this trend. That will indeed be a challenge for drug discovery research.

Therapeutics in the 19th century

Although preventive medicine had made some spectacular advances, for example in controlling scurvy (Lind, 1763) and in the area of infectious diseases, vaccination (Jenner, 1798), curtailment of the London cholera epidemic of 1854 by turning off the Broad Street Pump (Snow), and control of childbirth fever and surgical infections using antiseptic techniques (Semmelweis, 1861; Lister, 1867), therapeutic medicine was virtually non-existent until the end of the 19th century.

Oliver Wendell Holmes – a pillar of the medical establishment – wrote in 1860: '....I firmly believe that if the whole materia medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind – and the worse for the fishes' (see Porter, 1997). This may have been a somewhat ungenerous appraisal, for some contemporary medicines – notably digitalis, famously described by Withering in 1785, extract of willow bark (salicylic acid), and *Cinchona* extract (quinine) – had beneficial effects that were well documented. But on balance, Holmes was right – medicines did more harm than good.

We can obtain an idea of the state of therapeutics at the time from the first edition of the *British Pharmacopoeia*, published in 1864, which lists 311 preparations. Of these, 187 were plant-derived materials, only nine of which were purified substances. Most of the plant products – lemon juice, rose hips, yeast etc. – lacked any components we would now regard as therapeutically relevant, but some – digitalis, castor oil, ergot, colchicum – were pharmacologically active. Of the 311 preparations, 103 were 'chemicals' mainly inorganic – iodine, ferrous sulfate, sodium bicarbonate, and many toxic salts of bismuth, arsenic, lead and mercury – but also a few synthetic chemicals, such as diethyl ether and chloroform. The remainder were miscellaneous materials and a few animal products, such as lard, cantharidin and cochineal.

An industry begins to emerge

For the pharmaceutical industry, the transition from prehistory to actual history occurred late in the 19th century (3Q19C, as managers of today might like to call it), when three essential strands came together. These were: the evolving science of biomedicine (and especially pharmacology); the emergence of synthetic organic chemistry; and the development of a chemical industry in Europe, coupled with a medical supplies trade – the result of buoyant entrepreneurship, mainly in America.

Developments in biomedicine

Science began to be applied whole-heartedly to medicine – as to almost every other aspect of life – in the 19th century. Among the most important milestones from the point of view of drug discovery was the elaboration in 1858 of cell theory, by the German pathologist Rudolf Virchow. Virchow was a remarkable man: pre-eminent as a pathologist, he also designed the Berlin sewage system and instituted hygiene inspections in schools, and later became an active member of the Reichstag. The tremendous reductionist leap of the cell theory gave biology – and the pharmaceutical industry – the scientific foundation it needed. It is only by thinking of living systems in terms of the function of their cells that one can begin to understand how molecules affect them.

A second milestone was the birth of pharmacology as a scientific discipline when the world's first Pharmacological Institute was set up in 1847 at Dorpat by Rudolf Buchheim – literally by Buchheim himself, as the Institute was in his own house and funded by him personally. It gained such recognition that the university built him a new one 13 years later. Buchheim foresaw that pharmacology as a science was needed to exploit the knowledge of physiology, which was being advanced by pioneers such as Magendie and Claude Bernard, and link it to therapeutics. When one remembers that this was at a time when organic chemistry and physiology were both in their cradles, and therapeutics was ineffectual, Buchheim's vision seems bold, if not slightly crazy. Nevertheless, his Institute was a spectacular success. Although he made no truly seminal discoveries, Buchheim imposed on himself and his staff extremely high standards of experimentation and argument, which eclipsed the empiricism of the old therapeutic principles and attracted some exceptionally gifted students. Among these was the legendary Oswald Schmiedeberg, who later moved to Strasbourg, where he set up an Institute of Pharmacology of unrivalled size and grandeur, which soon became the Mecca for would-be pharmacologists all over the world.

A third milestone came with Louis Pasteur's germ theory of disease, proposed in Paris in 1878. A chemist by training, Pasteur's initial interest was in the process of fermentation of wine and beer, and the souring of milk. He showed, famously, that airborne infection was the underlying cause, and concluded that the air was actually alive with microorganisms. Particular types, he argued, were pathogenic to humans, and accounted for many forms of disease, including anthrax, cholera and rabies. Pasteur successfully introduced several specific immunization procedures to give protection against infectious diseases. Robert Koch, Pasteur's rival and near-contemporary, clinched the infection theory by observing anthrax and other bacilli in the blood of infected animals.

The founder of chemotherapy – some would say the founder of molecular pharmacology – was Paul Ehrlich (see Drews, 2004 for a mini-biography). Born in 1854 and trained in pathology, Ehrlich became interested in histological stains and tested a wide range of synthetic chemical dyes that were being produced at that time. He invented 'vital staining' – staining by dyes injected into living animals – and described how the chemical properties of the dyes, particularly their acidity and lipid solubility, influenced the distribution of dye to particular tissues and cellular structures. Thence came the idea of specific binding of molecules to particular cellular components, which directed not only Ehrlich's study of chemotherapeutic agents, but much of pharmacological thinking ever since. 'Receptor' and 'magic bullets' are Ehrlich's terms, though he envisaged receptors as targets for toxins, rather than physiological mediators. Working in Koch's Institute, Ehrlich developed diphtheria antitoxin for clinical use, and put forward a theory of antibody action based on specific chemical recognition of microbial macromolecules, work for which he won the 1908 Nobel Prize. Ehrlich became director of his own Institute in Frankfurt, close to a large dye works, and returned to his idea of using the specific binding properties of synthetic dyes to develop selective antimicrobial drugs.

At this point, we interrupt the biological theme at the end of the 19th century, with Ehrlich in full flood, on the verge of introducing the first designer drugs, and turn to the chemical and commercial developments that were going on simultaneously.

Developments in chemistry

The first synthetic chemicals to be used for medical purposes were, ironically, not therapeutic agents at all, but anaesthetics. Diethyl ether ('sweet oil of vitriol') was first made and described in 1540. Early in the 19th century, it and nitrous oxide (prepared by Humphrey Davy in 1799 and found – by experiments on himself – to have stupefacient properties) were used to liven up parties and sideshows; their usefulness as surgical anaesthetics was demonstrated, amid much controversy, only in the 1840s[3], by which time chloroform had also made its appearance. Synthetic chemistry at the time could deal only with very simple molecules, made

[3] An event welcomed, in his inimitable prose style, by Oliver Wendell Holmes in 1847: 'The knife is searching for disease, the pulleys are dragging back dislocated limbs – Nature herself is working out the primal curse which doomed the tenderest of her creatures to the sharpest of her trials, but the fierce extremity of suffering has been steeped in the waters of forgetfulness, and the deepest furrow in the knotted brow of agony has been smoothed forever'.

by recipe rather than reason, as our understanding of molecular structure was still in its infancy. The first therapeutic drug to come from synthetic chemistry was amyl nitrite, prepared in 1859 by Guthrie and introduced, on the basis of its vasodilator activity, for treating angina by Brunton in 1864 – the first example of a drug born in a recognizably 'modern' way, through the application of synthetic chemistry, physiology and clinical medicine. This was a landmark indeed, for it was nearly 40 years before synthetic chemistry made any further significant contribution to therapeutics, and not until well into the 20th century that physiological and pharmacological knowledge began to be applied to the invention of new drugs.

It was during the latter half of the 19th century that the foundations of synthetic organic chemistry were laid, the impetus coming from work on aniline, a copious byproduct of the coal-tar industry. An English chemist, Perkin, who in 1856 succeeded in preparing from aniline a vivid purple compound, *mauvein*, laid the foundations. This was actually a chemical accident, as Perkin's aim had been to synthesize quinine. Nevertheless, the discovery gave birth to the synthetic dyestuffs industry, which played a major part in establishing the commercial potential of synthetic organic chemistry – a technology which later became a linchpin of the evolving pharmaceutical industry. A systematic approach to organic synthesis went hand in hand with improved understanding of chemical structure. Crucial steps were the establishment of the rules of chemical equivalence (valency), and the elucidation of the structure of benzene by Von Kekulé in 1865. The first representation of a structural formula depicting the bonds between atoms in two dimensions, based on valency rules, also appeared in 1865[4].

The reason why Perkin had sought to synthesize quinine was that the drug, prepared from *Cinchona* bark, was much in demand for the treatment of malaria, one of whose effects is to cause high fever. So quinine was (wrongly, as it turned out) designated as an antipyretic drug, and used to treat fevers of all kinds. Because quinine itself could not be synthesized, fragments of the molecule were made instead. These included antipyrine, phenacetin and various others, which were introduced with great success in the 1880s and 1890s, the first drugs to be 'designed' on chemical principles[5].

The apothecaries' trade

Despite the lack of efficacy of the pharmaceutical preparations that were available in the 19th century, the apothecaries' trade flourished; then, as now, physicians felt themselves obliged to issue prescriptions to satisfy the expectations of their patients for some token of remedial intent. Early in the 19th century, when many small apothecary businesses existed to satisfy the demand on a local basis, a few enterprising chemists undertook the task of isolating the active substances from these plant extracts. This was a bold and inspired leap, and one that attracted a good deal of ridicule. Although the old idea of 'signatures', which held that plants owed their medicinal properties to their biological characteristics[6], was falling into disrepute, few were willing to accept that individual chemical substances could be responsible for the effects these plants produced, such as emesis, narcosis, purgation or fever. The trend began with Friedrich Sertürner, a junior apothecary in Westphalia, who in 1805 isolated and purified morphine, barely surviving a test of its potency on himself. This was the first 'alkaloid', so named because of its ability to neutralize acids and form salts. This discovery led to the isolation of several more plant alkaloids, including emetine, strychnine, caffeine and quinine, mainly by two remarkably prolific chemists, Caventou and Pelletier, working in Paris in the period 1810–1825. The recognition that medicinal plants owed their properties to their individual chemical constituents, rather than to some intangible property associated with their living nature, marks a critical point in the history of the pharmaceutical industry. It can be seen as the point of origin of two of the three strands from which the industry grew – namely the beginnings of the 'industrialization' of the apothecaries' trade, and the emergence of the science of pharmacology. And by revealing the chemical nature of medicinal preparations, it hinted at the future possibility of making medicines artificially. Even though, at that time, synthetic organic chemistry was barely out of its cradle, these discoveries provided the impetus that later caused the chemical industry to turn, at a very early stage in its history, to making drugs.

The first local apothecary business to move into large-scale production and marketing of pharmaceuticals was the old-established Darmstadt firm Merck,

[4]Its author, the Edinburgh chemist Alexander Crum Brown, was also a pioneer of pharmacology, and was the first person to use a chemical reaction – quaternization of amines – to modify naturally occurring substances such as strychnine and morphine. With Thomas Fraser, in 1868, he found that this drastically altered their pharmacological properties, changing strychnine, for example, from a convulsant to a paralysing agent. Although they knew neither the structures of these molecules nor the mechanisms by which they acted, theirs was the first systematic study of structure–activity relationships.

[5]These drugs belong pharmacologically to the class of non-steroidal anti-inflammatories (NSAIDs), the most important of which is aspirin (acetylsalicylic acid). Ironically, aspirin itself had been synthesized many years earlier, in 1855, with no pharmacological purpose in mind. Aspirin was not developed commercially until 1899, subsequently generating huge revenues for Bayer, the company responsible.

[6]According to this principle, pulmonaria (lungwort) was used to treat respiratory disorders because its leaves resembled lungs, saffron to treat jaundice, and so on.

founded in 1668. This development, in 1827, was stimulated by the advances in purification of natural products. Merck was closely followed in this astute business move by other German- and Swiss-based apothecary businesses, giving rise to some which later also became giant pharmaceutical companies, such as Schering and Boehringer. The American pharmaceutical industry emerged in the middle of the 19th century. Squibb began in 1858, with ether as its main product. Soon after came Parke Davis (1866) and Eli Lilly (1876); both had a broader franchise as manufacturing chemists. In the 1890s Parke Davis became the world's largest pharmaceutical company, one of whose early successes was to purify crystalline adrenaline from adrenal glands and sell it in ampoules for injection. The US scientific community contested the adoption of the word 'adrenaline' as a trade name, but industry won the day and the scientists were forced to call the hormone 'epinephrine'.

The move into pharmaceuticals was also followed by several chemical companies such as Bayer, Hoechst, Agfa, Sandoz, Geigy and others, which began, not as apothecaries, but as dyestuffs manufacturers. The dyestuffs industry at that time was also based largely on plant products, which had to be refined, and were sold in relatively small quantities, so the commercial parallels with the pharmaceutical industry were plain. Dye factories, for obvious reasons, were usually located close to large rivers, a fact that accounts for the present-day location of many large pharmaceutical companies in Europe. As we shall see, the link with the dyestuffs industry later came to have much more profound implications for drug discovery.

From about 1870 onwards – following the crucial discovery by Kekulé of the structure of benzene – the dyestuffs industry turned increasingly to synthetic chemistry as a source of new compounds, starting with aniline-based dyes. A glance through any modern pharmacopoeia will show the overwhelming preponderance of synthetic aromatic compounds, based on the benzene ring structure, among the list of useful drugs. Understanding the nature of aromaticity was critical. Though we might be able to dispense with the benzene ring in some fields of applied chemistry, such as fuels, lubricants, plastics or detergents, its exclusion would leave the pharmacopoeia bankrupt. Many of these dyestuffs companies saw the potential of the medicines business from 1880 onwards, and moved into the area hitherto occupied by the apothecaries. The result was the first wave of companies ready to apply chemical technology to the production of medicines. Many of these founder companies remained in business for years. It was only recently, when their cannibalistic urges took over in the race to become large, that mergers and take-overs caused many names to disappear.

Thus the beginnings of a recognizable pharmaceutical industry date from about 1860–1880, its origins being in the apothecaries and medical supplies trades on the one hand, and the dyestuffs industry on the other. In those early days, however, they had rather few products to sell; these were mainly inorganic compounds of varying degrees of toxicity, and others best described as concoctions. Holmes (see above) dismissed the pharmacopoeia in 1860 as worse than useless.

To turn this ambitious new industry into a source of human benefit, rather than just corporate profit, required two things. First, it had to embrace the principles of biomedicine, and in particular pharmacology, which provided a basis for understanding how disease and drugs, respectively, affect the function of living organisms. Second, it had to embrace the principles of chemistry, going beyond the descriptors of colour, crystallinity, taste, volatility etc. towards an understanding of the structure and properties of molecules, and how to make them in the laboratory. As we have seen, both of these fields had made tremendous progress towards the end of the 19th century, so at the start of the 20th century the time was right for the industry to seize its chance. Nevertheless, several decades passed before the inventions coming from the industry began to make a major impact on the treatment of disease.

The industry enters the 20th century

By the end of the 19th century various synthetic drugs had been made and tested, including the 'antipyretics' (see above) and also various central nervous system depressants. Chemical developments based on chloroform had produced chloral hydrate, the first non-volatile CNS depressant, which was in clinical use for many years as a hypnotic drug. Independently, various compounds based on urea were found to act similarly, and von Mering followed this lead to produce the first barbiturate, *barbitone* (since renamed barbital), which was introduced in 1903 by Bayer and gained widespread clinical use as a hypnotic, tranquillizer and antiepileptic drug – the first blockbuster. Almost simultaneously, Einthorn in Munich synthesized *procaine*, the first synthetic local anaesthetic drug, which followed the naturally occurring alkaloid cocaine. The local anaesthetic action of cocaine on the eye was discovered by Sigmund Freud and his ophthalmologist colleague Koeller in the late 19th century, and was heralded as a major advance for ophthalmic surgery. After several chemists had tried, with limited success, to make synthetic compounds with the same actions, procaine was finally produced and introduced commercially in 1905 by Hoechst. Barbitone and procaine were triumphs for chemical ingenuity, but owed little or nothing to physiology, or indeed pharmacology. The physiological site or sites of action of barbiturates remain unclear to this day, and their

mechanism of action at the molecular level was unknown until the 1980s.

From this stage, where chemistry began to make an impact on drug discovery, up to the last quarter of the 20th century, when molecular biology began to emerge as a dominant technology, we can discern three main routes by which new drugs were discovered, namely chemistry-driven approaches, target-directed approaches, and accidental clinical discoveries. In many of the most successful case histories, graphically described by Weatherall (1990), the three were closely interwoven. The remarkable family of diverse and important drugs that came from the original sulfonamide, lead, described below, exemplifies this pattern very well.

