

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

effects of the amino-acid substitutions on the stability of human protective protein cathepsin A (Rudenko *et al.*, 1998).

- (ii) The modification of Tyr393 α to Asn in the branched-chain 2-oxo acid dehydrogenase occurs at the interface of the α and β subunits in this $\alpha_2\beta_2$ heterotetramer, providing a nice explanation of the ‘mennonite’ variants of maple syrup urine disease (MSUD) (Ævarsson *et al.*, 2000).

Impressive as the insights obtained into the causes of diseases like these might be, there is almost a sense of tragedy associated with this detailed understanding of a serious, sometimes fatal, afflictions at the atomic and three-dimensional level: there is often so little one can *do* with this knowledge. There are at least two, very different, reasons for this. The first reason is that turning a malfunctioning protein or nucleic acid into one that functions properly is notoriously difficult. Treatment would generally require the oral use of small molecules that somehow counteract the effect of the mutation, *i.e.* the administration of the small molecule has to result in a functional complex of the drug with the mutant protein. This is in almost all cases far more difficult than finding compounds that *block* the activity of a protein or nucleic acid – which is the way in which most current drugs function. The second reason for the paucity of drugs for treating genetic diseases is very different in nature: the number of patients suffering from a particular mutation responsible for a genetic disease is very small in most cases. This means that market forces do not encourage funding the expensive steps of testing the toxicity and efficacy of potentially pharmaceutically active compounds. One of several exceptions is sickle cell anaemia, where significant efforts have been made to arrive at pharmaceutically active agents (Rolan *et al.*, 1993). In this case the mutation Glu6 β Val leads to deoxyhaemoglobin polymerization *via* the hydrophobic valine. In spite of several ingenious approaches based on the allosteric properties of haemoglobin (Wireko & Abraham, 1991), no successful compound seems to be on the horizon yet for the treatment of sickle cell anaemia.

More recently, the spectacular molecular mechanisms underlying genetic serpin deficiency diseases have been elucidated. A typical example is α_1 -antitrypsin deficiency, which leads to cirrhosis and emphysema. Normal α_1 -antitrypsin, a serine protease inhibitor, exposes a peptide loop as a substrate for the cognate proteinase in its active but metastable conformation. After cleavage of the loop, the protease becomes trapped as an acyl-enzyme with the serpin, and the cleaved serpin loop inserts itself as the central strand of one of the serpin β -sheets, accompanied by a dramatic change in protein stability. In certain mutant serpins, however, the exposed loop is conformationally more metastable and occasionally inserts itself into the β -sheet of a neighbouring serpin molecule, thereby forming serpin polymers with disastrous consequences for the patient (Carrell & Gooptu, 1998). *In vitro*, the polymerization of α_1 -antitrypsin can be reversed with synthetic homologues of the exposed peptide loop (Skinner *et al.*, 1998). This approach might be useful for other ‘conformational diseases’, which include Alzheimer’s and other neurodegenerative disorders.

Another frequently occurring genetic disease is cystic fibrosis. Here we face a more complex situation than that in the case of sickle cell anaemia: a range of different mutations causes a malfunctioning of the same ion channel, which, consequently, leads to a range of severity of the disease (Collins, 1992). Protein crystallography is currently helpful in an indirect way in alleviating the problems of cystic fibrosis patients, not by studying the affected ion channel itself, but by revealing the structure of

leukocyte elastase (Bode *et al.*, 1986), an enzyme responsible for much of the cellular damage associated with cystic fibrosis (Birrer, 1995). On the basis of the elastase structure, inhibitors were developed to combat the effects of the impaired ion channel (Warner *et al.*, 1994). Also, structures of key enzymes of *Pseudomonas aeruginosa*, a bacterium affecting many cystic fibrosis patients, form a basis for the design of therapeutics to treat infections by this pathogen. Yet, to the best of our knowledge, no compound has been developed so far that repairs the malfunctioning ion channel.

However, in some cases there might be more opportunities than assumed so far. Several mutations leading to genetic diseases result in a lack of stability of the affected protein. In instances when the mutant protein is still stable enough to fold, small molecules could conceivably be discovered that bind ‘anywhere’ to a pocket of these proteins, thereby stabilizing the protein. The same small molecule could even be able to increase the stability of proteins with *different* mildly destabilizing mutations. Such an approach, though not trivial by any means, might be worth pursuing. Proof of principle of this concept has recently been provided for several unstable p53 mutants, where the *same* small molecule enhanced the stability of *different* mutants (Foster *et al.*, 1999).

Of course, mutations that destroy cofactor binding or active sites, or destroy proper recognition of partner proteins, will be extremely difficult to correct by small molecules targeting the affected protein. In such instances, gene therapy is likely to be the way by which our and the next generation may be able to improve the lives of future generations.

1.3.4. Crystallography and development of novel pharmaceuticals

The impact of detailed knowledge of protein and nucleic acid structures on the design of new drugs has already been significant, and promises to be of tremendous importance in the next decades. The first structure of a known major drug bound to a target protein was probably that of methotrexate bound to dihydrofolate reductase (DHFR) (Matthews *et al.*, 1977). Even though the source of the enzyme was bacterial while methotrexate is used as a human anticancer agent, this protein–drug complex structure was nevertheless a hallmark achievement. It is generally accepted that the first protein-structure-inspired drug actually reaching the market was captopril, which is an anti-hypertensive compound blocking the action of angiotensin-converting enzyme, a metalloprotease. In this case, the structure of zinc-containing carboxypeptidase A was a guide to certain aspects of the chemical modification of lead compounds (Cushman & Ondetti, 1991). This success has been followed up by numerous projects specifically aimed at the design of new inhibitors, or activators, of carefully selected drug targets.

Structure-based drug design (SBDD) (Fig. 1.3.4.1) is the subject of several books and reviews that summarize projects and several success stories up until the mid-1990s (Kuntz, 1992; Perutz, 1992; Verlinde & Hol, 1994; Whittle & Blundell, 1994; Charifson, 1997; Veerapandian, 1997). Possibly the most dramatic impact made by SBDD has been on the treatment of AIDS, where the development of essentially all of the protease inhibitors on the market in 1999 has been guided by, or at least assisted by, the availability of numerous crystal structures of protease–inhibitor complexes.

The need for a large number of structures is common in all drug design projects and is due to several factors. One is the tremendous challenge for theoretical predictions of the correct

1. INTRODUCTION

binding mode and affinity of inhibitors to proteins. The current force fields are approximate, the properties of water are treacherous, the flexibility of protein and ligands lead quickly to a combinatorial explosion, and the free-energy differences between various binding modes are small. All this leads to the need for several experimental structures in a structure-based drug design cycle (Fig. 1.3.4.1). In this cycle, numerous disciplines are interacting in multiple ways. Many institutions, small and large, are following in one way or another this paradigm to speed up the lead discovery, lead optimization and even the bioavailability improvement steps in the drug development process. Moreover, a very powerful synergism exists between combinatorial chemistry and structure-based drug design. Structure-guided combinatorial libraries can utilize knowledge of ligand target sites in the design of the library [see *e.g.* Ferrer *et al.* (1999), Eckert *et al.* (1999) and Minke *et al.* (1999)]. Once tight-binding ligands are found by combinatorial methods, crystal structures of library compound–target complexes provide detailed information for new highly specific libraries.