Chemistry-driven drug discovery

Synthetic chemistry

The pattern of drug discovery driven by synthetic chemistry – with biology often struggling to keep up – became the established model in the early part of the 20th century, and prevailed for at least 50 years. The balance of research in the pharmaceutical industry up to the 1970s placed chemistry clearly as the key discipline in drug discovery, the task of biologists being mainly to devise and perform assays capable of revealing possible useful therapeutic activity among the many anonymous white powders that arrived for testing. Research management in the industry was largely in the hands of chemists. This strategy produced many successes, including benzodiazepine tranquillizers, several antiepileptic drugs, antihypertensive drugs, antidepressants and antipsychotic drugs. The surviving practice of classifying many drugs on the basis of their chemical structure (e.g. phenothiazines, benzodiazepines, thiazides etc.) rather than on the more logical basis of their site or mode of action stems from this era. The development of antiepileptic drugs exemplifies this approach well. Following the success of barbital (see above) several related compounds were made, including the phenyl derivative *phenobarbital*, first made in 1911. This proved to be an effective hypnotic (i.e. sleep-inducing) drug, helpful in allowing peaceful nights in a ward full of restive patients. By chance, it was found by a German doctor also to reduce the frequency of seizures when tested in epileptic patients – an example of clinical serendipity (see below), and it became widely used for this purpose, being much more effective in this regard than barbital itself. About 20 years later, Putnam, working in Boston, developed an animal model whereby epilepsy-like seizures could be induced in mice by electrical stimulation of the brain via extracranial electrodes. This simple model allowed hundreds of compounds to be tested for potential antiepileptic activity. *Phenytoin* was an early success of this programme, and several more compounds followed, as chemists from several companies embarked on synthetic programmes. None of this relied at all on an understanding of the mechanism of action of these compounds – which is still controversial; all that was needed were teams of green-fingered chemists, and a robust assay that fairly predicted efficacy in the clinic.

Natural product chemistry

We have mentioned the early days of pharmacology, with its focus on plant-derived materials, such as *atropine, tubocurarine, strychnine, digitalis* and *ergot alkaloids*, which were almost the only drugs that existed until well into the 20th century. Despite the rise of synthetic chemistry, natural products remain a significant source of new drugs, particularly in the field of chemotherapy, but also in other applications. Following the discovery of *penicillin* by Fleming in 1929 – described by Mann (1984) as 'the most important medical discovery of all time' – and its development as an antibiotic for clinical use by Chain and Florey in 1938, an intense search was undertaken for antibacterial compounds produced by fungi and other microorganisms, which yielded many useful antibiotics, including *chloramphenicol* (1947), *tetracyclines* (1948), *streptomycin* (1949) and others. The same fungal source that yielded streptomycin also produced *actinomycin D*, used in cancer chemotherapy. Higher plants have continued to yield useful drugs, including *vincristine* and *vinblastine* (1958), and *paclitaxel* (taxol, 1971).

Outside the field of chemotherapy, successful drugs derived from natural products include *ciclosporin* (1972) and *tacrolimus* (1993), both of which come from fungi and are used to prevent transplant rejection. Soon after came *mevastatin* (1976), another fungal metabolite, which was the first of the 'statin' series of cholesterol-lowering drugs which act by inhibiting the enzyme HMG CoA reductase.

Overall, the pharmaceutical industry continues to have something of a love–hate relationship with natural products. They often have weird and wonderful structures that cause hardened chemists to turn pale; they are often near-impossible to synthesize, troublesome to produce from natural sources, and 'optimizing' such molecules to make them suitable for therapeutic use is akin to remodelling Westminster Abbey to improve its acoustics. But the fact remains that Nature unexpectedly provides some of our most useful drugs, and most of its potential remains untapped.

Target-directed drug discovery

Although chemistry was the pre-eminent discipline in drug discovery until the 1970s, the seeds of the biological revolution had long since been sown, and within the chemistry-led culture of the pharmaceutical industry these developments began to bear fruit in certain areas

This happened most notably in the field of chemotherapy, where Ehrlich played such an important role as the first 'modernist' who defined the principles of drug specificity in terms of a specific interaction between the drug molecule and a target molecule – the 'receptive substance' – in the organism, an idea summarized in his famous Latin catchphrase *Corpora non agunt nisi fixata.* Although we now take it for granted that the chemical nature of the target molecule, as well as that of the drug molecule, determines what effects a drug will produce, nobody before Ehrlich had envisaged drug action in this way[7]. By linking chemistry and biology, Ehrlich effectively set the stage for drug discovery in the modern style. But despite Ehrlich's seminal role in the evolution of the pharmaceutical industry, discoveries in his favourite field of endeavour, chemotherapy, remained for many years empirical rather than target directed[8].

The fact is that Ehrlich's preoccupation with the binding of chemical dyes, as exemplified by biological stains, for specific constituents of cells and tissues, turned out to be misplaced, and not applicable to the problem of achieving selective toxicity. Although he soon came to realize that the dye-binding moieties of cells were not equivalent to the supposed drug-binding moieties, neither he nor anyone else succeeded in identifying the latter and using them as defined targets for new compounds. The history of successes in the field of chemotherapy prior to the antibiotic era, some of which are listed in Table 1.1, actually represents a series of striking achievements in synthetic chemistry, coupled to the development of assay systems in animals, according to the chemistry-led model that we have already discussed. The popular image of 'magic bullets' – a phrase famously coined by Ehrlich – designed to home in, like cruise missiles, on defined targets is actually a misleading one in the context of the early days of chemotherapy, but there is no doubt that Ehrlich's thinking prepared the ground for the steady advance of target-directed approaches to drug discovery, a trend that, from the

Table 1.1
Examples of drugs from different sources: natural products, synthetic chemistry and biopharmaceuticals

Natural products	Synthetic chemistry	Biopharmaceuticals produced by recombinant DNA technology
Antibiotics (penicillin, streptomycin, tetracyclines, cephalosporins etc.)	Early successes include:	Human insulin (the first biotech product, registered 1982)
Anticancer drugs (doxorubicin, bleomycin, actinomycin, vincristine, vinblastine, taxol etc.)	Antiepileptic drugs	Human growth hormone
Atropine, hyoscine	Antihypertensive drugs	α-interferon, γ-interferon
Ciclosporin	Antimetabolites	Hepatitis B vaccine
Cocaine	Barbiturates	Tissue plasminogen activator (t-PA)
Colchicine	Bronchodilators	Hirudin
Digitalis (digoxin)	Diuretics	Blood clotting factors
Ephedrine	Local anaesthetics	Erythropoietin
Heparin	Sulfonamides	G-CSF, GM-CSF
Human growth hormone*	*[Since c.1950, synthetic chemistry has accounted for the great majority of new drugs]*	
Insulin (porcine, bovine)*		
Opium alkaloids (morphine, papaverine)		
Physostigmine		
Rauwolfia alkaloids (reserpine)		
Statins		
Streptokinase		
Tubocurarine		
Vaccines		

*Now largely or entirely replaced by material prepared by recombinant DNA technology.

[7]Others came close at around the same time, particularly the British physiologist J. N. Langley (1905), who interpreted the neuromuscular blocking effect of 'curari' in terms of its interaction with a specific 'receptive substance' at the junction between the nerve terminal and the muscle fibre. This was many years before chemical transmission at this junction was discovered. Langley's student, A. V. Hill (1909), first derived the equations based on the Law of Mass Action, which describe how binding varies with drug concentration. Hill's quantitative theory later formed the basis of 'receptor theory', elaborated by pharmacologists from A. J. Clark (1926) onwards. Although this quantitative approach underlies much of our current thinking about drug–receptor interactions, it was Ehrlich's more intuitive approach that played the major part in shaping drug discovery in the early days.

[8]Even now, important new chemotherapeutic drugs, such as the taxanes, continue to emerge through a combination of a chance biological discovery and high-level chemistry.

1950s onwards, steadily shifted the industry's focus from chemistry to biology (Maxwell and Eckhardt, 1990; Lednicer, 1993). A few selected case histories exemplify this general trend.

The sulfonamide story

Ehrlich's major triumph was the discovery in 1910 of *Salvarsan* (Compound 606), the first compound to treat syphilis effectively, which remained in use for 40 years. Still, bacterial infections, such as pneumonia and wound infections, proved resistant to chemical treatments for many years, despite strenuous effort on the part of the pharmaceutical industry. In 1927, IG Farbenindustrie, which had a long-standing interest in discovering antimicrobial drugs, appointed Gerhard Domagk to direct their research. Among the various leads that he followed was a series of azo dyes, included among which were some sulfonamide derivatives (a modification introduced earlier into dyestuffs to improve their affinity for certain fibres). These were much more effective in animals, and less toxic, than anything that had gone before, and *Prontosil* – a dark-red azo dye – was introduced in 1935. In the same year, it saved the life of Domagk's daughter, who developed septicaemia after a needle prick. It was soon discovered that the azo linkage in the Prontosil molecule was rapidly cleaved in the body, yielding the colourless compound *sulfanilamide*, which accounted for the antibacterial effect of Prontosil[9].

With chemistry still firmly in the driving seat, and little concern about mechanisms or targets, many sulfonamides were made in the next few years and they dramatically improved the prognosis of patients suffering from infectious diseases.

The mechanistic light began to dawn in 1940, when D. D. Woods, a microbiologist in Oxford, discovered that the antibacterial effect of sulfonamides was antagonized by *p*-aminobenzoic acid (PABA), a closely related compound and a precursor in the biosynthesis of folic acid (Figure 1.1). Bacteria, but not eukaryotic cells, have to synthesize their own folic acid to support DNA synthesis. Woods deduced that sulfonamides compete with PABA for a target enzyme, now known to be dihydropteroate synthase, and thus prevent folic acid synthesis.

[9]Sulfanilamide, a known compound, could not be patented, and so many companies soon began to make and sell it in various formulations. In 1937 about 80 people who took the drug died as a result of solvent-induced liver and kidney damage. It was this accident that led to the US Food and Drug Act, with the Food & Drug Administration (FDA) to oversee it (see Chapter 20).

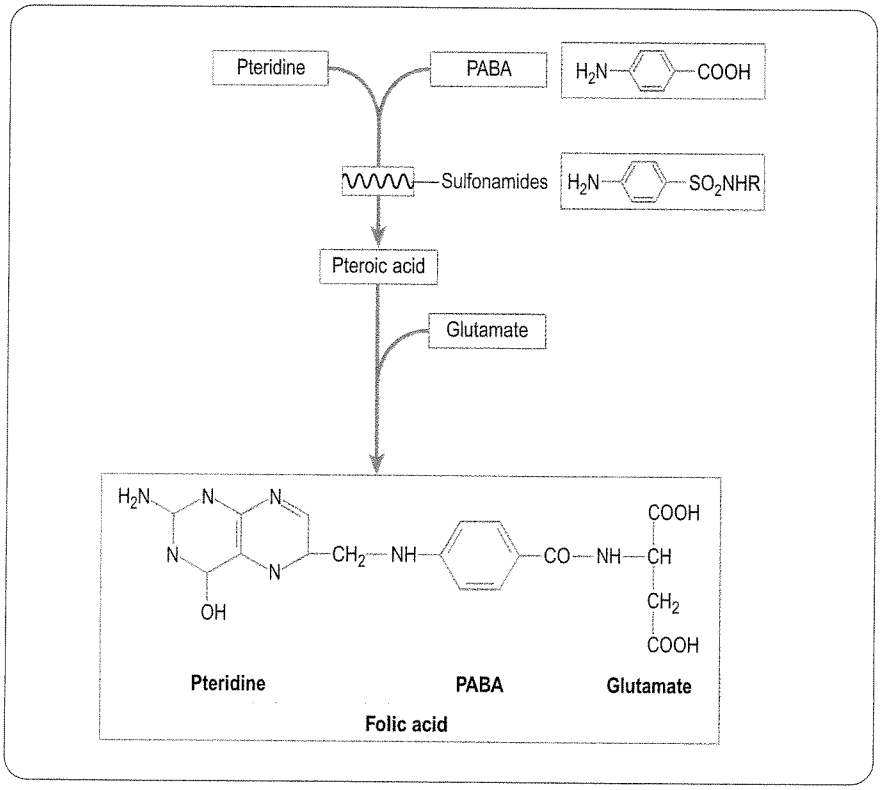


Fig. 1.1
Folic acid synthesis and PABA.

The discovery of sulfonamides and the elucidation of their mechanism of action had great repercussions, scientifically as well as clinically. In the drug discovery field, it set off two major lines of inquiry. First, the sulfonamide structure proved to be a rich source of molecules with many different, and useful, pharmacological properties – an exuberant vindication of the chemistry-led approach to drug discovery. Second, attacking the folic acid pathway proved to be a highly successful strategy for producing therapeutically useful drugs – a powerful boost for the 'targeteers', who were still few in number at this time.

The chemical dynasty originating with sulfanilamide is shown in Figure 1.2. An early spin-off came from the clinical observation that some sulfonamides produced an alkaline diuresis, associated with an increased excretion of sodium bicarbonate in the urine. Carbonic anhydrase, an enzyme which catalyses the interconversion of carbon dioxide and carbonic acid, was described in 1940, and its role in renal bicarbonate excretion was discovered a few years later, which prompted the finding that some, but not all, sulfonamides inhibit this enzyme. Modification of the sulfonamide structure led eventually to *acetazolamide* the first commercially available carbonic anhydrase inhibitor, as a diuretic in 1952. Following the diuretic trail led in turn to *chlorothiazide* (1957), the first of the thiazide diuretics, which, though devoid of carbonic anhydrase inhibitory activity, was much more effective than acetazolamide in increasing sodium excretion, and much safer than the earlier mercurial diuretics, which had until then been the best drugs available for treating oedema associated with heart failure and other conditions. Still further modifications led first to *frusemide* (1962) and later to *bumetanide* (1984), which were even more effective than the thiazides in producing a rapid diuresis – 'torrential' being the adjective applied by clinicians with vivid imaginations. Other modifications of the thiazide structures led to the accidental but important discovery of a series of hypotensive vasodilator drugs, such as *hydralazine* and *diazoxide*. In yet another development, *carbutamide*, one of the sulfonamides synthesized by Boehringer in 1954 as part of an antibacterial drug programme, was found accidentally to cause hypoglycaemia. This drug was the first of the sulfonylurea series, from which many further derivatives, such as *tolbutamide* and *glibenclamide*, were produced and used successfully to treat diabetes. All of these products of the sulfonamide dynasty are widely used today. Their chemical relationship to sulfonamides is clear, though none of them has antibacterial activity. Their biochemical targets in smooth muscle, the kidney, the pancreas and

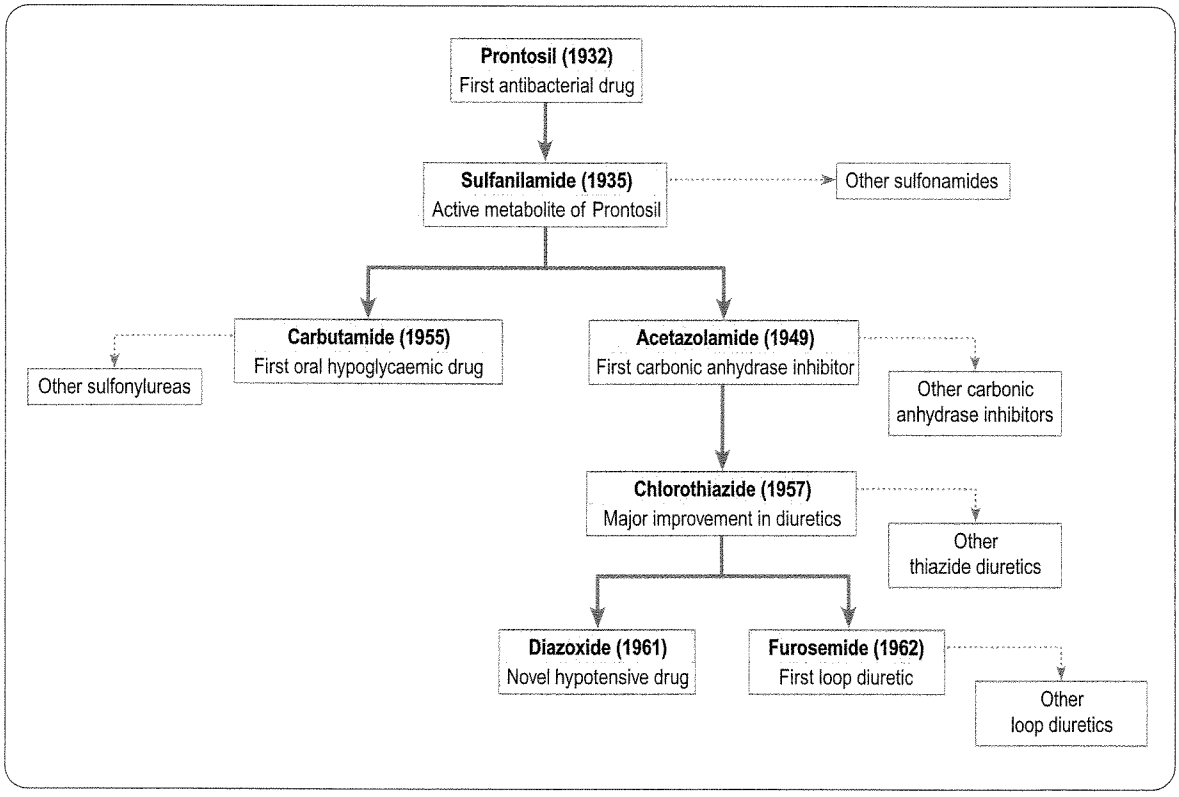


Fig. 1.2
Sulfonamide dynasty.

elsewhere, are all different. Chemistry, not biology, was the guiding principle in their discovery and synthesis.