The fate of a drug candidate during clinical tests can hinge on a single methyl group – just as a point mutation can alter the benefit of a wild-type protein molecule into the nightmare of a life-long genetic disease. Hence, many promising inhibitors eventually fail to be of benefit to patients. Nevertheless, knowledge of a *series* of protein structures in complex with inhibitors is of immense value in the design and development of future pharmaceuticals. In the following sections some examples will be looked at.

1.3.4.1. Infectious diseases

1.3.4.1.1. Viral diseases

Some icosahedral pathogenic viruses have all their capsid proteins elucidated, while for the more complex viruses like influenza virus, hepatitis C virus (HCV) and HIV, numerous individual protein structures have been solved (Table 1.3.4.1). However, not all 14 native proteins of the HIV genome have yet surrendered to the crystallographic community, nor to the NMR spectroscopists or the high-resolution electron microscopists, our partners in experimental structural biology (Turner & Summers, 1999). Nevertheless, the structures of HIV protease, reverse

transcriptase and fragments of HIV integrase and of HIV viral core and surface proteins are of tremendous value for developing novel anti-AIDS therapeutics [Arnold *et al.*, 1996; Lin *et al.*, 1998; Wlodawer & Vondrasek, 1998; see also references in Table 1.3.4.1(a)]. A similar situation occurs for hepatitis C virus. The protease structure of this virus has been solved recently (simultaneously by four groups!), as well as its helicase structure, providing platforms on the basis of which the design of novel drugs is actively pursued (Le *et al.*, 1998).

A quite spectacular example of how structural knowledge can lead to the synthesis of powerful inhibitors is provided by influenza virus neuraminidase. The structure of a neuraminidase–transition-state analogue complex suggested the addition of positively charged amino and guanidinium groups to the C4 position of the analogue, which resulted, in one step, in a gain of four orders of magnitude in binding affinity for the target enzyme (von Itzstein *et al.*, 1993).

1.3.4.1.2. Bacterial diseases

A very large number of structures of important drug target proteins of bacterial origin have been solved crystallographically (Table 1.3.4.2). Currently, the most important single infectious bacterial agent is *Mycobacterium tuberculosis*, with three million deaths and eight million new cases annually (Murray & Salomon, 1998). The crystal structures of several *M. tuberculosis* potential and proven drug target proteins have been elucidated (Table 1.3.4.2). The complete *M. tuberculosis* genome has been sequenced recently (Cole *et al.*, 1998), and this will undoubtedly have a tremendous impact on future drug development.

The crystal structures of many bacterial dihydrofolate reductases, the target of several antimicrobials including trimethoprim, have also been reported. Recently, the atomic structure of dihydropteroate synthase (DHPS), the target of sulfa drugs, has been elucidated, almost 60 years after the first sulfa drugs were used to treat patients (Achari *et al.*, 1997; Hampele *et al.*, 1997).

A very special set of bacterial proteins are the toxins. Some of these have dramatic effects, with even a single molecule able to kill a host cell. These toxins outsmart and (mis)use many of the defence systems of the host, and their structures are often most unusual and fascinating, as recently reviewed by Lacy & Stevens (1998). The structures of the toxins are actively used for the design of prophylactics and therapeutic agents to treat bacterial diseases [see *e.g.* Merritt *et al.* (1997), Kitov *et al.* (2000) and Fan *et al.* (2000)]. It is remarkable that the properties of these devastating toxins are also used, or at least considered, for the treatment of disease, such as in the engineering of diphtheria toxin to create immunotoxins for the treatment of cancer and the deployment of cholera toxin mutants as adjuvants in mucosal vaccines. Knowledge of the three-dimensional structures of these toxins assists in the design of new therapeutically useful proteins.

1.3.4.1.3. Protozoan infections

A major cause of death and world-wide suffering is due to infections by several protozoa, including:

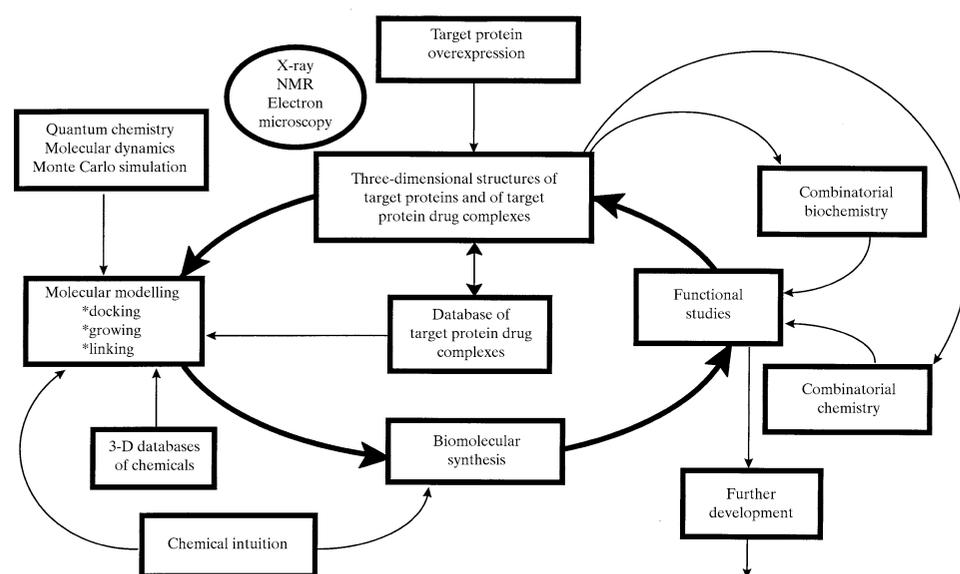


Figure 1.3.4.1
The structure-based drug design cycle.