Target-directed approaches to drug design have played a much more significant role in areas other than antibacterial chemotherapy, the approaches being made possible by advances on two important fronts, separated, as it happens, by the Atlantic Ocean. In the United States, the antimetabolite principle, based on interfering with defined metabolic pathways, proved to be highly successful, due largely to the efforts of George Hitchings and Gertrude Elion at Burroughs Wellcome. In Europe, drug discovery took its lead more from physiology than biochemistry, and sprung from advances in knowledge of chemical transmitters and their receptors. The names of Henry Dale and James Black deserve special mention here.

Hitchings and Elion and the antimetabolite principle

George Hitchings and Gertrude Elion came together in 1944 in the biochemistry department of Burroughs Wellcome in Tuckahoe, New York. Their biochemical interest lay in the synthesis of folic acid, based on the importance of this pathway for the action of sulfonamides, and they set about synthesizing potential 'antimetabolites' of purines and pyrimidines as chemotherapeutic agents. At the time, it was known that this pathway was important for DNA synthesis, but the role of DNA in cell function was uncertain. It turned out to be an inspired choice, and theirs was one of the first drug discovery programmes to focus on a biochemical pathway, rather than on a series of chemical compounds[10].

Starting from a series of purine and pyrimidine analogues, which had antibacterial activity, Hitchings and Rlion identified a key enzyme in the folic acid pathway, namely dihydrofolate reductase, which was necessary for DNA synthesis and was inhibited by many of their antibacterial pyrmidine analogues. Because all cells, not just bacteria, use this reaction to make DNA, they wondered why the drugs showed selectivity in their ability to block cell division, and found that the enzyme showed considerable species variation in its susceptibility to inhibitors. This led them to seek inhibitors that would selectively attack bacteria, protozoa and human neoplasms, which they achieved with great success. Drugs to emerge from this programme included the antituberculosis drug *pyrimethamine*, the antibacterial *trimethoprim*, and the anticancer drug *6-mercaptopurine*, as well as *azathioprine*, an immunosuppressant drug that was later widely used to prevent transplant rejection. Another spin-off from the work of Hitchings and Elion was *allopurinol*, an inhibitor of purine synthesis that is used in the treatment of gout. Later on, Elion – her enthusiasm for purines and pyrimidines undiminished – led the research group which in 1977 discovered one of the first effective antiviral drugs, *aciclovir*, an inhibitor of DNA polymerase, and later the first antiretroviral drug, *zidovudine* (AZT), which inhibits reverse transcriptase. Hitchings and Elion, by focusing on the metabolic pathways involved in DNA synthesis, invented an extraordinary range of valuable therapeutic drugs, an achievement unsurpassed in the history of drug discovery.

[10]They were not alone. In the 1940s, a group at Lederle laboratories made *aminopterin* and *methotrexate*, folic acid antagonists which proved effective in treating leukaemia.

James Black and receptor-targeted drugs

As already mentioned, the concept of 'receptors' as recognition sites for hormones and other physiological mediators came from J. N. Langley's analysis of the mechanism of action of 'curari'. Henry Dale's work on the distinct 'muscarinic' and 'nicotinic' actions of acetylcholine also pointed to the existence of two distinct types of cholinergic receptor, though Dale himself dismissed the receptor concept as an abstract and unhelpful cloak for ignorance. During the 1920s and 1930s, major discoveries highlighting the role of chemical mediators were made by physiologists, including the discovery of insulin, adrenal steroids and several neurotransmitters, and the realization that chemical signalling was crucial for normal function focused attention on the receptor mechanisms needed to decode these signals. Pharmacologists, particularly A. J. Clark, J. H. Gaddum and H. O. Schild, applied the Law of Mass Action to put ligand–receptor interactions on a quantitative basis. Schild's studies on drug antagonism in particular, which allowed the binding affinity of competitive antagonists to receptors to be estimated from pharmacological experiments, were an important step forward, which provided the first – and still widely used – quantitative basis for classifying drug receptors. On the basis of such quantitative principles, R. P. Ahlquist in 1948 proposed the existence of two distinct classes of adrenergic receptor, α and β, which accounted for the varied effects of epinephrine and norepinephrine on the cardiovascular system. This discovery inspired James Black, working in the research laboratories of Imperial Chemical Industries in the UK, to seek antagonists that would act selectively on β-adrenoceptors and thus block the effects of epinephrine on the heart, which were thought to be

harmful in patients with coronary disease. His chemical starting point was *dichloroisoprenaline*, which had been found by Slater in 1957 to block the relaxant effects of epinephrine on bronchial smooth muscle – of no interest to Slater at the time, as he was looking for compounds with the opposite effect. The result of Black's efforts was the first β-adrenoceptor blocking drug, *pronethalol* (1960), which had the desired effects in humans but was toxic. It was quickly followed by *propranolol* (registered in 1964[11]) – one of the earliest blockbusters, which found many important applications in cardiovascular medicine. This was the first time that a receptor, identified pharmacologically, had been deliberately targeted in a drug discovery project.

Black, after moving to Smith Kline and French, went on from this success to look for novel histamine antagonists that would block the stimulatory effect of histamine on gastric acid secretion, this effect being resistant to the then-known antihistamine drugs. The result of this project, in which the chemistry effort proved much tougher than the β-adrenoceptor antagonist project, was the first H_2-receptor antagonist, *burimamide* (1972). This compound was a major clinical advance, being the first effective drug for treating peptic ulcers, but (like pronethalol) was quickly withdrawn because of toxicity, to be replaced by *cimetidine* (1976). In 1988 Black, along with Hitchings and Elion, was awarded the Nobel Prize.

Black's work effectively opened up the field of receptor pharmacology as an approach to drug discovery, and the pharmaceutical industry quickly moved in to follow his example. Lookalike β-adrenoceptor antagonists and H_2-receptor antagonists followed rapidly during the 1970s and 1980s, and many other receptors were set up as targets for potential therapeutic agents, based on essentially the same approach – though with updated technology – that Black and his colleagues had introduced.

Drews (2000) estimated that of 483 identified targets on which the current set of approved drugs act, G-protein-coupled receptors – of which β-adrenoceptors and H_2-receptors are typical examples – account for 45%. Many other successful drugs have resulted from target-directed projects along the lines pioneered by Black and his colleagues. In recent years, of course, receptors have changed from being essentially figments in an operational scheme devised by pharmacologists to explain their findings, to being concrete molecular entities that can be labelled, isolated as proteins, cloned and expressed, just like many other proteins. As we shall see in later chapters, these advances have completely transformed the techniques employed in drug discovery research.

[11]Ironically, propranolol had been made in 1959 in the laboratories of Boehringer Ingelheim, as part of a different, chemistry-led project. Only when linked to its target could the clinical potential of propranolol be revealed – chemistry alone was not enough!

Accidental clinical discoveries

Another successful route to the discovery of new drugs has been through observations made in the clinic. Until drug discovery became an intentional activity, such serendipitous observations were the only source of knowledge. Withering's discovery in 1785 of the efficacy of digitalis in treating dropsy, and Wenkebach's discovery in 1914 of the antidysrhythmic effect of quinine, when he treated a patient with malaria who also happened to suffer from atrial tachycardia, are two of many examples where the clinical efficacy of plant-derived agents has been discovered by highly observant clinicians. More recently, clinical benefit of unexpected kinds has been discovered with synthetic compounds developed for other purposes. In 1937, for example, Bradley tried *amphetamine* as a means of alleviating the severe headache suffered by children after lumbar puncture (spinal tap), on the grounds that the drug's cardiovascular effects might prove beneficial. The headache was not alleviated, but Bradley noticed that the children became much less agitated. From this chance observation he went on to set up one of the first controlled clinical trials, which demonstrated unequivocally that amphetamine had a calming effect – quite unexpected for a drug known to have stimulant effects in other circumstances. From this developed the widespread use, validated by numerous controlled clinical trials, of amphetamine-like drugs, particularly *methylphenidate* (Ritalin) to treat attention deficit hyperactivity disorder (ADHD) in children. Other well-known examples include the discovery of the antipsychotic effects of phenothiazines by Laborit in 1949. Laborit was a naval surgeon, concerned that patients were dying from 'surgical shock' – circulatory collapse resulting in irreversible organ failure – after major operations. Thinking that histamine might be involved, he tested the antihistamine *promethazine* combined with autonomic blocking drugs to prevent this cardiovascular reaction. Although it was ineffective in treating shock, promethazine caused some sedation and Laborit tried some chemically related sedatives, notably *promazine*, which had little antihistamine activity. Patients treated with it fared better during surgery, but Laborit particularly noticed that they appeared much calmer postoperatively. He therefore persuaded his psychiatrist colleagues to test the drug on psychotic patients, tests that quickly revealed the drug's antipsychotic effects and led to the development of the antipsychotic *chlorpromazine*. In a sequel, other phenothiazine-like tricyclic compounds were tested for antipsychotic activity but were found accidentally to relieve the symptoms of depression. After

Bradley and Laborit, psychiatrists had become alert to looking for the unexpected.

Astute clinical observation has revealed many other unexpected therapeutic effects, for example the efficacy of various antidepressant and antiepileptic drugs in treating certain intractable pain states.

The regulatory process

In the mid-19th century restrictions on the sale of poisonous substances were imposed in the USA and UK, but it was not until the early 1900s that any system of 'prescription-only' medicines was introduced, requiring approval by a medical practitioner. Soon afterwards, restrictions began to be imposed on what 'cures' could be claimed in advertisements for pharmaceutical products and what information had to be given on the label; legislation evolved at a leisurely pace. Most of the concern was with controlling frankly poisonous or addictive substances or contaminants, not with the efficacy and possible harmful effects of new drugs.

In 1937, the use of diethylene glycol as a solvent for a sulfonamide preparation caused 107 deaths in the USA, and a year later the 1906 Food and Drugs Act was revised, requiring safety to be demonstrated before new products could be marketed, and also allowing federal inspection of manufacturing facilities. The requirement for proven efficacy, as well as safety, was added in the Kefauver–Harris amendment in 1962.

In Europe, preoccupied with the political events in the first half of the century, matters of drug safety and efficacy were a minor concern, and it was not until the mid-1960s, in the wake of the thalidomide disaster – a disaster averted in the USA by an assiduous officer, who used the provisions of the 1938 Food and Drugs Act to delay licensing approval – that the UK began to follow the US lead in regulatory laws. Until then, the ability of drugs to do harm – short of being frankly poisonous or addictive – was not really appreciated, most of the concern having been about contaminants. In 1959, when thalidomide was first put on the market by the German company Chemie Grünenthal, regulatory controls did not exist in Europe: it was up to the company to decide how much research was needed to satisfy itself that the drug was safe and effective. Grünenthal made a disastrously wrong judgement (see Sjöstrom and Nilsson, 1972, for a full account), which resulted in an estimated 10 000 cases of severe congenital malformation following the company's specific recommendation that the drug was suitable for use by pregnant women. This single event caused an urgent reappraisal, leading to the introduction of much tighter government controls.

In the UK, the Committee on the Safety of Drugs was established in 1963. For the first time, as in the USA, all new drugs (including new mixtures and formulations) had to be submitted for approval before clinical trials could begin, and before they could be marketed. Legally, companies could proceed even if the Committee did not approve, but very few chose to do so. This loophole was closed by the Medicines Act (1968), which made it illegal to proceed without approval. Initially, safety alone was the criterion for approval; in 1970, under the Medicines Act, evidence of efficacy was added to the criteria for approval. It was the realization that all drugs, not just poisons or contaminants, have the potential to cause harm that made it essential to seek proof of therapeutic efficacy to ensure that the net effect of a new drug was beneficial.

In the decade leading up to 1970, the main planks in the regulatory platform – evidence of safety, efficacy and chemical purity – were in place in most developed countries. Subsequently, the regulations have been adjusted in various minor ways, and adopted with local variations in most countries.

A progressive tightening of the restrictions on the licensing of new drugs continued for about two decades after the initial shock of thalidomide, as public awareness of the harmful effects of drugs became heightened, and the regulatory bodies did their best to respond to public demand for assurance that new drugs were 'completely safe'. The current state of licensing regulations is described in Chapter 20.

Concluding remarks

In this chapter we have followed the evolution of ideas and technologies that have led to the state of drug discovery research that existed circa 1970. The main threads, which came together, were:

- Clinical medicine, by far the oldest of the antecedents, which relied largely on herbal remedies right up to the 20th century;
- Pharmacy, which began with the apothecary trade in the 17th century, set up to serve the demand for herbal preparations;
- Organic chemistry, beginning in the mid-19th century and evolving into medicinal chemistry via dyestuffs;
- Pharmacology, also beginning in the mid-19th century and setting out to explain the effects of plant-derived pharmaceutical preparations in physiological terms.

Some of the major milestones are summarized in Table 1.2.

The pharmaceutical industry as big business began around the beginning of the 20th century, and for 60 or more years was dominated by chemistry. Gradually, from the middle of the century onwards, the balance

Table 1.2
Milestones in the development of the pharmaceutical industry

Year	Event	Notes
c. 1550 BC	Ebers papyrus	The earliest known compendium of medical remedies
1540	Diethyl ether synthesized	'Sweet oil of vitriol', arguably the first synthetic drug
1668	Merck (Darmstadt) founded	The apothecary business which later (1827) evolved into the first large-scale pharmaceutical company
1775	Nitrous oxide synthesized	
1785	Withering describes use of digitalis extract to treat 'dropsy'	The first demonstration of therapeutic efficacy
1803	Napoleon established examination and licensing scheme for doctors	
1763	Lind shows that lack of fruit causes scurvy	
1798	Jenner shows that vaccination prevents smallpox	
1799	Humphrey Davy demonstrates anaesthetic effect of nitrous oxide	
1806	Sertürner purifies morphine and shows it to be the active principle of opium	A seminal advance – the first evidence that herbal remedies contain active chemicals. Many other plant alkaloids isolated 1820–1840
1846	Morton administers ether as anaesthetic at Massachusets General Hospital	The first trial of surgical anaesthesia
1847	Chloroform administered to Queen Victoria to control labour pain	
1847	The first Pharmacological Institute set up by Bucheim	
mid-19C	The first pharmaceutical companies formed: Merck (1827) Squibb (1858) Hoechst (1862) Parke Davis (1866) Lilley (1876) Burroughs Wellcome (1880)	In many cases, pharmaceutical companies evolved from dyestuffs companies or apothecaries
1858	Virchow proposes cell theory	
1859	Amyl nitrite synthesized	
1865	Benzene structure elucidated (Kekule), and first use of structural fromulae to describe organic molecules	Essential foundations for the development of organic synthesis
1867	Brunton demonstrates use of amyl nitrite to relieve anginal pain	
1878	Pasteur proposes germ theory of disease	
1898	Heroin (diacetylmorphine) developed by Bayer	The first synthetic derivative of a natural product. Heroin was marketed as a safe and non-addictive alternative to morphine
1899	Aspirin developed by Bayer	
1903	Barbital developed by Bayer	
1904	Elliott demonstrates biological activity of extracts of adrenal glands, and proposes adrenaline release as a physiological mechanism	The first evidence for a chemical mediator – the basis of much modern pharmacology

Table 1.2
Milestones in the development of the pharmaceutical industry—cont'd

Year	Event	Notes
1910	Ehrlich discovers Salvarsan	The first antimicrobial drug, which revolutionized the treatment of syphilis
1912	Starling coins the term 'hormone'	
1921	MacLeod, Banting and Best discover insulin	Produced commercially by Lilly (1925)
1926	Loewi demonstrates release of 'Vagusstoff' from heart	The first clear evidence for chemical neurotransmission
1929	Fleming discovers penicillin	Penicillin was not used clinically until Chain and Florey solved production problems in 1938
1935	Domagk discovers sulfonamides	The first effective antibacterial drugs, and harbingers of the antimetabolite era
1936	Steroid hormones isolated by Upjohn company	
1937	Bovet discovers antihistamines	Subsequently led to discovery of antipsychotic drugs
1946	Gilman and Philips demonstrate anticancer effect of nitrogen mustards	The first anticancer drug
1951	Hitchings and Elion discover mercaptopurine	The first anticancer drug from the antimetabolite approach
1961	Hitchings and Schwartz discover azathioprine	Also from the antimetabolite programme, the first effective immunosuppressant able to prevent transplant rejection
1962	Black and his colleagues discover pronethalol	The first β-adrenoceptor antagonist to be used clinically
1972	Black and his colleagues discover burimamide	The first selective H_2 antagonist
1976	Genentech founded	The first biotech company, based on recombinant DNA technology
c. 1990	Introduction of combinatorial chemistry	

shifted towards pharmacology until, by the mid-1970s, chemistry and pharmacology were evenly balanced. This was a highly productive period for the industry, which saw many new drugs introduced, some of them truly novel but also many copycat drugs, which found an adequate market despite their lack of novelty. The maturation of the scientific and technological basis of the discovery process to its 1970s level coincided with the development of much more stringent regulatory controls, which also reached a degree of maturity at this time, and an acceptable balance seemed to be struck between creativity and restraint.