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

Table 1.3.4.1

Important human pathogenic viruses and their proteins

(a) RNA viruses
(i) Single-stranded

Family	Example	Protein structures solved	Reference
<i>Arenaviridae</i>	Lassa fever virus	None	
<i>Bunyaviridae</i>	Hantavirus	None	
<i>Caliciviridae</i>	Hepatitis E virus, Norwalk virus	None	
<i>Coronaviridae</i>	Corona virus	None	
<i>Deltaviridae</i>	Hepatitis D virus	Oligomerization domain of antigen	[1]
<i>Filoviridae</i>	Ebola virus	GP2 of membrane fusion glycoprotein	[2]
<i>Flaviviridae</i>	Dengue	NS3 protease	[3]
	Hepatitis C	NS3 protease	[4], [5]
		RNA helicase	[6]
	Yellow fever	None	
<i>Orthomyxoviridae</i>	Tick-borne encephalitis virus	Envelope glycoprotein	[7]
	Influenza virus	Neuraminidase	[8]
		Haemagglutinin	[9]
		Matrix protein M1	[10]
<i>Paramyxoviridae</i>	Measles, mumps, parainfluenza, respiratory syncytial virus	None	
<i>Picornaviridae</i>	Hepatitis A virus	3C protease	[11]
	Poliovirus	Capsid	[12]
		RNA-dependent polymerase	[13]
	Rhinovirus	Capsid	[14]
		3C protease	[15]
<i>Retroviridae</i>	Echovirus	Capsid	[16]
	HIV	Capsid protein	[17]
		Matrix protein	[18]
		Protease	[19], [20], [21]
		Reverse transcriptase	[22], [23], [47], [48], [49]
		Integrase	[24]
		gp120	[25]
		NEF	[26]
		gp41	[27]
<i>Rhabdovirus</i>	Rabies virus	None	
<i>Togaviridae</i>	Rubella	None	

(ii) Double-stranded

Family	Example	Protein structures solved	Reference
<i>Reoviridae</i>	Rotavirus	None	

(b) DNA viruses
(i) Single-stranded

Family	Example	Protein structures solved	Reference
<i>Parvoviridae</i>	B 19 virus	None	

- (a) *Plasmodium falciparum* and related species, causing various forms of malaria;
- (b) *Trypanosoma cruzi*, the causative agent of Chagas' disease in Latin America;
- (c) *Trypanosoma brucei*, causing sleeping sickness in Africa;
- (d) Some eleven different *Leishmania* species, responsible for several of the most horribly disfiguring diseases known to mankind.

Drug resistance, combined with other factors, has been the cause of a major disappointment for the early hopes of a 'malaria eradication campaign'. Fortunately, new initiatives have been launched recently under the umbrella of the 'Malaria roll back' program and the 'Multilateral Initiative for Malaria' (MIM). We are facing a formidable challenge, however, since the parasite is very clever at evading the immune response of the human host. Drugs are the mainstay of current treatments and may well be so for a long time to come. Protein crystallographic studies of *Plasmodium* proteins are hampered by the unusual codon usage of the *Plasmodium* species, coupled with a tendency to contain

insertions of numerous hydrophilic residues in the polypeptide chain (Gardner *et al.*, 1998) which provide sometimes serious obstacles to obtaining large amounts of *Plasmodium* proteins for structural investigations.

However, the structures of an increasing number of potential drug targets from these protozoan parasites are being unravelled. These include the variable surface glycoproteins (VSGs) and glycolytic enzymes of *Trypanosoma brucei*, crucial malaria proteases, and the remarkable trypanothione reductase (Table 1.3.4.3). The structures of nucleotide phosphoribosyl transferases of a variety of protozoan parasites have also been elucidated. Moreover, the structure of DHFR from *Pneumocystis carinii*, the major opportunistic pathogen in AIDS patients in the United States, has been determined. Several of these structures are serving as starting points for the development of new drugs.

1.3.4.1.4. Fungi

In general, the human immune system is able to keep the growth of fungi under control, but in immuno-compromised

1. INTRODUCTION

Table 1.3.4.1 (continued)

(ii) Double-stranded

Family	Example	Protein structures solved	Reference
<i>Adenoviridae</i>	Adenovirus	Protease Capsid Knob domain of fibre protein	[28] [29] [30]
<i>Hepadnaviridae</i>	Hepatitis B	Capsid	[31]
<i>Herpesviridae</i>	Cytomegalovirus	Protease	[32], [33], [34]
	Epstein–Barr virus	Domains of nuclear antigen 1 BCRF1	[35] [36]
	Herpes simplex	Protease Thymidine kinase Uracyl-DNA glycosylase Core of VP16	[37] [38] [39] [40]
<i>Papovaviridae</i>	Varicella zoster	Protease	[42]
	Papillomavirus	DNA-binding domain of E2 Activation domain of E2	[43] [44]
<i>Poxviridae</i>	Smallpox virus	None	
	Vaccinia virus (related to smallpox but non-pathogenic)	Methyltransferase VP39 Domain of topoisomerase	[45] [46]

References: [1] Zuccola *et al.* (1998); [2] Weissenhorn *et al.* (1998); [3] Murthy *et al.* (1999); [4] Love *et al.* (1996); [5] Yan *et al.* (1998); [6] Yao *et al.* (1997); [7] Rey *et al.* (1995); [8] Varghese *et al.* (1983); [9] Wilson *et al.* (1981); [10] Sha & Luo (1997); [11] Allaire *et al.* (1994); [12] Hogle *et al.* (1985); [13] Hansen *et al.* (1997); [14] Rossmann *et al.* (1985); [15] Matthews *et al.* (1994); [16] Filman *et al.* (1998); [17] Worthylake *et al.* (1999); [18] Hill *et al.* (1996); [19] Navia, Fitzgerald *et al.* (1989); [20] Wlodawer *et al.* (1989); [21] Erickson *et al.* (1990); [22] Rodgers *et al.* (1995); [23] Ding *et al.* (1995); [24] Dyda *et al.* (1994); [25] Kwong *et al.* (1998); [26] Lee *et al.* (1996); [27] Chan *et al.* (1997); [28] Ding *et al.* (1996); [29] Roberts *et al.* (1986); [30] Xia *et al.* (1994); [31] Wynne *et al.* (1999); [32] Tong *et al.* (1996); [33] Qiu *et al.* (1996); [34] Shieh *et al.* (1996); [35] Bochkarev *et al.* (1995); [36] Zdanov *et al.* (1997); [37] Hoog *et al.* (1997); [38] Wild *et al.* (1995); [39] Savva *et al.* (1995); [40] Liu *et al.* (1999); [42] Qiu *et al.* (1997); [43] Hegde & Androphy (1998); [44] Harris & Botchan (1999); [45] Hodel *et al.* (1996); [46] Sharma *et al.* (1994); [47] Kohlstaedt *et al.* (1992); [48] Jacobo-Molina *et al.* (1993); [49] Ren *et al.* (1995).

patients (*e.g.* as a result of cancer chemotherapy, HIV infection, transplantation patients receiving immunosuppressive drugs, genetic disorders) such organisms are given opportunities they usually do not have. *Candida albicans* is an opportunistic fungal organism which causes serious complications in immunocompromised patients. Several of its proteins have been structurally characterized (Table 1.3.4.3) and provide opportunities for the development of selectively active inhibitors.

1.3.4.1.5. Helminths

Worms affect the lives of billions of human beings, causing serious morbidity in many ways. *Onchocerca volvulus* is the cause of river blindness, which resulted in the virtual disappearance of entire villages in West Africa, until ivermectin appeared. This remarkable compound dramatically reduced the incidence of the disease, even though it does not kill the adult worms. Therefore, a biological clock is ticking, waiting until resistance occurs against this single compound available for treatment. *Schistosoma* species are responsible for a wide variety of liver diseases and are spreading continuously since irrigation schemes provide a perfect environment for the intermediate snail vector. Other medically important helminths are *Wuchereria bancrofti* and *Brugia malayi*. Only a few protein structures from these very important disease-causing organisms have been unravelled so far (Table 1.3.4.3).

1.3.4.2. Resistance

Resistance to drugs in infectious organisms, as well as in cancers, is a fascinating subject, since it demonstrates the action and reaction of biological systems in response to environmental challenges. Life, of course, has been evolving to do just that – and the arrival of new chemicals, termed ‘drugs’, on the scene is nothing new to organisms that are the result of evolutionary processes involving billions of years of chemical warfare. Populations of organisms span a wide range of variation at the genetic and protein levels, and the chance that one of the variants is able to cope with drug pressure is nonzero. The variety of mechanisms

observed to be responsible for drug resistance is impressive (Table 1.3.4.4).

Crystallography has made major contributions to the detailed molecular understanding of resistance in the case of detoxification, mutation and enzyme replacement mechanisms. Splendid examples are:

- The beta-lactamases: These beta-lactam degrading enzymes, of which there are four classes, are produced by many bacteria to counteract penicillins and cephalosporins, the most widely used antibiotics on the planet. No less than 53 structures of these enzymes reside in the PDB.
- HIV protease mutations: Tens of mutations have been characterized structurally. Many alter the active site at the site of mutation, thereby preventing drug binding. Other mutations rearrange the protein backbone, reshaping entire pockets in the binding site (Erickson & Burt, 1996).
- HIV reverse transcriptase mutations: *Via* structural studies, at least three mechanisms of drug resistance have been elucidated: direct alteration of the binding sites for the nucleoside analogue or non-nucleoside inhibitors, mutations that change the position of the DNA template, and mutations that induce conformational changes that propagate into the active site (Das *et al.*, 1996; Hsiou *et al.*, 1998; Huang, Chopra *et al.*, 1998; Ren *et al.*, 1998; Sarafianos *et al.*, 1999).
- Resistance to vancomycin: In non-resistant bacteria, vancomycin stalls the cell-wall synthesis by binding to the D-Ala-D-Ala terminus of the lipid-PP-disaccharide-pentapeptide substrate of the bacterial transglycosylase/transpeptidase, thereby sterically preventing the approach of the substrate. Resistant bacteria, however, have acquired a plasmid-borne transposon encoding for five genes, *vanS*, *vanR*, *vanH*, *vanA* and *vanX*, that allows them to synthesise a substrate ending in D-Ala-D-lactate. This minute difference, an oxygen atom replacing an NH, leads to a 1000-fold reduced affinity for vancomycin, explaining the resistance (Walsh *et al.*, 1996). Thus far, the structures of vanX (Bussiere *et al.*, 1998) and D-Ala-D-Ala ligase as a model for vanA (Fan *et al.*, 1994) have

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

Table 1.3.4.2

Protein structures of important human pathogenic bacteria

Organism	Disease(s)	Protein structures solved	Reference
<i>Staphylococcus aureus</i>	Abscesses	Alpha-haemolysin	[1]
	Endocarditis	Aureolysin	[2]
	Gastroenteritis	Beta-lactamase	[3]
	Toxic shock syndrome	Collagen adhesin	[4]
		7,8-Dihydroneopterin aldolase	[5]
		Dihydropteroate synthetase	[6]
		Enterotoxin A	[7]
		Enterotoxin B	[8]
		Enterotoxin C2	[9]
		Enterotoxin C3	[10]
		Exfoliative toxin A	[11]
		Ile-tRNA-synthetase	[12]
		Kanamycin nucleotidyltransferase	[13]
		Leukocidin F	[14]
		Nuclease	[15]
		Staphopain	[16]
		Staphylokinase	[17]
		Toxic shock syndrome toxin-1	[18]
<i>Staphylococcus epidermidis</i>	Implant infections	None	
<i>Enterococcus faecalis</i> (<i>Streptococcus faecalis</i>)	Urinary tract and biliary tract infections	NADH peroxidase	[19]
<i>Streptococcus mutans</i>	Endocarditis	Histidine-containing phosphocARRIER protein	[20]
<i>Streptococcus pneumoniae</i>	Pneumonia	Glyceraldehyde-3-phosphate dehydrogenase	[21]
	Meningitis, upper respiratory tract infections	Penicillin-binding protein PBP2x	[22]
<i>Streptococcus pyogenes</i>	Pharyngitis	Dpm DNA adenine methyltransferase	[23]
	Scarlet fever, toxic shock syndrome, immunologic disorders (acute glomerulonephritis and rheumatic fever)	Inosine monophosphate dehydrogenase	[24]
		Pyrogenic exotoxin C	[25]
<i>Bacillus anthracis</i>	Anthrax	Anthrax protective antigen	[26]
<i>Bacillus cereus</i>	Food poisoning	Beta-amylase	[27]
		Beta-lactamase II	[28]
		Neutral protease	[29]
		Oligo-1,6-glucosidase	[30]
		Phospholipase C	[31]
<i>Clostridium botulinum</i>	Botulism	Neurotoxin type A	[32]
<i>Clostridium difficile</i>	Pseudomembranous colitis	None	
<i>Clostridium perfringens</i>	Gas gangrene	Alpha toxin	[33]
	Food poisoning	Perfringolysin O	[34]
<i>Clostridium tetani</i>	Tetanus	Toxin C fragment	[35]
<i>Corynebacterium diphtheriae</i>	Diphtheria	Toxin	[36]
		Toxin repressor	[37]
<i>Listeria monocytogenes</i>	Meningitis, sepsis	Phosphatidylinositol-specific phospholipase C	[38]
<i>Actinomyces israelii</i>	Actinomycosis	None	
<i>Nocardia asteroides</i>	Nocardiosis	None	
<i>Neisseria gonorrhoeae</i>	Gonorrhoea	Type 4 pilin	[39]
		Carbonic anhydrase	[40]
<i>Neisseria meningitidis</i>	Meningitis	Dihydroipoamide dehydrogenase	[41]
<i>Bordetella pertussis</i>	Whooping cough	Toxin	[42]
		Virulence factor P.69	[43]
<i>Brucella sp.</i>	Brucellosis	None	
<i>Campylobacter jejuni</i>	Enterocolitis	None	
<i>Enterobacter cloacae</i>	Urinary tract infection, pneumonia	Beta-lactamase: class C	[44]
		UDP-N-acetylglucosamine enolpyruvyltransferase	[45]
<i>Escherichia coli</i>			
ETEC (enterotoxigenic)	Traveller's diarrhoea	Heat-labile enterotoxin	[46]
		Heat-stable enterotoxin (is a peptide)	[47]
EHEC (enterohaemorrhagic)	HUS	Verotoxin-1	[48]
EPEC (enteropathogenic)	Diarrhoea	None	
EAEC (enteroaggregative)	Diarrhoea	None	
EIEC (enteroinvasive)	Diarrhoea	None	
UPEC (uropathogenic)		FimH adhesin	[49]
		FimC chaperone	[49]
		PapD	[50]
NMEC (neonatal meningitis)	Meningitis	None	
<i>Francisella tularensis</i>	Tularemia	None	
<i>Haemophilus influenzae</i>	Meningitis, otitis media, pneumonia	Diaminopimelate epimerase	[51]
		6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase	[52]
		Ferric iron binding protein Mirp	[53]
<i>Klebsiella pneumoniae</i>	Urinary tract infection, pneumonia, sepsis	β -Lactamase SHV-1	[54]
<i>Legionella pneumophila</i>	Pneumonia	None	
<i>Pasteurella multocida</i>	Wound infection	None	