We ended our historical account in the mid-1970s, when drug discovery seemed to have found a fairly serene and successful equilibrium, and products and profits flowed at a healthy rate. Just around the corner, however, lay the arrival on the drug discovery scene of molecular biology and its commercial wing, the biotechnology industry, which over the next 20 years were to transform the process and diversify its products in a dramatic fashion (see Chapters 3 and 12). Starting in 1976, when the first biotechnology companies (Cetus and Genentech) were founded in the USA, there are now about 1300 such companies in the USA and another 900 in Europe, and the products of such enterprises account for a steadily rising proportion – currently about 25% – of new therapeutic agents registered. As well as contributing directly in terms of products, biotechnology is steadily and radically transforming the ways in which conventional drugs are discovered.

The last quarter of the century has been a turbulent period which has affected quite radically the scientific basis of the development of new medicines, as well as the commercial environment in which the industry operates. The changes that are occurring show no sign of slowing down, and it is too soon to judge which developments will prove genuinely successful in terms of drug discovery, and which will not. In later chapters we discuss in detail the present state of the art with respect to the science and technology of drug discovery. The major discernible trends are as follows:

- Genomics as an approach to identifying new drug targets (Chapters 6 and 7);
- Increasing use of informatics technologies to store and interpret data (Chapter 7);

- High-throughput screening of large compound libraries as a source of chemical leads (Chapter 8);
- Combinatorial chemistry as a means of efficiently and systematically synthesizing collections of related compounds (Chapter 9);
- Increased emphasis on 'drugability' – mainly centred on pharmacokinetics and toxicology – in the selection of lead compounds (Chapter 10);
- Increased use of transgenic animals as disease models for drug testing (Chapter 11);
- The growth of biopharmaceuticals (Chapter 12).

What effect are these changes having on the success of the industry in finding new drugs? Despite the difficulty of defining and measuring such success, the answer seems to be 'not much so far'. Productivity, measured by the flow of new drugs (Figure 1.3) seems, if anything, to have drifted downwards over the last 15 years, and very markedly so since the 1960s, when regulatory controls began to be tightened. This measure, of course, takes no account of whether new drugs have inherent novelty and represent a significant therapeutic advance, or are merely the result of one company copying another. There are reasons for thinking that new drugs are now more innovative than they used to be, but clear evidence for this is hard to find. One sign is the growth of biopharmaceuticals relative to conventional drugs (Figure 1.4; see Chapters 12 and 21); most biopharmaceuticals represent novel therapeutic strategies, and there is less scope for copycat projects, which still account for a substantial proportion of new synthetic drugs. Adding to the disquiet caused by the downward trend in Figure 1.3 is the fact that research and development expenditure has steadily increased over the same period, and that development times – from discovery of a new molecule to market – have remained at 10–12 years since 1982. Costs, times and success rates are discussed in more detail in Chapter 21.

Nobody really understands why the apparent drop in research productivity has occurred, but speculations abound. One factor may be the increasing regulatory hurdles, which mean that development takes longer and costs more than it used to, so that companies have become more selective in choosing which compounds to develop. Another factor may be the trend away from 'me-too' drugs (drugs that differ little, if at all, from those already in use, but which nevertheless provide the company with a profitable share of the market while providing little or no benefit to patients).

The hope is that the downward trend will be reversed as the benefit of new technologies to the drug discovery process works through the system, as long development times mean that technologies introduced since 1990 have not yet had time to make an impact on registrations.

In the remainder of this book, we describe drug discovery at the time of writing (2003) – a time when the molecular biology revolution is in full swing. In a few years' time our account will undoubtedly look as dated as the 1970s scenario seems to us today.

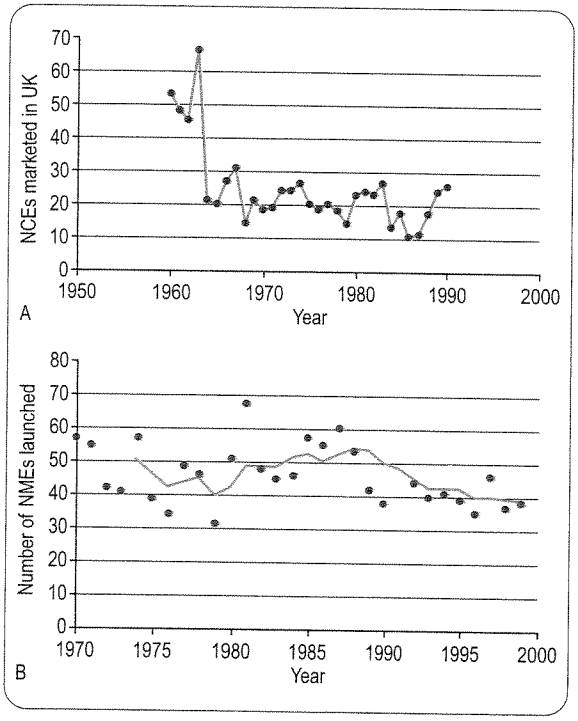


Fig. 1.3
Productivity – new drugs introduced 1970–2000. (a) Number of new chemical entities (NCEs) marketed in the UK 1960–1990, showing the dramatic effect of the thalidomide crisis in 1961. (Data from Griffin (1991) International Journal of Pharmacy 5: 206–209.) (b) Annual number of new molecular entities marketed in 20 countries worldwide 1970–1999. The line represents a 5-year moving average. (Data from Centre for Medicines Research Pharma R&D Compendium, 2000.)

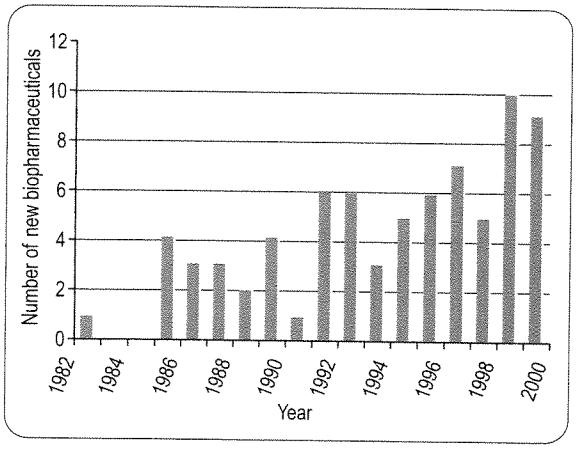


Fig. 1.4
Growth of biopharmaceuticals.

References

Drews J (2000) Drug discovery: a historical perspective. Science 287; 1960–1964.

Drews J (2003) In quest of tomorrow's medicines, 2nd edn. New York: Springer.

Drews J (2004) Paul Ehrlich: magister mundi. Nature Reviews Drug Discovery 3: 797–801.

Lednicer D (ed) (1993) Chronicles of drug discovery. Vol 3. Washington, DC: American Chemical Society.

Mann R D (1984) Modern drug use: an enquiry on historical principles. Lancaster: MTP Press.

Maxwell R A, Eckhardt S B (1990) Drug discovery: a casebook and analysis. Clifton, NJ: Humana Press.

Office of Health Economics (1983) Pharmaceutical innovation: recent trends, future prospects. Report No. 74. London, Office of Health Economics.

Porter R (1997) The greatest benefit to mankind: a medical history of humanity from antiquity to the present. London: Harper Collins.

Sjöstrom H, Nilsson R (1972) Thalidomide and the power of the drug companies. Harmondsworth: Penguin.

Sneader W (1985) Drug discovery: the evolution of modern medicines. Chichester: John Wiley.

Weatherall M (1990) In search of a cure: a history of pharmaceutical discovery. Oxford: Oxford University Press.

2 The nature of disease and the purpose of therapy

H P Rang

Introduction

In this book, we are concerned mainly with the drug discovery process itself, which those involved in research proudly regard as the mainspring of the pharmaceutical industry. In this chapter we consider the broader context of the human environment into which new drugs and medicinal products are launched, and where they must find their proper place. Most pharmaceutical companies place at the top of their basic mission statement a commitment to improve the public's health, to relieve the human burden of disease. Few would argue with the spirit of this commitment. Nevertheless, we need to look more closely at what it means, how disease is defined, what medical therapy aims to alter, and how – and by whom – the effects of therapy are judged and evaluated. Here we outline some of the basic principles underlying these broader issues.

Concepts of disease

The practice of medicine predates by thousands of years the science of medicine, and the application of 'therapeutic' procedures by professionals similarly predates any scientific understanding of how the human body works, or what happens when it goes wrong. As discussed in Chapter 1, the ancients defined disease not only in very different terms, but also on a quite different basis from what we would recognize today. The origin of disease and the measures needed to counter it were generally seen as manifestations of divine will and retribution, rather than of physical malfunction. The scientific revolution in medicine, which began in earnest during the 19th century and has been steadily accelerating since, has changed our concept of disease quite drastically, and continues to challenge it, raising new ethical problems and thorny discussions of principle. For the centuries of prescientific medicine, codes of practice based on honesty, integrity and professional relationships were quite sufficient: as therapeutic interventions were ineffective anyway, it mattered little to what situations they were applied. Now, quite suddenly, the language of disease has changed and interventions have become effective; not surprisingly, we have to revise our

ideas about what constitutes disease, and how medical intervention should be used. In this chapter, we will try to define the scope and purpose of therapeutics in the context of modern biology. In reality, however, those in the science-based drug discovery business have to recognize the strong atavistic leanings of many healthcare professions[1], whose roots go back much further than the age of science.

Therapeutic intervention, including the medical use of drugs, aims to prevent, cure or alleviate disease states. The question of exactly what we mean by disease, and how we distinguish disease from other kinds of human affliction and dysfunction, is of more than academic importance, because policy and practice with respect to healthcare provision depend on where we draw the line between what is an appropriate target for therapeutic intervention and what is not. The issue concerns not only doctors, who have to decide every day what kind of complaints warrant treatment, but all those involved in the healthcare business – including, of course, the pharmaceutical industry. Much has been written on the difficult question of how to define health and disease, and what demarcates a proper target for therapeutic intervention (Reznek, 1987; Caplan, 1993; Caplan et al., 2004); nevertheless, the waters remain distinctly murky.

One approach is to define what we mean by health, and to declare the attainment of health as the goal of all healthcare measures, including therapeutics.

What is health?

In everyday parlance we use the words 'health', 'fitness', 'wellbeing' on the one hand, and 'disease', 'illness', 'sickness', 'ill-health' etc. on the other, more or less interchangeably, but these words become slippery and evasive when we try to define them. The World Health Organization (WHO), for example, defines health as 'a state of complete physical, mental and social wellbeing and not merely the absence of sickness or infirmity'. On this basis, few humans could claim to possess health, although the majority may not be in the grip of obvious sickness or infirmity. Who is to say what constitutes 'complete physical, mental and social well-being' in a human being? Does physical well being imply an ability to run a marathon? Does a shy and self-effacing person lack social well-being?

We also find health defined in functional terms, less idealistically than in the WHO's formulation: '...health consists in our functioning in conformity with our natural design with respect to survival and reproduction, as determined by natural selection...' (Caplan, 1993). Here the implication is that evolution has brought us to an optimal – or at least an acceptable – compromise with our environment, with the corollary that healthcare measures should properly be directed at restoring this level of functionality in individuals who have lost some important element of it. This has a fashionably 'greenish' tinge, and seems more realistic than the WHO's chillingly utopian vision, but there are still difficulties in trying to use it as a guide to the proper application of therapeutics. Environments differ. A black-skinned person is at a disadvantage in sunless climates, where he may suffer from vitamin D deficiency, whereas a white-skinned person is liable to develop skin cancer in the tropics. The possession of a genetic abnormality of haemoglobin, known as sickle-cell trait, is advantageous in its heterozygous form in the tropics, as it confers resistance to malaria, whereas homozygous individuals suffer from a severe form of haemolytic anaemia (sickle-cell disease). Hyperactivity in children could have survival value in primitive societies, whereas in western countries it disrupts families and compromises education. Obsessionality and compulsive behaviour are quite normal in early motherhood, and may serve a good biological purpose, but in other walks of life can be a severe handicap, warranting medical treatment.

Health cannot therefore be regarded as a definable state – a fixed point on the map, representing a destination which all are seeking to reach. Rather, it seems to be a continuum, through which we can move in either direction, becoming more or less well adapted for survival in our particular environment. Although we could argue that the aim of healthcare measures is simply to improve our state of adaptation to our present environment, this is obviously too broad. Other factors than health – for example wealth, education, peace, and the avoidance of famine – are at least as important, but lie outside the domain of medicine. What actually demarcates the work of doctors and healthcare workers from that of other caring professionals – all of whom may contribute to health in different ways – is that the former focus on *disease*.

What is disease?

Consider the following definitions of disease:

- A condition which alters or interferes with the normal state of an organism and is usually characterized by the abnormal functioning of one or more of the host's systems, parts or organs (*Churchill's Medical Dictionary*, 1989).
- A morbid entity characterized usually by at least two of these criteria: recognized aetiologic agents,

[1]The upsurge of 'alternative' therapies, many of which owe nothing to science – and indeed reject the relevance of science to what its practitioners do – perhaps reflects an urge to return to the prescientific era of medical history.

identifiable groups of signs and symptoms, or consistent anatomical alterations (elsewhere, 'morbid' is defined as diseased or pathologic) (*Stedman's Medical Dictionary*, 1990).

We sense the difficulty that these thoughtful authorities found in pinning down the concept. The first definition emphasizes two aspects, namely *deviation from normality*, and *dysfunction*; the second emphasizes *aetiology* (i.e. causative factors) and *phenomenology* (signs, symptoms etc.), which is essentially the manifestation of dysfunction.

Deviation from normality does not define disease

The criterion of deviation from normality begs many questions. It implies that we know what the 'normal state' is, and can define what constitutes an alteration of it. It suggests that if our observations were searching enough, we could unfailingly distinguish disease from normality. But we know, for example, that the majority of 50-year-olds will have atherosclerotic lesions in their arteries, that some degree of osteoporosis is normal in postmenopausal women. These are not deviations from normality, nor do they in themselves cause dysfunction, and so they do not fall within these definitions of disease, yet both are seen as pathological and as legitimate – indeed important – targets for therapeutic intervention. Furthermore, as discussed below, deviations from normality are often beneficial and much prized.

Phenomenology and aetiology are important factors – the naturalistic view

Setting aside the normality criterion, the definitions quoted above are examples of the *naturalistic*, or observation-based, view of disease, defined by phenomenology and backed up in many cases by an understanding of aetiology; it is now generally agreed that this by itself is insufficient, for there is no *general* set of observable characteristics that distinguishes disease from health. Although individual diseases of course have their defining characteristics, which may be structural, biochemical or physiological, there is no common feature. Further, there are many conditions, particularly in psychiatry, but also in other branches of medicine, where such physical manifestations are absent, even though their existence as diseases is not questioned. Examples would include obsessive–compulsive disorder, schizophrenia, chronic fatigue syndrome, and low back pain. In such cases, of which there are many examples, the disease is defined by symptoms, of which only the patient is aware, or altered behaviour, of which he and those around him are aware: defining features at the physical, biochemical or physiological level are absent, or at least not yet recognized.

Harm and disvalue – the normative view

The shortcomings of the naturalistic view of disease, which is in principle value free, have led some authors to take the opposite view, to the extent of denying the relevance of any kind of objective criteria to the definition of disease. Crudely stated, this value-based (or *normative*) view holds that disease is simply any condition the individual or society finds disagreeable or harmful (i.e. *disvalues*). Taken to extremes by authors such as Szasz and Illich, this view denies the relevance of the physical manifestations of illness, and focuses instead on illness only as a manifestation of *social* intolerance or malfunction. Although few would go this far – and certainly modern biologists would not be among them – it is clear that value-laden judgements play a significant role in determining what we choose to view as disease. In the mid-19th century masturbation was regarded as a serious disease, to be treated if necessary by surgery, and this view persisted well into the 20th century. 'Drapetomania', defined as a disease of American slaves, was characterized by an obsessional desire for freedom. Homosexuality was seen as pathological, and determined attempts were made to treat it.

A definition of disease which tries to combine the concepts of biological malfunction and harm (or disvalue) was proposed by Caplan et al. (1981):

> 'States of affairs are called diseases when they are due to physiological or psychological processes that typically cause states of disability, pain or deformity, and are generally held to be below acceptable physiological or psychological norms'.

What is still lacking is any reference to aetiology, yet this can be important in recognizing disease, and indeed is increasingly so as we understand more about the underlying biological mechanisms. A patient who complains of feeling depressed may be reacting quite normally to a bereavement, or may come from a suicide-prone family, suggestive of an inherited tendency to depressive illness. The symptoms might be very similar, but the implications, based on aetiology, would be different.

In conclusion, disease proves extremely difficult to define (Scully, 2004). The closest we can get at present to an operational definition of disease rests on a combination of three factors: phenomenology, aetiology and disvalue, as summarized in Figure 2.1.