1. INTRODUCTION

Table 1.3.4.2 (continued)

Organism	Disease(s)	Protein structures solved	Reference
<i>Proteus mirabilis</i>	Pneumonia, urinary tract infection	Catalase	[55]
<i>Proteus vulgaris</i>	Urinary tract infections	Glutathione S-transferase	[56]
		Pvu II DNA-(cytosine N4) methyltransferase	[57]
		Pvu II endonuclease	[58]
		Tryptophanase	[59]
<i>Salmonella typhi</i>	Typhoid fever	None, but many for <i>S. typhimurium</i> linked with zoonotic disease	
<i>Salmonella enteridis</i>	Enterocolitis	None	
<i>Serratia marcescens</i>	Pneumonia, urinary tract infection	Serralysin	[60]
		Aminoglycoside 3- <i>N</i> -acetyltransferase	[61]
		Chitinase A	[62]
		Chitinase	[63]
		Endonuclease	[64]
		Hasa (haemophore)	[65]
		Prolyl aminopeptidase	[66]
<i>Shigella sp.</i>	Dysentery	Chloramphenicol acetyltransferase	[67]
		Shiga-like toxin I	[68]
<i>Vibrio cholerae</i>	Cholera	Cholera toxin	[69], [70]
		DSBA oxidoreductase	[71]
		Neuraminidase	[104]
<i>Yersinia enterocolitica</i>	Enterocolitis	Protein-Tyr phosphatase YOPH	[72]
<i>Yersinia pestis</i>	Plague	None	
<i>Pseudomonas aeruginosa</i>	Wound infection, urinary tract infection, pneumonia, sepsis	Alkaline metalloprotease	[73]
		Amidase operon	[74]
		Azurin	[75]
		Cytochrome 551	[76]
		Cytochrome <i>c</i> peroxidase	[77]
		Exotoxin A	[78]
		<i>p</i> -Hydroxybenzoate hydroxylase	[79]
		Hexapeptide xenobiotic acetyltransferase	[80]
		Mandelate racemase	[81]
		Nitrite reductase	[82]
		Ornithine transcarbamoylase	[83]
		Porphobilinogen synthase	[84]
		Pseudolysin	[85]
<i>Burkholderia cepacia</i>	Wound infection, urinary tract infection, pneumonia, sepsis	Biphenyl-cleaving extradiol dioxygenase	[86]
		<i>cis</i> -Biphenyl-2,3-dihydrodiol-2,3-dehydrogenase	[87]
		Dialkylglycine decarboxylase	[88]
		Lipase	[89]
		Phthalate dioxygenase reductase	[90]
<i>Stenotrophomonas maltophilia</i> (= <i>Pseudomonas maltophilia</i>)	Sepsis	None	
<i>Bacteroides fragilis</i>	Intra-abdominal infections	Beta-lactamase type 2	[91]
<i>Mycobacterium leprae</i>	Leprosy	Chaperonin-10 (GroES homologue)	[92]
		RUVA protein	[93]
<i>Mycobacterium tuberculosis</i>	Tuberculosis	3-Dehydroquininate dehydratase	[94]
		Dihydrofolate reductase	[95]
		Dihydropteroate synthase	[96]
		Enoyl acyl-carrier-protein reductase (InhA)	[97]
		Mechanosensitive ion channel	[98]
		Quinolinate phosphoribosyltransferase	[99]
		Superoxide dismutase (iron dependent)	[100]
		Iron-dependent repressor	
<i>Mycobacterium bovis</i>	Tuberculosis	Tetrahydrodipicolinate- <i>N</i> -succinyltransferase	[102]
<i>Chlamydia psittaci</i>	Psittacosis	None	
<i>Chlamydia pneumoniae</i>	Atypical pneumonia	None	
<i>Chlamydia trachomatis</i>	Ocular, respiratory and genital infections	None	
<i>Coxiella burnetii</i>	Q fever	None	
<i>Rickettsia sp.</i>	Rocky Mountain spotted fever	None	
<i>Borrelia burgdorferi</i>	Lyme disease	Outer surface protein A	[103]
<i>Leptospira interrogans</i>	Leptospirosis	None	
<i>Treponema pallidum</i>	Syphilis	None	
<i>Mycoplasma pneumoniae</i>	Atypical pneumonia	None	