Labelling human afflictions as diseases (i.e. 'medicalizing' them) has various beneficial and adverse consequences, both for the affected individuals and for healthcare providers. It is of particular relevance to the pharmaceutical industry, which stands to benefit from the labelling of borderline conditions as diseases meriting therapeutic intervention. Strong criticism has been levelled at the industry for the way in which it uses its resources to promote the recognition of questionable

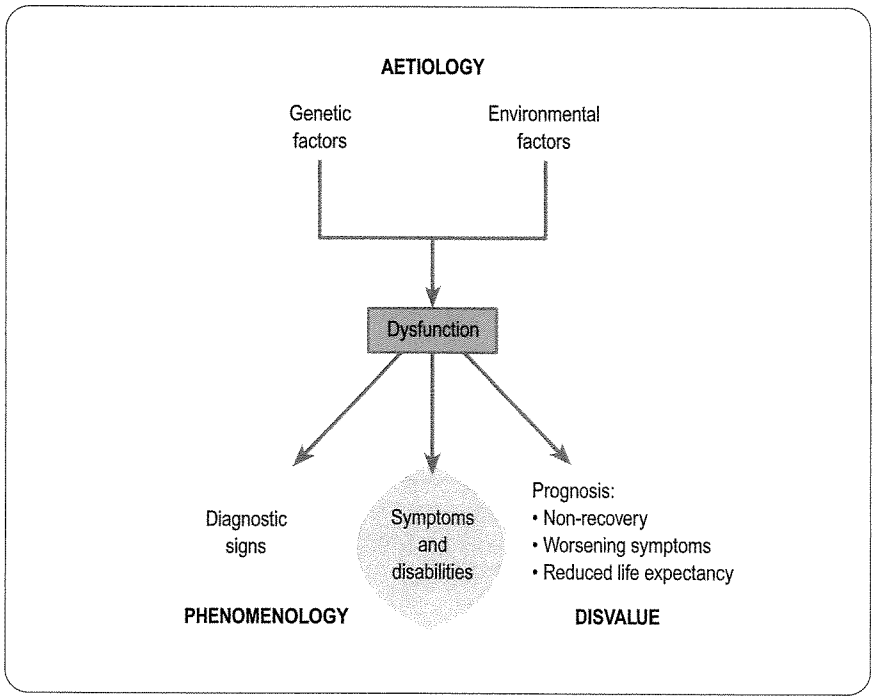


Fig. 2.1
Three components of disease.

disorders, such as female sexual dysfunction or social phobia, as diseases, and to elevate identified risk factors – asymptomatic in themselves but increasing the likelihood of disease occurring later – to the level of diseases in their own right. A recent polemic (Moynihan et al., 2004) starts with the sentence: 'there's a lot of money to be made from telling healthy people they're sick', and emphasizes the thin line that divides medical education from marketing.

The aims of therapeutics

Components of disvalue

The discussion so far leads us to the proposition that the proper aim of therapeutic intervention is to minimize the disvalue associated with disease. The concept of disvalue is therefore central, and we need to consider what comprises it. The disvalue experienced by a sick individual has two distinct components[2] (Figure 2.1), namely *present symptoms and disabilities* (collectively termed *morbidity*), and future *prognosis* (namely the likelihood of increasing morbidity, or premature death). An individual who is suffering no abnormal symptoms or disabilities, and whose prognosis is that of an average individual of the same age, we call 'healthy'. An individual with a bad cold or a sprained ankle has symptoms and disabilities, but probably has a normal prognosis. An individual with asymptomatic lung cancer or hypertension has no symptoms but a poor prognosis. Either case constitutes disease, and warrants therapeutic intervention. Very commonly, both components of disvalue are present and both need to be addressed with therapeutic measures – different measures may be needed to alleviate morbidity and to im-

[2]These concepts apply in a straightforward way to many real-life situations, but there are exceptions and difficulties. For example, in certain psychiatric disorders the patient's judgement of his or her state of morbidity is itself affected by the disease. Patients suffering from mania, paranoid delusions or severe depression may pursue an extremely disordered and self-destructive lifestyle, while denying that they are ill and resisting any intervention. In such cases, society often imposes its own judgement of the individual's morbidity, and may use legal instruments such as the Mental Health Act to apply therapeutic measures against the patient's will.

Vaccination represents another special case. Here, the disvalue being addressed is the theoretical risk that a healthy individual will later contract an infectious disease such as diphtheria or measles. This risk can be regarded as an adverse factor in the prognosis of a perfectly normal individual.

Similarly, a healthy person visiting the tropics will, if he is wise, take antimalarial drugs to avoid infection – in other words, to improve his prognosis.

prove prognosis. Of course, such measures need not be confined to physical and pharmacological approaches.

The proposition at the beginning of this section sets clear limits to the aims of therapeutic intervention, which encompass the great majority of non-controversial applications. Real life is, of course, not so simple, and in the next section we consider some of the important exceptions and controversies that healthcare professionals and policy-makers are increasingly having to confront.

Therapeutic intervention is not restricted to treatment or prevention of disease

The term 'lifestyle drugs' is a recent invention, but the concept of using drugs, and other types of intervention in a medical setting for purposes unrelated to the treatment of disease is by no means new.

Pregnancy is not by any definition a disease, nor are skin wrinkles, yet contraception, abortion and plastic surgery are well established practices in the medical domain. Why are we prepared to use drugs as contraceptives or abortifacients, but condemn using them to enhance sporting performance? The basic reason seems to be that we attach disvalue to unwanted pregnancy (i.e. we consider it harmful). We also attach disvalue to alternative means of avoiding unwanted pregnancy, such as sexual abstinence or using condoms. Other examples, however, such as cosmetic surgery to remove wrinkles or reshape breasts, seem to refute the disvalue principle: minor cosmetic imperfections are in no sense harmful, but society none the less concedes to the demand of individuals that medical technology should be deployed to enhance their beauty. In other cases, such as the use of sildenafil (Viagra) to improve male sexual performance, there is ambivalence about whether its use should be confined to those with evidence for erectile dysfunction (i.e. in whom disvalue exists) or whether it should also be used in normal men.

It is obvious that departures from normality can bring benefit as well as disadvantage. Individuals with above-average IQs, physical fitness, ball-game skills, artistic talents, physical beauty or charming personalities have an advantage in life. Is it, then, a proper role of the healthcare system to try to enhance these qualities in the average person? Our instinct says not, because the average person cannot be said to be diseased or suffering. There may be value in being a talented footballer, but there is no harm in not being one. Indeed, the value of the special talent lies precisely in the fact that most of us do not possess it. Nevertheless, a magical drug that would turn anyone into a brilliant footballer would certainly sell extremely well, at least until footballing skills became so commonplace that they no longer had any value[3].

Football skills may be a fanciful example; longevity is another matter. The 'normal' human lifespan varies enormously in different countries, and in the west it has increased dramatically during our own lifetime (Figure 2.2). Is lifespan prolongation a legitimate therapeutic aim? Our instinct – and certainly medical tradition – suggests that delaying premature death from disease is one of the most important functions of healthcare, but we are very ambivalent when it comes to prolonging life in the aged. Our ambivalence stems from the fact that the aged are often irremediably infirm, not merely chronologically old. In the future we may understand better why humans become infirm, and hence more vulnerable to the environmental and genetic circumstances that cause them to become ill and die. And beyond that we may discover how to retard or prevent aging, so that the 'normal' lifespan will be much prolonged. Opinions will differ as to whether this will be the ultimate triumph of medical science or the ultimate social disaster[4].

Conclusions

We have argued that that disease can best be defined in terms of three components, aetiology, phenomenology and disvalue, and that the element of disvalue is the

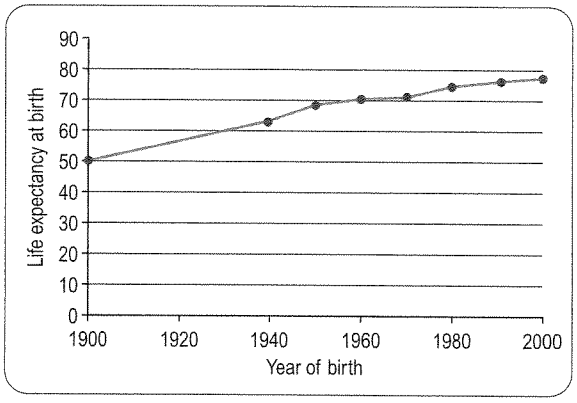


Fig. 2.2
Human lifespan in the USA. (Data from National Centre for Health Statistics, 1998.)

[3]The use of drugs to improve sporting performance is one of many examples of 'therapeutic' practices that find favour among individuals, yet are strongly condemned by society. We do not, as a society, attach disvalue to the possession of merely average sporting ability, even though the individual athlete may take a different view.

[4]Jonathan Swift, in *Gulliver's Travels*, writes of the misery of the Struldbrugs, rare beings with a mark on their forehead who, as they grew older, lost their youth but never died, and who were declared 'dead in law' at the age of 80.

most important determinant of what is considered appropriate to treat. In the end, though, medical practice evolves in a more pragmatic fashion, and such arguments prove to be of limited relevance to the way in which medicine is actually practised, and hence to the therapeutic goals the drug industry sees as commercially attractive. Politics, economics, and above all, social pressures are the determinants, and the limits are in practice set more by our technical capabilities than by issues of theoretical propriety.

Although the drug industry has so far been able to take a pragmatic view in selecting targets for therapeutic intervention, things are changing as technology advances. The increasing cost and sophistication of what therapeutics can offer mean that healthcare systems the world over are being forced to set limits, and are having to go back to the issue of what constitutes disease. Furthermore, by invoking the concept of disease, governments control access to many other social resources (e.g. disability benefits, entry into the armed services, insurance pay-outs, access to life insurance, exemption from legal penalties etc.).

So far, we have concentrated mainly on the impact of disease on individuals and societies. We now need to adopt a more biological perspective, and attempt to put the concept of disease into the framework of contemporary ideas about how biological systems work.

Function and dysfunction: the biological perspective

The dramatic revelations of the last few decades about the molecular basis of living systems have provided a new way of looking at function and dysfunction, and the nature of disease. Needless to say, molecular biology could not have developed without the foundations of scientific biology that were built up in the 19th century. As we saw in Chapter 1, this was the period in which science came to be accepted as the basis on which medical practice had to be built. Particularly significant was cell theory, which established the cell as the basic building block of living organisms. In the words of the pioneering molecular biologist, François Jacob: 'With the cell, biology discovered its atom'. It is by focusing on the instruction sets that define the form and function of cells, and the ways in which these instructions are translated in the process of generating the structural and functional phenotypes of cells, that molecular biology has come to occupy centre stage in modern biology. Genes specify proteins, and the proteins a cell produces determine its structure and function.

From this perspective, deviations from the norm, in terms of structure and function at the cellular level, arise through deviations in the pattern of protein expression by individual cells, and they may arise either through faults in the instruction set itself (genetic mutations) or through environmental factors that alter the way in which the instruction set is translated (i.e. that affect gene expression). We come back to the age-old distinction between inherited and environmental factors (nature and nurture) in the causation of disease, but with a sharper focus: altered gene expression, resulting in altered protein synthesis, is the mechanism through which all these factors operate. Conversely, it can be argued[5] that all therapeutic measures (other than physical procedures, such as surgery) also work at the cellular level, by influencing the same fundamental processes (gene expression and protein synthesis), although the link between a drug's primary target and the relevant effect(s) on gene expression that account for its therapeutic effect may be very indirect. We can see how it has come about that molecular biology, and in particular genomics, has come to figure so largely in the modern drug discovery environment.

Levels of biological organization

Figure 2.3 shows schematically the way in which the genetic constitution of a human being interacts with his or her environment to control function at many different levels, ranging from protein molecules, through single cells, tissues and integrated physiological systems, to the individual, the family and the population at large. For simplicity, we will call this the *bioaxis* . 'Disease', as we have discussed, consists of alterations of function sufficient to cause disability or impaired prognosis at the level of the individual. It should be noted that the arrows along the bioaxis in Figure 2.3 are bidirectional – that is, disturbances at higher levels of organization will in general affect function at lower levels, and vice versa. Whereas it is obvious that genetic mutations can affect function further up the bioaxis (as in many inherited diseases, such as muscular dystrophy, cystic fibrosis or thalassaemia), we should not forget that environmental influences also affect gene function. Indeed, we can state that any long-term phenotypic change (such as weight gain, muscle weakness or depressed mood) *necessarily* involves alterations of gene expression. For example:

[5]This represents the ultimate reductionist view of how living organisms work, and how they respond to external influences. Many still hold out against it, believing that the 'humanity' of man demands a less concrete explanation, and that 'alternative' systems of medicine, not based on our scientific understanding of biological function, have equal validity. Many doctors apparently feel most comfortable somewhere on the middle ground, and society at large tends to fall in behind doctors rather than scientists.

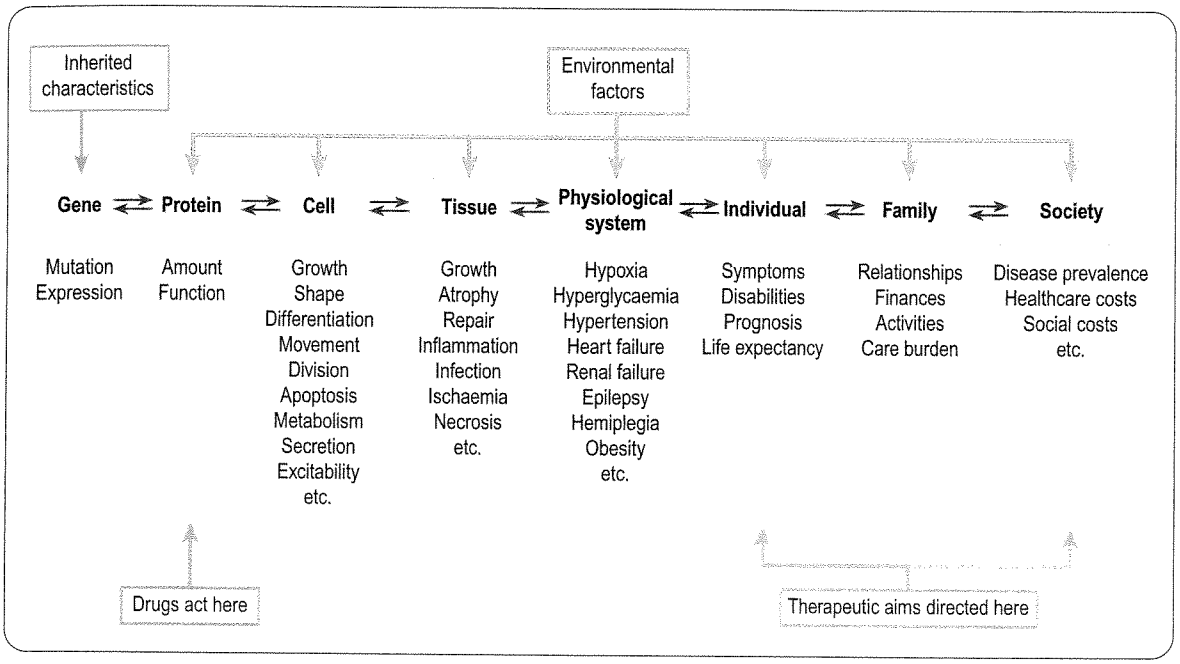


Fig. 2.3
The nature of disease.

- Exposure to a stressful environment will activate the hypothalamopituitary system and thereby increase adrenal steroid secretion, which in turn affects gene transcription in many different cells and tissues, affecting salt metabolism, immune responses and many other functions.
- Smoking, initiated as result of social factors such as peer pressure or advertising, becomes addictive as a result of changes in brain function, phenotypic changes which are in turn secondary to altered gene expression.
- Exposure to smoke carcinogens then increases the probability of cancer-causing mutations in the DNA of the cells of the lung. The mutations, in turn, result in altered protein synthesis and malignant transformation, eventually producing a localized tumour and later, disseminated cancer, with damage to the function of tissues and organs leading to symptoms and premature death.

The pathogenesis of any disease state reveals a similar level of complexity of such interactions between different levels of the bioaxis.

There are two important conclusions to be drawn from the bidirectionality of influence between events at different levels of the bioaxis. One is that it is difficult to pinpoint the *cause* of a given disease. Do we regard the cause of lung cancer in an individual patient as the lack of control over tobacco advertising, the individual's susceptibility to advertising and peer pressure, the state of addiction to nicotine, the act of smoking, the mutational event in the lung epithelial cell, or the individual's inherited tendency to lung cancer? There is no single answer, and the uncertainty should make us wary of the stated claim of many pharmaceutical companies that their aim is to correct the causes rather than the symptoms of disease. The truth, more often than not, is that we cannot distinguish them. Rather, the aim should be to intervene in the disease process in such a way as to minimize the disvalue (disability and impaired prognosis) experienced by the patient.

The second conclusion is that altered gene expression plays a crucial role in pathogenesis and the production of any long-term phenotypic change. If we are thinking of timescales beyond, at maximum, a few hours, any change in the structure and function of cells and tissues will be associated with changes in gene expression. These changes will include those responsible for the phenotypic change (e.g. upregulation of cytokine genes in inflammation, leading to leukocyte accumulation), and those that are consequences of it (e.g. loss of bone matrix following muscle paralysis); some of the latter will, in turn, lead to secondary phenotypic changes, and so on. The pattern of genes expressed in a cell or tissue (sometimes called the 'transcriptome', as distinct from the 'genome', which represents all of the genes present, whether expressed or not), together with the 'proteome' (which describes the array of proteins present in a cell or tissue), provides a uniquely detailed description of how the cell or tissue is behaving. Molecular biology is providing us with powerful methods for mapping the changes in gene and protein expression

associated with different functional states – including disease states and therapeutic responses – and we discuss in more detail in Chapters 6 and 7 the way these new windows on function are influencing the drug discovery process (see Debouck and Goodfellow, 1999).

Therapeutic targets

Traditionally, medicine has regarded the interests of the individual patient as paramount, putting them clearly ahead of those of the community or general population. The primacy of the patient's interests remains the guiding principle for the healthcare professions; in other words, their aim is to address disvalue as experienced by the patient, not to correct biochemical abnormalities, nor to put right the wrongs of society. The principal aim of therapeutic intervention, as shown in Figure 2.3, is therefore to alleviate the condition of the individual patient. Genetic, biochemical or physiological deviations which are not associated with any disvalue for the patient (e.g. possession of a rare blood group, an unusually low heart rate or blood pressure, or blood cholesterol concentration) are not treated as diseases because they neither cause symptoms nor carry an unfavourable prognosis. High blood pressure, or high blood cholesterol, on the other hand, do confer disvalue because they carry a poor prognosis, and are targets for treatment – surrogate targets, in the sense that the actual aim is to remedy the unfavourable prognosis, rather than to correct the physiological abnormality per se.

Although the present and future wellbeing of the individual patient remains the overriding priority for medical care, the impact of disease is felt not only by individuals, but also by society in general, partly for economic reasons, but also for ideological reasons. Reducing the overall burden of disease, as measured by rates of infant mortality, heart disease or AIDS, for example, is a goal for governments throughout the civilized world, akin to the improvement of educational standards. The disease-related disvalue addressed in this case, as shown by the secondary arrow in Figure 2.3, is experienced at the national, rather than the individual level, for individuals will in general be unaware of whether or not they have benefited personally from disease prevention measures. As the therapeutic target has come to embrace the population as a whole, so the financial burden of healthcare has shifted increasingly from individuals to institutional providers of various kinds, mainly national agencies or large-scale commercial healthcare organizations. Associated with this change, there has been a much more systematic focus on assessment in economic terms of the burden of disease (disvalue, to return to our previous terminology) in the community, and the economic cost of healthcare measures. The new and closely related disciplines of *pharmacoeconomics* and *pharmacoepidemiology*, discussed later, reflect the wish (a) to quantify disease-related disvalue and therapeutic benefit in economic terms, and (b) to assess the impact of disease and therapy for the population as a whole, and not just for the individual patient.

The relationship between drug targets and therapeutic targets

There are very few exceptions to the rule, shown in Figure 2.3, that protein molecules are the primary targets of drug molecules. We will come back to this theme repeatedly later, because of its prime importance for the drug discovery process. We should note here that many complex biological steps intervene between the primary drug target and the therapeutic target. Predicting, on the one hand, whether a drug that acts specifically on a particular protein will produce a worthwhile therapeutic effect, and in what disease state, or, on the other hand, what protein we should choose to target in order to elicit a therapeutic effect in a given disease state, are among the thorniest problems for drug discoverers. Molecular biology is providing new insights into the nature of genes and proteins and the relationship between them, whereas time-honoured biochemical and physiological approaches can show how disease affects function at the level of cells, tissues, organs and individuals. The links between the two nevertheless remain tenuous, a fact which greatly limits our ability to relate drug targets to therapeutic effects. Not surprisingly, attempts to bridge this Grand Canyon form a major part of the work of many pharmaceutical and biotechnology companies. Afficionados like to call themselves 'postgenomic' biologists; Luddites argue that they are merely coming down from a genomic 'high' to face once more the daunting complexities of living organisms. We patient realists recognize that a biological revolution has happened, but do not underestimate the time and money needed to bridge the canyon. More of this later.

Therapeutic interventions

Therapeutics in its broadest sense covers all types of intervention aimed at alleviating the effects of disease. The term 'therapeutics' generally relates to procedures based on accepted principles of medical science, that is, on 'conventional' rather than 'alternative' medical

practice[6]. The account of drug discovery presented in this book relates exclusively to conventional medicine – and for this we make no apology – but it needs to be realized that the therapeutic landscape is actually much broader, and includes many non-pharmacological procedures in the domain of conventional medicine, as well as quasi-pharmacological practices (e.g. homeopathy and herbalism) in the 'alternative' domain.

As discussed above, the desired effect of any therapeutic interventions is to improve *symptoms* or *prognosis* or both. From a pathological point of view, therapeutic interventions may be directed at *disease prevention, alleviation* of the effects of existing disease, or permanent *cure* (i.e. restoration to a state of function and prognosis equivalent to those of a healthy individual of the same age, without the need for continuing therapeutic intervention). In practice, there are relatively few truly curative interventions, and they are mainly confined to certain surgical procedures (e.g. removal of circumscribed tumours, fixing of broken bones) and chemotherapy of some infectious and malignant disorders. Most therapeutic interventions aim to alleviate symptoms and/or improve prognosis, and there is increasing emphasis on disease prevention as an objective.

It is important to realize that many types of interventions are carried out with therapeutic intent whose efficacy has not been rigorously tested. This includes not only the myriad alternative medical practices, but also many accepted conventional therapies for which a good scientific basis may exist but which have not been subjected to rigorous clinical trials.

Measuring therapeutic outcome

Effect, efficacy, effectiveness and benefit

These terms have acquired particular meanings – more limited than their everyday meanings – in the context of therapeutic trials.

Pharmacological *effects* of drugs (i.e. their effects on cells, organs and systems) are, in principle, simple to measure in animals, and often also in humans. We can measure effects on blood pressure, plasma cholesterol concentration, cognitive function etc. without difficulty. Such measures enable us to describe quantitatively the pharmacological properties of drugs, but say nothing about their usefulness as therapeutic agents.

Efficacy describes the ability of a drug to produce a desired therapeutic effect in patients under carefully controlled conditions. The gold standard for measurements of efficacy is the randomized controlled clinical trial, described in more detail in Chapter 18. The aim is to discover whether, based on a strictly defined outcome measure, the drug is more or less beneficial than a standard treatment or placebo, in a selected group of patients, under conditions which ensure that the patients actually receive the drug in the specified dose. Proof of efficacy, as well as proof of safety, is required by regulatory authorities as a condition for a new drug to be licensed. Efficacy tests what the drug can do under optimal conditions, which is what the prescriber usually wants to know.

Effectiveness describes how well the drug works in real life, where the patients are heterogeneous, are not randomized, are aware of the treatment they are receiving, are prescribed different doses, which they may or may not take, often in combination with other drugs. The desired outcome is generally less well defined than in efficacy trials, related to general health and freedom from symptoms, rather than focusing on a specific measure. The focus is not on the response of individual patients under controlled conditions, but on the overall usefulness of the drug in the population going about its normal business. Studies of effectiveness are of increasing interest to the pharmaceutical companies themselves, because effectiveness rather than efficacy alone ultimately determines how well the drug will sell, and because effectiveness may depend to a considerable extent on the companies' marketing strategies. Effectiveness measures are also becoming increasingly important to the many agencies that now regulate the provision of healthcare, such as formulary committees, insurance companies, health management organizations, and bodies such as the grandly titled National Institute for Clinical Excellence (NICE), set up by the UK Government in 1999 to advise, on the basis of cost-effectiveness, which drugs and other therapeutic procedures should be paid for under the National Health Service.

Benefit comprises effectiveness expressed in monetary terms. It is popular with economists, as it allows cost and benefit to be compared directly, but treated with deep suspicion by many who find the idea of assigning monetary value to life and wellbeing fundamentally abhorrent.

Returning to the theme of Figure 2.3, we can see that whereas *effect* and *efficacy* are generally measured at the level of cells, tissues, systems and individuals, *effective-*

[6]Scientific doctors rail against the term 'alternative', arguing that if a therapeutic practice can be shown to work by properly controlled trials, it belongs in mainstream medicine. If such trials fail to show efficacy, the practice should not be adopted. Paradoxically, whereas 'therapeutics' generally connotes conventional medicine, the term 'therapy' tends to be used most often in the 'alternative' field.

ness and *benefit* are measures of drug action as it affects populations and society at large. We next consider two growing disciplines that have evolved to meet the need for information at these levels, and some of the methodological problems that they face.

Pharmacoepidemiology and pharmacoeconomics

Pharmacoepidemiology (Strom, 2000) is the study of the use and effects of drugs in human populations, as distinct from individuals, the latter being the focus of clinical pharmacology. The subject was born in the early 1960s, when the problem of adverse drug reactions came into prominence, mainly as a result of the infamous thalidomide disaster. The existence of rare but serious adverse drug reactions which can be detected only by the study of large numbers of subjects, was the initial stimulus for the development of pharmacoepidemiology, and the detection of adverse drug reactions remains an important concern. The identification of Reyes' syndrome as a serious, albeit rare, consequence of using aspirin in children is just one example of a successful pharmacoepidemiological study carried out under the auspices of the US Department of Health and published in 1987. The subject has gradually become broader, however, to cover aspects such as the variability of drug responses between individuals and population groups, the level of compliance of individual patients in taking drugs that are prescribed, and the overall impact of drug therapies on the population as a whole, taking all of these factors into account. The widely used antipsychotic drug *clozapine* provides an interesting example of the importance of pharmacoepidemiological issues in drug evaluation. Clozapine, first introduced in the 1970s, differed from its predecessors, such as haloperidol, in several ways, some good and some bad. On the good side, clozapine has a much lower tendency than haloperidol to cause extrapyramidal motor effects (a serious problem with many antipsychotic drugs), and it appeared to have the ability to improve not only the positive symptoms of schizophrenia (hallucinations, delusions, thought disorder, stereotyped behaviour) but also the negative symptoms (social withdrawal, apathy). Compliance is also better with clozapine, because the patient usually has fewer severe side effects. On the bad side, in about 1% of patients clozapine causes a fall in the blood white cell count (leukopenia), which can progress to an irreversible state of agranulocytosis unless the drug is stopped in time. Furthermore, clozapine does not produce benefit in all schizophrenic patients – roughly one-third fail to show improvement, and there is currently no way of knowing in advance which patients will benefit. Clozapine is also much more expensive than haloperidol. Considered from the perspective of an individual patient, and with hindsight, it is straightforward to balance the pros and cons of using clozapine rather than haloperidol, based on the severity of the extrapyramidal side effects, the balance of positive and negative symptoms which the patient has, whether clozapine is affecting the white cell count, and whether the patient is a responder or a non-responder. From the perspective of the overall population, evaluating the pros and cons of clozapine and haloperidol (or indeed of any two therapies) requires epidemiological data: how frequent are extrapyramidal side effects with haloperidol, what is the relative incidence of positive and negative symptoms, what is the incidence of agranulocytosis with clozapine, what proportion of patients are non-responders, what is the level of patient compliance with haloperidol and clozapine?

In summary, pharmacoepidemiology is a special area of clinical pharmacology which deals with population, rather than individual, aspects of drug action, and provides the means of quantifying *variability* in the response to drugs. Its importance for the drug discovery process is felt mainly at the level of clinical trials and regulatory affairs, for two reasons (Dieck et al., 1994). First, allowing for variability is essential in drawing correct inferences from clinical trials (see Chapter 18). Second, variability in response to a drug is per se disadvantageous, as drug A, whose effects are unpredictable, is less useful than drug B which acts consistently, even though the mean balance between beneficial and unwanted effects may be the same for both. From the population perspective, drug B looks better than drug A, even though for many individual patients the reverse may be true.

Pharmacoeconomics, a branch of health economics, is a subject that grew up around the need for healthcare providers to balance the ever-growing costs of healthcare against limited resources. The arrival of the welfare state, which took on healthcare provision as a national rather than an individual responsibility, was the signal for economists to move in. Good accounts of the basic principles and their application to pharmaceuticals are given by Gold et al. (1996), Johannesson (1996) and McCombs (1998). The aim of pharmacoeconomics is to measure the benefits and costs of drug treatments, and in the end to provide a sound basis for comparing the value for money of different treatments. As might be expected, the subject arouses fierce controversy. Economics in general is often criticized for defining the price of everything but appreciating the value of nothing, and health economics particularly tends to evoke this reaction, as health and quality of life are such ill-defined and subjective, yet highly emotive, concepts. Nevertheless, pharmacoeconomics is a rapidly growing discipline and will undoubtedly have an increasing influence on healthcare provision.

Pharmacoeconomic evaluation of new drugs is often required by regulatory authorities, and is increasingly being used by healthcare providers as a basis for choosing how to spend their money. Consequently, pharmaceutical companies now incorporate such studies into the clinical trials programmes of new drugs. The trend can be seen as a gradual progression towards the right-hand end of the bioaxis in Figure 2.3 in our frame of reference for assessing the usefulness of a new drug. Before 1950, new drugs were often introduced into clinical practice on the basis of studies in animals and a few human volunteers; later, formal randomized controlled clinical trials on carefully selected patient populations, with defined outcome measures, became the accepted standard, along with postmarketing pharmacoepidemiological studies to detect adverse reactions. Pharmacoeconomics represents the further shift of focus to include society in general and its provisions for healthcare. A brief outline of the main approaches used in pharmacoeconomic analysis follows.

Pharmacoeconomics covers four levels of analysis:

- Cost identification
- Cost-effectiveness analysis
- Cost–utility analysis
- Cost–benefit analysis.

Cost identification consists of determining the full cost in monetary units of a particular therapeutic intervention, including hospitalization, working days lost etc., as well as direct drug costs. It pays no attention to outcome, and its purpose is merely to allow the costs of different procedures to be compared. The calculation is straightforward, but deciding exactly where to draw the line (e.g. whether to include indirect costs, such as loss of income by patients and carers) is somewhat arbitrary. Nevertheless, cost identification is the least problematic part of pharmacoeconomics.

Cost-effectiveness analysis aims to quantify outcome as well as cost. This is where the real problems begin. The outcome measure most often used in cost-effectiveness analysis is based on prolongation of life, expressed as *life-years saved per patient treated*. Thus if treatment prolongs the life expectancy of patients, on average, from 3 years to 5 years, the number of life-years gained per patient is 2. Comparing cost and outcome for different treatments then allows the cost per life-year saved to be determined for each. For example, a study of various interventions in coronary heart disease, cited by McCombs (1998), showed that the cost per life-year saved was $5900 for use of a β-adrenoceptor blocker in patients who had suffered a heart attack, the corresponding figure for use of a cholesterol-lowering drug in patients with coronary heart disease was $7200, whereas coronary artery bypass surgery cost $34 000 per life-year saved. Any kind of all-or-nothing event, such as premature births prevented, hospital admissions avoided etc., can be used for this kind of analysis. Its weakness is that it is a very crude measure, making no distinction between years of life spent in a healthy and productive mode and those spent in a state of chronic illness.

Cost–utility analysis is designed to include allowance for quality of life, as well as survival, in the calculation, and is yet more controversial, for it becomes necessary somehow to quantify quality – not an endeavour for the faint-hearted. What the analysis seeks to arrive at is an estimate known as *quality-adjusted life-years (QALYs)*. Thus if the quality of life for a given year, based on the results of the questionnaire, comes out at 70% of the value for an average healthy person of the same age, that year represents 0.7 QALYs, compared with 1 QALY for a year spent in perfect health, the assumption being that 1 year spent at this level of illness is 'worth' 0.7 years spent in perfect health.

Many different questionnaire-based rating scales have been devised to reflect different aspects of an individual's state of health or disability, such as ability to work, mobility, mental state, pain etc. Some relate to specific disease conditions, whereas others aim to provide a general 'quality-of-life' estimate (Jaeschke and Guyatt, 1994), some of the best-known being the *Sickness Impact Profile*, the *Nottingham Health Profile*, the *McMaster Health Index*, and a 36-item questionnaire known as *SF-36*. In addition to these general quality-of-life measures, a range of disease-specific questionnaires have been devised which give greater sensitivity in measuring the specific deficits associated with particular diseases. Standard instruments of this kind are now widely used in pharmacoeconomic studies.

To use such ratings in estimating QALYs it is necessary to position particular levels of disability on a life/death scale, such that 1 represents alive and in perfect health and 0 represents dead. This is where the problems begin in earnest. How can we possibly say what degree of pain is equivalent to what degree of memory loss, for example, or how either compares with premature death? This problem has, of course, received a lot of expert attention (Gold et al., 1996; Drummond et al., 1997; Johannesson, 1996) and various solutions have been proposed, some of which, to the untrained observer, have a distinctly chilling and surreal quality. For example, the standard gamble approach, which is well grounded in the theory of welfare economics, involves asking the individual a question of the following kind:

> Imagine you have the choice of remaining in your present state of health for 1 year or taking a gamble between dying now and living in perfect health for 1 year. What odds would you need to persuade you to take the gamble?[7]

[7]Imagine being asked this by your doctor! 'But I only wanted something for my sore throat', you protest weakly.