References: [1] Song *et al.* (1996); [2] Banbula *et al.* (1998); [3] Herzberg & Moulton (1987); [4] Symersky *et al.* (1997); [5] Hennig *et al.* (1998); [6] Hampele *et al.* (1997); [7] Sundstrom *et al.* (1996); [8] Papageorgiou *et al.* (1998); [9] Papageorgiou *et al.* (1995); [10] Fields *et al.* (1996); [11] Vath *et al.* (1997); [12] Silvan *et al.* (1999); [13] Pedersen *et al.* (1995); [14] Pedelacq *et al.* (1999); [15] Loll & Lattman (1989); [16] Hofmann *et al.* (1993); [17] Rabijns *et al.* (1997); [18] Prasad *et al.* (1993); [19] Yeh *et al.* (1996); [20] Jia *et al.* (1993); [21] Cobessi *et al.* (1999); [22] Pares *et al.* (1996); [23] Tran *et al.* (1998); [24] Zhang, Evans *et al.* (1999); [25] Roussel *et al.* (1997); [26] Petosa *et al.* (1997); [27] Mikami *et al.* (1999); [28] Carli *et al.* (1995); [29] Paupit *et al.* (1988); [30] Watanabe *et al.* (1997); [31] Hough *et al.* (1989); [32] Lacy *et al.* (1998); [33] Naylor *et al.* (1998); [34] Rossjohn, Feil, McKinstry *et al.* (1997); [35] Umland *et al.* (1997); [36] Choe *et al.* (1992); [37] Qiu *et al.* (1995); [38] Moser *et al.* (1997); [39] Parge *et al.* (1995); [40] Huang, Xue *et al.* (1998); [41] Li de la Sierra *et al.* (1997); [42] Stein *et al.* (1994); [43] Emsley *et al.* (1996); [44] Lobkovsky *et al.* (1993); [45] Schonbrunn *et al.* (1996); [46] Sixma *et al.* (1991); [47] Ozaki *et al.* (1991); [48] Stein *et al.* (1992); [49] Choudhury *et al.* (1999); [50] Sauer *et al.* (1999); [51] Cirilli *et al.* (1993); [52] Hennig *et al.* (1999); [53] Bruns *et al.* (1997); [54] Kuzin *et al.* (1999); [55] Gouet *et al.* (1995); [56] Rossjohn, Polekhina *et al.* (1998); [57] Gong *et al.* (1997); [58] Athanasiadis *et al.* (1994); [59] Isupov *et al.* (1998); [60] Baumann (1994); [61] Wolf *et al.* (1998); [62] Perrakis *et al.* (1994); [63] Tews *et al.* (1996); [64] Miller *et al.* (1994); [65] Arnoux *et al.* (1999); [66] Yoshimoto *et al.* (1999); [67] Murray *et al.* (1995); [68] Ling *et al.* (1998); [69] Merritt *et al.* (1994); [70] Zhang *et al.* (1995); [71] Hu *et al.* (1997); [72] Stuckey *et al.* (1994); [73] Miyatake *et al.* (1995); [74] Pearl *et al.* (1994); [75] Adman *et al.* (1978); [76] Almasy & Dickerson (1978); [77] Fulop *et al.* (1995); [78] Allured *et al.* (1986); [79] Gatti *et al.* (1994); [80] Beaman *et al.* (1998); [81] Kallarakal *et al.* (1995); [82] Nurizzo *et al.* (1997); [83] Villeret *et al.* (1995); [84] Frankenberg *et al.* (1999); [85] Thayer *et al.* (1991); [86] Han *et al.* (1995); [87] Hulsmeyer *et al.* (1998); [88] Toney *et al.* (1993); [89] Kim *et al.* (1997); [90] Correll *et al.* (1992); [91] Concha *et al.* (1996); [92] Mande *et al.* (1996); [93] Roe *et al.* (1998); [94] Gourley *et al.* (1999); [95] Li *et al.* (2000); [96] Baca *et al.* (2000); [97] Dessen *et al.* (1995); [98] Chang, Spencer *et al.* (1998); [99] Sharma *et al.* (1998); [100] Cooper *et al.* (1995); [102] Beaman *et al.* (1997); [103] Li *et al.* (1997); [104] Crennell *et al.* (1994).

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

Table 1.3.4.3

Protein structures of important human pathogenic protozoa, fungi and helminths

(a) Protozoa

Organism	Disease	Protein structures solved	Reference	
<i>Acanthamoeba sp.</i> <i>Cryptosporidium parvum</i> <i>Entamoeba histolytica</i> <i>Giardia lamblia</i> <i>Leishmania sp.</i>	Opportunistic meningoencephalitis, corneal ulcers	Actophorin	[1]	
		Profilin	[2]	
	Cryptosporidiosis	None		
		None		
	Amoebic dysentery, liver abscesses	None		
		Adenine phosphoribosyltransferase	[3]	
	Giardiasis	Dihydrofolate reductase-thymidylate synthase	[4]	
		Glyceraldehyde-3-phosphate dehydrogenase	[5]	
	<i>Plasmodium sp.</i>	Malaria	Leishmanolysin	[6]
			Nucleoside hydrolase	[7]
Pyruvate kinase			[8]	
Triosephosphate isomerase			[9]	
Fructose-1,6-bisphosphate aldolase			[10]	
Lactate dehydrogenase			[11]	
MSP1			[12]	
Plasmepsin II			[13]	
Purine phosphoribosyltransferase			[14]	
Triosephosphate isomerase			[15]	
<i>Pneumocystis carinii</i> <i>Toxoplasma gondii</i>	Pneumonia	Dihydrofolate reductase	[16]	
	Toxoplasmosis	HGXPRtase	[17]	
<i>Trichomonas vaginalis</i> <i>Trypanosoma brucei</i>	Trichomoniasis	UPRTase	[18]	
		None		
	Sleeping sickness	Fructose-1,6-bisphosphate aldolase	[19]	
		Glyceraldehyde-3-phosphate dehydrogenase	[20]	
		6-Phosphogluconate dehydrogenase	[21]	
		Phosphoglycerate kinase	[22]	
		Triosephosphate isomerase	[23]	
		VSG	[24]	
		Cruzain (cruzipain)	[25]	
		Glyceraldehyde-3-phosphate dehydrogenase	[26]	
<i>Trypanosoma cruzi</i>	Chagas' disease	Hypoxanthine phosphoribosyltransferase	[27]	
		Triosephosphate isomerase	[28]	
		Trypanothione reductase	[29]	
		Tyrosine aminotransferase	[30]	

(b) Fungi

Organism	Disease	Protein structures solved	Reference
<i>Aspergillus fumigatus</i> <i>Blastomyces dermatidis</i> <i>Candida albicans</i>	Aspergillosis	Restrictocin	[31]
	Blastomycosis	None	
	Candidiasis	Dihydrofolate reductase	[32]
		<i>N</i> -Myristoyl transferase	[33]
		Phosphomannose isomerase	[34]
		Secreted Asp protease	[35]
<i>Coccidioides immitis</i> <i>Cryptococcus neoformans</i> <i>Histoplasma capsulatum</i> <i>Mucor sp.</i> <i>Paracoccidioides brasiliensis</i> <i>Rhizopus sp.</i>	Coccidioidomycosis	None	
	Cryptococcosis	None	
	Histoplasmosis	None	
	Mucormycosis	None	
	Paracoccidioidomycosis	None	
	Phycomycosis	Lipase II	[36]
		Rhizopuspepsin	[37]
		RNase Rh	[38]

been solved. They provide an exciting basis for arriving at new antibiotics against vancomycin-resistant bacteria.

(e) DHFR: Some bacteria resort to the 'ultimate mutation' in order to escape the detrimental effects of antibiotics. They simply replace the entire targeted enzyme by a functionally identical but structurally different enzyme. A prime example is the presence of a dimeric plasmid-encoded DHFR in certain trimethoprim-resistant bacteria. The structure proved to be unrelated to that of the chromosomally encoded monomeric DHFR (Narayana *et al.*, 1995).