If the subject says 50:50, the implication is that he values a year of life in his present state of health at 0.5 QALYs. An alternative method involves asking the patient how many years of life in their present condition he or she would be prepared to forfeit in exchange for enjoying good health until they die. Although there are subtle ways of posing this sort of question, such an evaluation, which most ordinary people find unreal, is implicit in the QALY concept. Figure 2.4 shows schematically the way in which quality of life, as a function of age, may be affected by disease and treatment, the area between the curves for untreated and treated patients representing the QALYs saved by the treatment. In reality, of course, continuous measurements spanning several decades are not possible, so the actual data on which QALY estimates are based in practice are much less than is implied by the idealized diagram in Figure 2.4. Cost–utility analysis results in an estimate of monetary cost per QALY gained, and is becoming widely accepted as a standard method for pharmacoeconomic analysis. Examples of cost per QALY gained range from £3700 for the use of sildenafil (Viagra) in treating erectile dysfunction (Stolk et al., 2000) to £328 000 for the treatment of multiple sclerosis with β-interferon (Parkin et al., 2000), this high value being accounted for by the high cost and limited therapeutic efficacy of the drug. 'Acceptable' thresholds for cost-effectiveness are suggested to be in the range of £8000–£25 000 per QALY gained. In principle, cost–utility analysis allows comparison of one form of treatment against another, and this explains its appeal to those who must make decisions about the allocation of healthcare resources. It has been adopted as the method of choice for pharmacoeconomic analysis of new medicines by several agencies, such as the US Public Health Service and the Australian Pharmaceutical Benefits Advisory Committee.

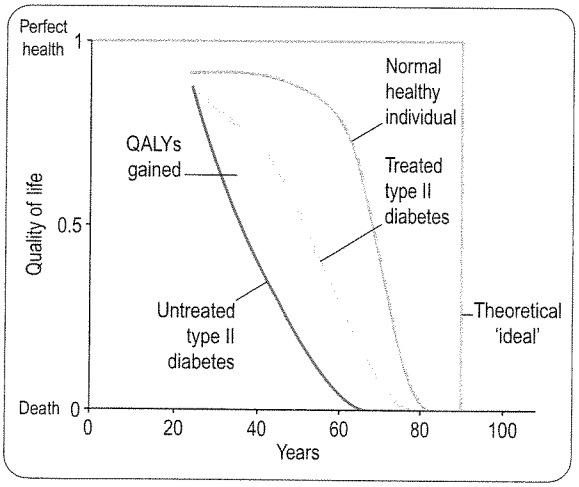


Fig. 2.4
Quality of life affected by disease and treatment.

Hardline economists strive for an absolute scale by which to judge the value of healthcare measures compared with other resource-consuming initiatives that societies choose to support. *Cost–benefit* analysis fulfils this need in principle, by translating healthcare improvements into monetary units that can be directly balanced against costs, to assess whether any given procedure is, on balance, 'profitable'. The science of welfare economics has provided various tools for placing a monetary value on different experiences human beings find agreeable or disagreeable, based generally on the 'willingness-to-pay' principle. Not surprisingly, attempts to value human life and health in cash terms lead rapidly into an ethical and moral minefield, dangerous enough in the context of a single nation and its economy, but much more so in the global context. As a result, cost–benefit analysis has been largely shunned as a practical approach for evaluating medicines.

Summary

In this chapter we have discussed concepts of disease and the aims of therapeutics, the needs newly introduced drugs have to satisfy, and the ways in which their ability to satisfy those needs are judged in practice. There are many uncertainties and ambiguities surrounding the definition of disease, and ideas are constantly shifting, but the two components that most satisfactorily define it are *dysfunction* and *disvalue*. Disvalue, which therapeutic interventions aim to mitigate, in turn has two main components, namely *morbidity* and *prognosis*.

We have described the bioaxis, which represents the various levels in the organizational heirarchy of living systems in general, and human beings in particular, and emphasized that disease inevitably affects all levels on the bioaxis. The drugs that we invent home in very specifically at one level, namely proteins, although the effects we want to produce are at another level, namely individuals. Furthermore, we emphasize that healthcare issues are increasingly being viewed from the perspective of populations and societies, and so the impact of drugs at these levels – even further removed from their primary targets – has to be evaluated. Evaluation of drug effects from these rather lofty perspectives, through the application of the emerging disciplines of pharmacoepidemiology and pharmacoeconomics, although fraught with problems, is an important trend the pharmaceutical industry cannot ignore.

Having taken this rather nervy look at the world about us, we turn in the next chapter to discuss the different therapeutic modalities on which the pharmaceutical and biotechnology industries have focused, before retreating to the safer ground at the left-hand end of the bioaxis, where the modern-day drug discovery business begins.

References

Caplan AL, Engelhardt HT Jr, McCartney JJ (eds) (1981) Concepts of health and disease: interdisciplinary perspectives. London: Addison-Wesley.

Caplan AL (1993) The concepts of health, illness, and disease. In: Bynum WF, Porter R, eds. Companion encyclopedia of the history of medicine. Vol. 1. London: Routledge.

Caplan AL, McCartney JJ, Sisti DA (eds) (2004) Health, disease, and illness: concepts in medicine. Washington, DC: Georgetown University Press.

Debouck C, Goodfellow P N (1999) DNA microarrays in drug discovery and development. Nature Genetics Suppl 21: 48–50.

Dieck G S, Glasser D B, Sachs R M (1994) Pharmacoepidemiology: a view from industry. In: Strom B L, ed. Pharmacoepidemiology. Chichester: John Wiley; 73–85.

Drummond M F, O'Brien B, Stoddart G I, Torrance G W (1997) Methods for the economic evaluation of healthcare programmes. Oxford: Oxford University Press.

Gold M R et al (ed) (1996) Cost-effectiveness in health and medicine. New York: Oxford University Press.

Jaeschke R, Guyatt G H (1994) Using quality-of-life measuremnets in pharmacoepidemiology research. In: Strom BL, ed. Pharmacoepidemiology. Chichester: John Wiley; 495–505.

Johannesson M (1996) Theory and methods of economic evaluation of health care. Dordrecht: Kluwer Academic.

McCombs J S (1998) Pharmacoeconomics: what is it and where is it going? American Journal of Hypertension 11: 112S–119S.

Moynihan R, Heath I, Henry D (2004) Selling sickness: the pharmaceutical industry and disease mongering. British Medical Journal 324: 886–891.

Parkin D, Jacoby A, McNamee P, Miller P, Thomas S, Bates D (2000) Treatment of multiple sclerosis with interferon beta: an appraisal of cost-effectiveness and quality of life. Journal of Neurology, Neurosurgery and Psychiatry 68: 144–149.

Reznek L (1987) The nature of disease. New York: Routledge & Kegan Paul.

Scully JL (2004) What is a disease? EMBO Reports 5: 650–653.

Stolk EA, Busschbach JJ, Caffa M, Meuleman EJ, Rutten FF (2000) Cost utility analysis of sildenafil compared with papaverine–phentolamine injections. British Medical Journal 320: 1156–1157.

Strom B L (ed) (2000) Pharmacoepidemiology. Chichester: John Wiley.

3 Therapeutic modalities

H P Rang
H LeVine

Introduction

Therapeutics in its broadest sense covers all types of intervention aimed at alleviating the effects of disease. The term 'therapeutics' generally relates to procedures based on accepted principles of medical science, that is, on 'conventional' rather than 'alternative' medical practice.

The account of drug discovery presented in this book relates exclusively to conventional medicine – and for this we make no apology – but it needs to be realized that the therapeutic landscape is actually much broader, and includes many non-pharmacological procedures in the domain of conventional medicine, as well as quasi-pharmacological practices in the 'alternative' domain.

As discussed in Chapter 2, the desired effect of any therapeutic intervention is to improve *symptoms* or *prognosis* or both. From a pathological point of view, therapeutic interventions may be directed at *disease prevention, alleviation* of the effects of existing disease, or permanent *cure* (i.e. restoration to a state of function and prognosis equivalent to those of a healthy individual of the same age, without the need for continuing therapeutic intervention). In practice, there are relatively few truly curative interventions, and they are mainly confined to certain surgical procedures (e.g. removal of circumscribed tumours, fixing of broken bones) and chemotherapy of some infectious and malignant disorders. Most therapeutic interventions aim to alleviate symptoms and/or improve prognosis, and there is increasing emphasis on disease prevention as an objective.

It is important to realize that many types of intervention are carried out with therapeutic intent whose efficacy has not been rigorously tested. This includes not only the myriad alternative medical practices, but also many accepted conventional therapies for which a sound scientific basis may exist but which have not been subjected to rigorous clinical trials.

Therapeutic interventions that lie within the field of conventional medicine can be divided into the following broad categories:

- Advice and counselling (e.g. genetic counselling)
- Psychological treatments (e.g. cognitive therapies for anxiety disorders, depression etc.)

- Dietary and nutritional treatments (e.g. gluten-free diets for celiac disease, diabetic diets etc.)
- Physical treatments, including surgery, radiotherapy
- Pharmacological treatments – encompassing the whole of conventional drug therapy
- Biological and biopharmaceutical treatments, a broad category including vaccination, transplantation, blood transfusion, biopharmaceuticals (see Chapter 12), in vitro fertilization etc.

On the fringe of conventional medicine are preparations that fall into the category of 'nutriceuticals' or 'cosmeceuticals'. Nutriceuticals include a range of dietary preparations, such as slimming diets, and diets supplemented with vitamins, minerals, antioxidants, unsaturated fatty acids, fibre etc. These preparations generally have some scientific rationale, although their efficacy has not, in most cases, been established by controlled trials. They are not subject to formal regulatory approval, so long as they do not contain artificial additives other than those that have been approved for use in foods. Cosmeceuticals is a fancy name for cosmetic products similarly supplemented with substances claimed to reduce skin wrinkles, promote hair growth etc. These products achieve very large sales, and some pharmaceutical companies have expanded their business in this direction. We do not discuss these fringe 'ceuticals' in this book, as most pharmaceutical and biotechnology companies restrict themselves to mainstream therapeutic products.

Within each of the medical categories listed above lies a range of procedures: at one end of the spectrum are procedures that have been fully tried and tested and are recognized by medical authorities; at the other is outright quackery of all kinds. Somewhere between lie widely used 'complementary' procedures, practised in some cases under the auspices of officially recognized bodies, which have no firm scientific foundation. Here we find, among psychological treatments, hypnotherapy and analytical psychotherapy; among nutritional treatments, 'health foods', added vitamins, and diets claimed to avoid ill-defined food allergies; among physical treatments, acupuncture and osteopathy; among chemical treatments, homeopathy, herbalism and aromatherapy. Biological procedures lying in this grey area between scientific medicine and quackery are uncommon (and we should probably be grateful for this) – unless one counts colonic irrigation and swimming with dolphins.

In this book we are concerned with the last two treatment categories on the list, summarized in Table 3.1, and in this chapter we consider the current status and future prospects of the three main fields. namely 'conventional' therapeutic drugs, biopharmaceuticals and various biological therapies.

Conventional therapeutic drugs

Small-molecule drugs, either synthetic compounds or natural products have for long been the mainstay of therapeutics and are likely to remain so, despite the rapid growth of biopharmaceuticals in recent years. For their advantages and disadvantages see Box 3.1.

Although the pre-eminent role of conventional small-molecule drugs may decline as biopharmaceutical products grow in importance, few doubt that they will continue to play a major role in medical treatment. New technologies described in Section 2, particularly combinatorial chemistry, high-throughput screening and genomic approaches to target identification, have already brought about a revolution in drug discovery, the fruits of which are only just beginning to appear. There are also high expectations that more sophisticated drug delivery systems (see Chapter 17) will allow drugs to act much more selectively where they are needed, and thus reduce the burden of side effects.

Biopharmaceuticals

For the purposes of this book, biopharmaceuticals are therapeutic protein or nucleic acid preparations made by techniques involving recombinant DNA technology (Walsh, 2003). Although proteins such as insulin and growth hormone, extracted from human or animal tissues, have long been used therapeutically, the era of biopharmaceuticals began in 1982 with the development by Eli Lilly of recombinant human insulin (Humulin), made by genetically engineered *Escherichia coli*. Recombinant human growth hormone (also produced in *E. coli*), erythropoietin (Epogen) and tissue plasminogen activator (tPA) made by engineered mammalian cells followed during the 1980s. This was the birth of the biopharmaceutical industry, and since then new bioengineered proteins have contributed an increasing proportion of new medicines to be registered (see Table 3.1 for some examples, and Chapters 12 and 22 for more details). The scope of protein biopharmaceuticals includes copies of endogenous mediators, blood clotting factors, enzyme preparations and monoclonal antibodies, as well as vaccines. See Box 3.2 for their advantages and disadvantages.

Immunization against infectious diseases dates from 1796, when Jenner first immunized patients against smallpox by infecting them with the relatively harmless cowpox. Many other immunization procedures were developed in the 19th century, and from the 20th century onwards pharmaceutical companies began producing standardized versions of the antigens, often the attenuated or modified organisms themselves, as well

Table 3.1
The main types of chemical therapeutic agent

Type	Source	Examples	Notes
Conventional small-molecule drugs	Synthetic organic compounds*	Most of the pharmacopoeia	The largest category of drugs in use, and of new registrations
	Natural products	Paclitaxel (Taxol) Many antibiotics and anticancer drugs (e.g. penicillins, aminoglycosides, erythromycin) Opiates (e.g. morphine) Statins (e.g. lovastatin) Ciclosporin, fujimycin	Continues to be an important source of new therapeutic drugs
	Semisynthetic compounds (i.e. compounds made by derivatizing natural products)	Penicillin derivatives (e.g. ampicillin) second-generation statins (e.g. simvastatin)	Strategy for generating improved 'second-generation' drugs from natural products
Peptide and protein mediators	Synthetic	Somatostatin Calcitonin Vasopressin	Peptides up to approximately 20 residues can be reliably made by solid-phase synthesis
	Extracted from natural sources (human, animal, microbial)	Insulin, growth hormone, human γ-globulins, botulinum toxin	At one time the only source of such hormones. Now largely replaced by recombinant biotechnology products. γ-globulins still obtained from human blood
	Recombinant DNA technology	Human insulin, erythropoietin, human growth hormone, GM-CSF TNF-α, hirudin	Many different expression systems in use and in development
Antibodies	Animal antisera, human immunoglobulins	Antisera used to treat infections such as hepatitis A and B, diphtheria, rabies, tetanus. Also poisoning by botulinum, snake and spider venoms etc.	
	Monoclonal antibodies	Trastuzumab (directed against epidermal growth factor receptor) Rituximab (directed against B-cell surface antigen)	A rapidly growing class of biopharmaceuticals, with many products in development.
Enzymes	Recombinant DNA technology	Cerebrosidase Dornase Galactosidase	
Vaccines	Infecting organism (killed, attenuated or non-pathogenic strains)	Smallpox, diphtheria, measles, tuberculosis, tetanus, influenza and many others	The conventional approach, still widely used. Some risk of introducing viable pathogens
	Antigens produced by recombinant DNA technology	Many of the above vaccines now available as recombinant antigens	Advantages are greater consistency and elimination of risk of introducing pathogens
DNA products	Recombinant DNA technology	Antisense oligonucleotides (e.g. Vitravene)	Many products in clinical development. Vitravene (for treating cytomegalovirus infection) is the only marketed product so far.

Table 3.1
The main types of chemical therapeutic agent—cont'd

Type	Source	Examples	Notes
Cells	Human donors Engineered cell lines	Various stem cell therapies in development	
Tissues	Human donors Animal tissues Engineered tissues	Apligraf	Bilayer of human skin cells
Organs	Human donors	Transplant surgery	

*Not considered here are many 'adjunct' therapies, such as oxygen, antiseptic agents, anaesthetic agents, intravenous salts etc., which are beyond the scope of this book.

Box 3.1 Advantages and disadvantages of small-molecule drugs

Advantages

- 'Chemical space' is so vast that synthetic chemicals, according to many experts, have the potential to bind specifically to any chosen biological target: the right molecule exists; it is just a matter of finding it.
- Doctors and patients are thoroughly familiar with conventional drugs as medicines, and the many different routes of administration that are available. Clinical pharmacology in its broadest sense has become part of the knowledge base of every practising doctor, and indeed, part of everyday culture. Although sections of the public may remain suspicious of drugs, there are few who will refuse to use them when the need arises.
- Oral administration is often possible, as well as other routes where appropriate.
- From the industry perspective, small-molecule drugs make up more than three-quarters of new products registered over the past decade. Pharmaceutical companies have long experience in developing, registering, producing, packaging and marketing such products.
- Therapeutic peptides are generally straightforward to design (as Nature has done the job), and are usually non-toxic.

Disadvantages

- As emphasized elsewhere in this book, the flow of new small-molecule drugs seems to be diminishing, despite increasing R&D expenditure.
- Side effects and toxicity remain a serious and unpredictable problem, causing failures in late development, or even after registration. One reason for this is that the selectivity of drug molecules with respect to biological targets is by no means perfect, and is in general less good than with biopharmaceuticals.
- Humans and other animals have highly developed mechanisms for eliminating foreign molecules, so drug design often has to contend with pharmacokinetic problems.
- Oral absorption is poor for many compounds. Peptides cannot be given orally.

as antisera which would give immediate passive protection against disease organisms. Vaccines and immune approaches to controlling disease are still a major concern, and increasingly biotechnology-derived vaccines are being developed to improve the efficacy of and reduce the risks associated with preparations made from infectious material.

Overall, biopharmaceuticals offer great promise for the future, and rapid progress is being made in the technologies used to produce them (Scrip Report, 2001). Currently, nearly all approved biopharmaceuticals are proteins, the majority being copies of endogenous mediators, monoclonal antibodies or vaccines. It is possible that most of the clinically useful hormones and mediators that we currently know about have already been produced as biopharmaceuticals, so future advances in this direction are likely to depend on progress in discovering new protein signalling mechanisms. Monoclonal antibodies may offer much broader possibilities, and progress will be greatly facilitated by identifying the genes for important functional proteins, such as key enzymes, transporters etc. Once the DNA sequence of a putative target is known, its amino acid sequence can be inferred and an antibody produced, even if the target protein is of such low abundance that it cannot be isolated biochemically.

Following the wave of successes by the biotechnology industry in producing biopharmaceuticals such as human insulin, erythropoietin and growth hormone during the 1980s and 1990s, medical biotechnology expanded into many other fields, including the development of therapeutic modalities beyond therapeutic proteins and antibodies. Next we briefly discuss two important developments still in the experimental phase, namely gene-based and cell-based therapies, which are under very active investigation.

Gene therapy

Recombinant DNA technology offers the promise of altering the genetic material of cells and thereby correct-

Box 3.2 Advantages and disadvantages of biopharmaceuticals

Advantages

- The main benefit offered by biopharmaceutical products is that they open up the scope of protein therapeutics, which was previously limited to proteins that could be extracted from animal or human sources.
- The discovery process for new biopharmaceuticals is often quicker and more straightforward than is the case with synthetic compounds, as screening and lead optimization are not required.
- Unexpected toxicity is less common than with synthetic molecules.
- The risk of immune responses to non-human proteins – a problem with porcine or bovine insulins – is avoided by expressing the human sequence.
- The risk of transmitting virus or prion infections is avoided.

Disadvantages

- Producing biopharmaceuticals on a commercial scale is expensive, requiring complex purification and quality control procedures.
- The products are not orally active and often have short plasma half-lives, so special delivery systems may be required, adding further to costs. Like other proteins, biopharmaceutical products do not cross the blood–brain barrier.
- For the above reasons, development generally costs more and takes longer, than it does for synthetic drugs.
- Many biopharmaceuticals are species specific in their effects, making tests of efficacy in animal models difficult or impossible.

ing the results of genetic defects, whether inherited or acquired. The techniques for manipulating cellular DNA that underpin much of modern molecular biology have great versatility, and can in principle be applied to therapeutic as well as experimental endeavours. Even where the genetic basis of the disease is not well understood, it should be possible to counteract its effects by genetic, as distinct from pharmacological, means. Further technical information about gene therapy is given in Chapter 12, and in reference works such as Meager (1999), Templeton and Lasic (2000), Kresina (2001), and Brooks (2002). Gene therapy has been actively investigated for more than two decades, and many clinical trials have been performed. So far, however, the results have proved disappointing, and there are currently (2004) no gene therapy products approved for clinical use.

The most widely investigated approach involves introducing new genes to replace missing or dysfunctional ones; this is most commonly done by engineering the new gene into a modified virus (the vector), which has the ability to enter the host cell, causing expression of the artificially introduced gene until the cell dies or expels the foreign DNA. Such non-integrated DNA is usually eliminated quite quickly and is not passed on to the cell's progeny, and so this type of transfection is generally only appropriate in situations where transient expression is all that is required. Retroviral vectors are able to incorporate the new DNA into the host cell's chromosomes, where it will remain and be expressed during the lifetime of the cell and will be passed on to any progeny of that cell. More elaborate gene therapy protocols for treating single-gene disorders are designed actually to correct the disease-producing sequence mutation in the host genome, or to alter gene expression so as to silence dysfunctional genes.

At one time gene therapy directed at germline cells was considered a possibility, the advantage being that an inherited gene defect could be prevented from affecting progeny, and effectively eliminated for good. The serious risks and ethical objections to such human genetic engineering, however, have led to a worldwide ban on germ-cell gene therapy experiments, and efforts are restricted to somatic cell treatments.

How much impact has gene therapy had so far as a therapeutic approach, and what can be expected of it in the future? The first successful trial of gene therapy to be reported was by Anderson and colleagues, who used it in 1990 to replace the dysfunctional gene for the enzyme adenosine deaminase (ADA). ADA deficiency causes *severe combined immunodeficiency syndrome* (SCID), a rare condition which prevents the normal immune response to pathogens, and means that the child can only survive in a germ-free environment. This first gene therapy trial was successful in partly restoring ADA function, but by no means curative. Hundreds of clinical trials were performed during the 1990s, mainly in three clinical areas, namely cancer, AIDS and single-gene inherited disorders such as cystic fibrosis, haemophilia and SCID. Most of these used viral vectors to deliver the DNA, though some used liposome-packaged DNA or other non-viral vectors (see Chapter 17) for this purpose. The genetic material was delivered systemically in some cases, by intravenous or subcutaneous injection; in other cases it was injected directly into solid tumours. An alternative strategy was to harvest bone marrow cells from the patient, transfect these with the necessary DNA construct ex vivo, and return them to the patient so that the genetically modified cells would recolonize the bone marrow and provide the required protein. These techniques had been extensively worked out in laboratory animals, but the clinical results were uniformly disappointing, mainly because transfection rates were too low and expression was too transient. Repeat administration of viral vectors often elicited an immune response which inactivated the vector. So the very high expectation in the early 1990s that gene therapy would revolutionize treatment in many areas of medicine, from arthritis to mental

illness, quickly gave way to a much more guarded optimism, and in some cases a pessimistic dismissal of the whole concept. There were, however, a few cases in which SCID in children was successfully – and apparently permanently – cured by gene therapy, and there were other trials in haemophilia and certain cancers where results looked promising. Alarm bells sounded, first in 1999 when a teenager, Jesse Gelsinger, who was participating in a gene therapy trial in Philadelphia, developed an intense immunological reaction and suddenly died 4 days after treatment. Official scrutiny uncovered many other cases of adverse reactions that had not been reported as they should have been. Many ongoing trials were halted, and much tighter controls were imposed. Subsequently, in 2000, immune function was successfully restored in 18 SCID children, 17 of whom are alive 5 years later (the first therapeutic success for human gene therapy), but two later developed leukaemia, thought to be because integration of the retroviral transgene occurred in a way that activated a cancer-promoting gene, raising even more serious concerns about the long-term side effects of gene therapy.

In the much more cautious atmosphere now prevailing, some carefully controlled trials are beginning to give positive results, mainly in the treatment of haemophilia, but overall, the general view is that gene therapy, while showing great theoretical potential, has so far proved disappointing in its clinical efficacy, amid concerns about its long-term safety and ongoing problems in designing effective delivery systems. (see commentaries by Cavazzana-Calvo et al, 2004 and Relph et al, 2004). Pessimists refer to a decade of failure and note that hundreds of trials have failed so far to produce a single approved therapy. A quick survey of the literature, however, shows a profusion of laboratory studies aimed at improving the technology, and exploring many new ideas for using gene therapy in numerous conditions, ranging from transplant rejection to psychiatric disorders.

The main problems to be overcome are (a) to find delivery vectors that are efficient and selective enough to transfect most or all of the target cells without affecting other cells; (b) to produce long-lasting expression of the therapeutic gene; and (c) to avoid serious adverse effects. Additionally, a method for reversing the effect by turning the foreign gene off if things go wrong would be highly desirable, but has not so far been addressed in trials.

Antisense DNA has been investigated as an alternative to the DNA strategies outlined above. Antisense DNA consists of an oligonucleotide sequence complementary to part of a known mRNA sequence. The antisense DNA binds to the mRNA and, by mechanisms that are not fully understood, blocks expression very selectively, though only for as long as the antisense DNA remains in the cell. The practical problems of developing therapeutic antisense reagents are considerable, as unmodified oligonucleotides are quickly degraded in plasma and do not enter cells readily, so either chemical modification or special delivery systems such as liposomal packaging are required. So far only one antisense preparation has been approved for clinical use, an oligonucleotide used to treat an ocular virus infection in AIDS patients. *Ribozymes*, specific mRNA sequences that inactivate genes by catalysing DNA cleavage, are being investigated as an alternative to antisense DNA, but so far none has been approved for clinical use.

In addition to their chequered clinical trials history, gene therapy products share with other biopharmaceuticals many features that cause major pharmaceutical companies to shy away from investing heavily in such products. The reagents are large molecules, or viruses, that have to be delivered to the appropriate sites in tissues, often to particular cells and with high efficiency. Supplying gene therapy reagents via the bloodstream is only effective for luminal vascular targets, and topical administration is usually needed. Viral vectors do not spread far from the site of injection, nor do they infect all cell types. The vectors have their separate toxicology issues. Commercial production, quality control, formulation and delivery often present problems.

In summary, the theoretical potential of gene therapy is enormous, and the ingenuity being applied to making it work is very impressive. Still, after 25 years of intense research effort no product has been developed, and many of the fundamental problems in delivering genes effectively and controllably still seem far from solution. Most likely, a few effective products for a few specific diseases will be developed and marketed in the next few years, and this trickle will probably grow until gene therapy makes a significant contribution to mainstream therapeutics. Whether it will grow eventually to a flood that supplants much of conventional therapeutics, or whether it will remain hampered by technical problems, nobody can say at this stage. In the foreseeable future, gene therapy is likely to gain acceptance as a useful adjunct to conventional chemotherapy for cancer and viral infections, particularly AIDS. The Holy Grail of a cure for inherited diseases such as cystic fibrosis still seems some way off.

Cell-based therapies

Cell replacement therapies offer the possibility of effective treatment for various kinds of degenerative disease, and much hope currently rests on the potential uses of stem cells, which are undifferentiated progenitor cells that can be maintained in tissue culture and, by the application of appropriate growth factors, be induced to differentiate into functional cells of various kinds. Their ability to divide in culture means that the stock of cells can be expanded as required.

Autologous cell grafts (i.e. returning treated cells to the same individual) are quite widely used for treating leukaemias and similar malignancies of bone marrow

cells. A sample of the patient's bone marrow is taken, cleansed of malignant cells, expanded, and returned to colonize the bone marrow after the patient has been treated with high-dose chemotherapy or radiotherapy to eradicate all resident bone marrow cells. Bone marrow is particularly suitable for this kind of therapy because it is rich in stem cells, and can be recolonized with 'clean' cells injected into the bloodstream.

Apart from this established procedure for treating bone marrow malignancies, only two cell-based therapeutic products have so far gained FDA approval: preparations of autologous chondrocytes used to repair cartilage defects, and autologous keratinocytes, used for treating burns. Other potential applications which have been the focus of much experimental work are reviewed by Fodor (2003). They include:

- Neuronal cells injected into the brain (Isaacson, 2003) to treat neurodegenerative diseases such as Parkinson's disease (loss of dopaminergic neurons), amyotrophic lateral sclerosis (loss of cholinergic neurons) and Huntington's disease (loss of GABA neurons);
- Insulin-secreting cells to treat insulin-dependent diabetes mellitus;
- Cardiac muscle cells to restore function after myocardial infarction.

The major obstacle to further development of such cell-based therapies is that the use of embryonic tissues – the preferred source of stem cells – is severely restricted for ethical reasons. Although stem cells can be harvested from adult tissues and organs, they are less satisfactory. Like gene therapy, cell-based therapeutics could in principle have many important applications, and the technical problems that currently stand in the way are the subject of intensive research efforts. Biotechnology companies are active in developing the necessary tools and reagents that are likely to be needed to select and prepare cells for transplantation.

Tissue and organ transplantation

Transplantation of human organs, such as heart, liver, kidneys and corneas, is of course a well established procedure, many of the problems of rejection having been largely solved by the use of immunosuppressant drugs such as ciclosporin and fujimycin. Better techniques for preventing rejection, including procedures based on gene therapy, are likely to be developed, but the main obstacle remains the limited supply of healthy human organs, and there is little reason to think that this will change in the foreseeable future. The possibility of xenotransplantation – the use of non-human organs, usually from pigs – has received much attention. Cross-species transplants are normally rejected within minutes by a process known as hyperacute rejection. Transgenic pigs whose organs are rendered resistant to hyperacute rejection have been produced, but trials in humans have so far been ruled out because of the risk of introducing pig retroviruses into humans. Despite much discussion and arguments on both sides, there is no sign of this embargo being lifted. Organ transplantation requires such a high degree of organization to get the correct matched organs to the right patients at the right time, as well as advanced surgical and follow-up resources, that it will remain an option only for the privileged minority.

Building two- (e.g. skin) and three-dimensional (e.g. a heart valve) structures that are intended to function mechanically, either from host cells or from banked, certified primordial or stem cell populations, is at the cutting edge of tissue engineering efforts. The aim is to fashion these tissues and organ parts around artificial scaffold materials, and to do this in culture under the control of appropriate growth and differentiation factors. The development of biocompatible scaffolding materials, and achieving the right growth conditions, are problems where much remains to be done. Artificial skin preparations recently became available and others will probably follow.

Also in an early, albeit encouraging, state of development are bionic devices – the integration of mechanical and electronic prostheses with the human body – which will go beyond what has already been accomplished with devices such as cochlear implants and neurally controlled limb prostheses. The role of the major pharmaceutical companies in this highly technological area is likely to be small. The economic realities of the relatively small patient populations will most likely be the limiting factor governing the full implementation of integrated bionics.

Summary

Small organic molecule drugs of molecular weight <500 Da are the preferred therapeutic modality of the major pharmaceutical companies for most disease applications. The advantages summarized above drive this choice. The development over the years of large, chemically diverse small-molecule libraries, many already with 'drug-like' properties (see Chapters 9 and 10) built into their structure, reinforces the commitment. Protein and peptide therapeutics also have their place in the pharmaceutical armamentarium, especially with respect to the immune system and hormonal dysregulation. Many pharmaceutical companies began with immune antisera and vaccines, but the first specialized biotechnology companies took advantage of recombinant DNA methods to produce therapeutic proteins. Although the major pharmaceutical companies have not completely

abandoned protein therapeutics, many of the advances in the field have been made by biotechnology companies.

Protein- and DNA-based biopharmaceuticals often face difficult pharmacokinetic problems, in particular poor absorption, rapid degradation, and inability to enter cells or cross the blood–brain barrier. Their successful development therefore often depends on developing suitable delivery systems that help to overcome these problems. For this reason (and also to improve the performance of conventional therapeutic drugs) drug delivery technology (see Chapter 17) is currently receiving a great deal of attention, with many new polymer- and liposome-based formulations being invented and tested. The right delivery system is as necessary as the right drug, and for biopharmaceuticals the two will generally need to be developed in tandem, rather than first developing a compound and then optimizing the delivery system (which is the development strategy usually adopted for small-molecule drugs).

Somatic (non-germline) gene therapy initially was thought to have great promise for curing inborn errors that lead to disease. Thirty years later, although the technology and our understanding have greatly improved, clinical success has proved elusive. Optimizing vectors and delivery systems so as to produce long-lasting gene expression in the tissues where it is needed has proved much more difficult than expected. Nevertheless, there is reason for optimism in the long term. Currently, gene therapy development is being directed mainly at life-threatening disorders such as cancer, AIDs and haemophilia, where the need is greatest and the risks are balanced by the severity of the diseases. It is likely to be another decade or two before gene therapy begins to make a broader clinical impact.

The involvement of pharmaceutical companies in the transplantation field is largely confined to improving the immunosuppressant drugs that are needed to protect transplants from immune rejection. The use of transplants is severely restricted by the availability of human organs, and hopes for improving the situation by the use of xenografts are unlikely to be realized in the foreseeable future. Stem-cell technologies are likely to be used successfully for certain kinds of tissue repair and cell replacement; biotechnology companies, rather than pharmaceutical companies, are likely to make the running in these new fields. Currently, techniques such as bone marrow transplants are being developed and used successfully by clinical teams without any necessary input from commercial research. Probably their use will become more routine, but it seems unlikely that the market size for commercial products in this area will be enough for a large pharmaceutical company.

References

Brooks G (ed) (2002) Gene therapy: the use of DNA as a drug. New York: John Wiley and Sons.

Cavazzano-Calvo M, Thrasher A, Mavilio F (2004) The future of gene therapy. Nature 427: 779–781.

Fodor W L (2003) Tissue engineering and cell based therapies, from the bench to the clinic: the potential to replace, repair and regenerate. Reproductive Biology and Endocrinology 1: 102–107.

Isaacson O (2003) The production and use of cells as therapeutic agents in neurodegenerative diseases. Lancet Neurology 2: 417–424.

Kresina T F (ed) (2001) An introduction to molecular medicine and gene therapy. New York: John Wiley and Sons.

Meager A (1999) Gene therapy technologies: Applications and regulations from laboratory to clinic. Chichester: John Wiley and Sons.

Relph K, Harrington K, Pandha H (2004) Recent developments and current status of gene therapy using viral vectors in the United Kingdom. Br Med J 329: 839–842.

Scrip Report (2001) Biopharmaceuticals: a new era of discovery in the biotechnology revolution.

Templeton N S, Lasic D D (2000) Gene therapy: therapeutic mechanisms and strategies. New York: Marcel Dekker.

Walsh G (2003) Biopharmaceuticals, 2nd edn. Chichester: John Wiley and Sons Ltd.