Clearly, the structural insight gained from these studies provides us with avenues towards methods for coping with the

rapid and alarming spread of resistance against available anti-biotics that threatens the effective treatment of bacterial infections of essentially every person on this planet. This implies that we will constantly have to be aware of the potential occurrence of mono- and also *multi*-drug resistance, which has profound consequences for drug-design strategies for essentially all infectious diseases. It requires the development of *many* different compounds attacking many different target proteins and nucleic acids in the infectious agent. It appears to be crucial to use, from the very beginning, several new drugs *in combination* so that the chances of the occurrence of resistance are minimal. Multi-drug regimens have been spectacularly successful in the case of leprosy and HIV. Obviously, the development of vaccines is by far the

1. INTRODUCTION

Table 1.3.4.3 (continued)

(c) Helminths

Organism	Disease	Protein structures solved	Reference	
<i>Clonorchis sinensis</i>	Clonorchiasis	None	[39]	
<i>Fasciola hepatica</i>	Fascioliasis	Glutathione S-transferase		
<i>Fasciolopsis buski</i>	Fasciolopsiasis	None		
<i>Paragonimus westermani</i>	Paragonimiasis	None	[39], [40]	
<i>Schistosoma sp.</i>	Schistosomiasis	Glutathione S-transferase		
		Hexokinase	[41]	
<i>Diphyllobothrium latum</i>	Diphyllobothriasis	None		
<i>Echinococcus granulosus</i>	Unilocular hydatid cyst disease	None		
<i>Taenia saginata</i>	Taeniasis	None		
<i>Taenia solium</i>	Taeniasis	None		
<i>Ancylostoma duodenale</i>	Old World hookworm disease	None		
<i>Anisakis</i>	Anisakiasis	None		
<i>Ascaris lumbricoides</i>	Ascariasis	Haemoglobin		[42]
		Major sperm protein		[43]
		Trypsin inhibitor		[44]
<i>Enterobius vermicularis</i>	Pinworm infection	None		
<i>Necator</i>	New World hookworm disease	None		
<i>Strongyloides stercoralis</i>	Strongyloidiasis	None		
<i>Trichinella spiralis</i>	Trichinosis	None		
<i>Trichuris trichiura</i>	Whipworm infection	None		
<i>Brugia malayi</i>	Filariasis	Peptidylprolyl isomerase	[45], [46]	
<i>Dracunculus medinensis</i>	Guinea worm disease	None		
<i>Loa loa</i>	Loiasis	None		
<i>Onchocerca volvulus</i>	River blindness	None		
<i>Toxocara canis</i>	Visceral larva migrans	None		
<i>Wuchereria bancrofti</i>	Lymphatic filariasis (elephantiasis)	None		

References: [1] Leonard *et al.* (1997); [2] Liu *et al.* (1998); [3] Phillips *et al.* (1999); [4] Knighton *et al.* (1994); [5] Kim *et al.* (1995); [6] Schlagenhauf *et al.* (1998); [7] Shi, Schramm & Almo (1999); [8] Rigden *et al.* (1999); [9] Williams *et al.* (1999); [10] Kim *et al.* (1998); [11] Read *et al.* (1999); [12] Chitarra *et al.* (1999); [13] Silva *et al.* (1996); [14] Shi, Li *et al.* (1999); [15] Velanker *et al.* (1997); [16] Champness *et al.* (1994); [17] Schumacher *et al.* (1996); [18] Schumacher *et al.* (1998); [19] Chudzik *et al.* (2000); [20] Vellieux *et al.* (1993); [21] Phillips *et al.* (1998); [22] Bernstein *et al.* (1998); [23] Wierenga *et al.* (1987); [24] Freymann *et al.* (1990); [25] McGrath *et al.* (1995); [26] Souza *et al.* (1998); [27] Focia *et al.* (1998); [28] Maldonado *et al.* (1998); [29] Lantwin *et al.* (1994); [30] Blankenfeldt *et al.* (1999); [31] Yang & Moffat (1995); [32] Whitlow *et al.* (1997); [33] Weston *et al.* (1998); [34] Cleasby *et al.* (1996); [35] Cutfield *et al.* (1995); [36] Kohno *et al.* (1996); [37] Suguna *et al.* (1987); [38] Kurihara *et al.* (1992); [39] Rossjohn, Feil, Wilce *et al.* (1997); [40] McTigue *et al.* (1995); [41] Mulichak *et al.* (1998); [42] Yang *et al.* (1995); [43] Bullock *et al.* (1996); [44] Huang *et al.* (1994); [45] Mikol *et al.* (1998); [46] Taylor *et al.* (1998).

better solution, but it is not always possible. Antigenic variation, see *e.g.* the influenza virus, requires global vigilance and constant re-engineering of certain vaccines every year. Moreover, for higher organisms, and even for many bacterial species like *Shigella* (Levine & Noriega, 1995), with over 50 serotypes per species, the development of successful vaccines has, unfortunately, proved to be very difficult indeed. For sleeping sickness, the development of a vaccine is generally considered to be impossible. It is most likely, therefore, that world health will depend for centuries on a wealth of therapeutic drugs, together with many other measures, in order to keep the immense number of pathogenic organisms under control.

1.3.4.3. Non-communicable diseases

Of this large and diverse category of human afflictions we have already touched upon genetic disorders in Section 1.3.3 above. Other major types of non-communicable diseases include cancer, aging disorders, diabetes, arthritis, and cardiovascular and neurological illnesses. The field of non-communicable diseases is immense. Describing in any detail the current projects in, and potential impact of, protein and nucleic acid crystallography on these diseases would need more space than this entire volume on macromolecular crystallography. Hence, only a few selected examples out of the hundreds which could be described can be discussed here. Table 1.3.4.5 lists many examples of human protein structures elucidated without any claim as to completeness – it is simply impossible to keep up with the fountain of structures being determined at present. Yet, such tables do provide, it is hoped, an overview of what has been achieved and what needs to be done.

1.3.4.3.1. Cancers

Over a hundred different cancers have been described and clearly the underlying defect, loss of control of cell division, can be the result of many different shortcomings in different cells. The research in this area proceeds at a feverish pace, yet the development, discovery and design of effective but safe anti-cancer agents are unbelievably difficult challenges. The modifications needed to turn a normal cell into a malignant one are very small, hence the chance of arriving at ‘true’ anti-cancer drugs that exploit such small differences between normal and abnormal cells is exceedingly small. Nevertheless, such selective anti-cancer agents would leave normal cells essentially unaffected and are therefore the holy grail of cancer therapy. Few if any such compounds have been found so far, but cancer therapy is benefiting from a gradual increase in the number of useful compounds. Many have serious side effects, weaken the immune system and are barely tolerated by patients. However, they rescue large numbers of patients and hence it is of interest that many targets of these compounds, proteins and DNA molecules, have been structurally elucidated by crystallographic methods – often in complex with the cancer drug. The mode of action of many anti-cancer compounds is well understood, *e.g.* methotrexate targeting dihydrofolate reductase, and fluorouracil targeting thymidilate synthase. These are specific enzyme inhibitors acting along principles well known in other areas of medicine. Several anti-cancer drugs display unusual modes of action, such as:

- (a) the DNA intercalators daunomycin (Wang *et al.*, 1987) and adriamycin (Zhang *et al.*, 1993);
- (b) cisplatin, which forms DNA adducts (Giulian *et al.*, 1996);

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

Table 1.3.4.4

Mechanisms of resistance

Overexpress target protein
Mutate target protein
Use other protein with same function
Remove target altogether
Overexpress detoxification enzyme
Mutate detoxification enzyme
Create new detoxification enzyme
Mutate membrane porin protein
Remove or underexpress membrane porin protein
Overexpress efflux pumps
Mutate efflux pumps
Create/steal new efflux pumps
Improve DNA repair
Mutate prodrug conversion enzyme

- (c) taxol, which not only binds to tubulin but also to bcl-2, thereby blocking the machinery of cancer cells in two entirely different ways (Amos & Lowe, 1999);
- (d) camptothecin analogues, such as irinotecan and topotecan, which have the unusual property of prolonging the lifetime of a covalent topoisomerase–DNA complex, generating major road blocks on the DNA highway and causing DNA breakage and cell death;
- (e) certain compounds function as minor-groove binders, *e.g.* netropsin and distamycin (Kopka *et al.*, 1985);
- (f) completely new drugs which were developed based on the structures of matrix metalloproteinases, purine nucleotide phosphorylase and glycinamide ribonucleotide formyl-transferase and which are in clinical trials (Jackson, 1997).

Meanwhile, it is sad that crystallography has not yet made any contribution to the molecular understanding of multi-drug resistance in cancer. The resistance is caused by cellular pumps that efficiently pump out the drugs, often leading to failed chemotherapy (Borst, 1999). On the other hand, the structures of major oncogenic proteins such as p21 (DeVos *et al.*, 1988; Pai *et al.*, 1989; Krenkel *et al.*, 1990; Scheffzek *et al.*, 1997) and p53 (Cho *et al.*, 1994; Gorina & Pavletich, 1996) are of tremendous importance for future structure-based design of anti-neoplastic agents.

1.3.4.3.2. Diabetes

The hallmark characteristic of type I diabetes is a lack of insulin. A major therapeutic approach to this problem is insulin replacement therapy. Unfortunately, the insulin requirements of the body vary dramatically during the course of a day, with high concentrations needed at meal times and a basal level during the rest of the day. Only monomeric insulin is active at the insulin receptor level, but insulin has a natural tendency to form dimers and hexamers that dissociate upon dilution. Thanks to the three-dimensional insight obtained from dozens of insulin crystal structures, as wild-type (Hodgkin, 1971), mutants (Whittingham *et al.*, 1998) and in complex with zinc ions and small molecules such as phenol (Derewenda *et al.*, 1989), it has been possible to fine-tune the kinetics of insulin dissociation. The resulting availability of a variety of insulin preparations with rapid or prolonged action profiles has improved the quality of life of millions of people (Brange, 1997).

1.3.4.3.3. Blindness

The main causes of blindness worldwide are cataract, trachoma, glaucoma and onchocerciasis (Thylefors *et al.*, 1995). Trachoma and onchocerciasis are parasitic diseases that destroy

the architecture of the eye; they were discussed in Section 1.3.4.1. The other two are discussed here. During cataract development, the lens of the eye becomes non-transparent as a result of aggregation of crystallins, preventing image formation. Crystal structures of several mammalian beta- and gamma-crystallins are known, but no human ones yet. In glaucoma, the optic nerve is destroyed by high intra-ocular pressure. One way to lower the pressure is to inhibit carbonic anhydrase II, a pivotal enzyme in maintaining the intra-ocular pressure. On the basis of the carbonic anhydrase crystal structure, researchers at Merck Research Laboratories were able to guide the optimization of an *S*-thienothiopyran-2-sulfonamide lead into a marketed drug for glaucoma: dorzolamide (Baldwin *et al.*, 1989).

1.3.4.3.4. Cardiovascular disorders

Thrombosis is a major cause of morbidity and mortality, especially in the industrial world. Hence, major effort is expended by pharmaceutical industries in the development of new classes of anti-coagulants with fewer side effects than available drugs, such as heparins and coumarins. Because blood coagulation is the result of an amplification cascade of enzymatic reactions, many potential targets are available. At present most of the effort is directed towards thrombin (Weber & Czarniecki, 1997) and factor Xa (Ripka, 1997), responsible for the penultimate step and the step immediately preceding it in the cascade, respectively. Thrombin is especially fascinating owing to the presence of at least three subsites: a primary specificity pocket with the catalytic serine-protease machinery, an exosite for recognizing extended fibrinogen and an additional pocket for binding heparin. This knowledge has led to the design of bivalent inhibitors which occupy two sites with ultra-high affinity and exquisite specificity. Several of these agents are in clinical trials (Pineo & Hull, 1999).

1.3.4.3.5. Neurological disorders

Even a quick glance at Table 1.3.4.5 shows that crystallography contributes to new therapeutics for numerous human afflictions and diseases. Yet there are major gaps in our understanding of protein functions, in particular of those involved in development and in neurological functions. These proteins are the target of many drugs obtained by classical pre-crystal-structure methods. These proven drug targets are very often membrane proteins involved in neuronal functions, and the diseases concerned are some of the most prevalent in mankind. A non-exhaustive list includes cerebrovascular disease (strokes), Parkinson's, epilepsy, schizophrenia, bipolar disease and depression.

Some of these diseases are heart-breaking afflictions, where parents have to accept the suicidal tendencies of their children, often with fatal outcomes; where partners have to endure the tremendous mood swings of their bipolar spouses and have to accept extreme excesses in behaviour; where a happy evening of life is turned into the gradual and sad demise of human intellect due to the progression of Alzheimer's, or to the loss of motor functions due to Parkinson's, or into the tragic stare of a victim of deep depression. Human nature, in all its shortcomings, has the tendency to try to help such tragic victims, but drugs for neurological disorders are rare, drug regimens are difficult to optimize and the commitment to follow a drug regimen – often for years, and often with major side effects – is a next to impossible task in many cases. New, better drugs are urgently needed and hence the structure determinations of the 'molecules of the brain' are major scientific as well as medical challenges of the next decades. Such

1. INTRODUCTION

Table 1.3.4.5

Important human protein structures in drug design

Proteins from other species that might have been studied as substitutes for human ones were left out because of space limitations. We apologize to the researchers affected.

Pharmacological category	Protein	Reference	
Synaptic and neuroeffector junctional function	None		
	None		
Central nervous system function			
Inflammation	Fibroblast collagenase (MMP-1) (also important in cancer)	[1], [2], [3], [4]	
	Gelatinase	[5], [6]	
	Stromelysin-1 (MMP-3) (also important in cancer)	[7], [8], [9], [10], [11]	
	Matrilysin (MMP-7) (also important in cancer)	[12]	
	Neutrophil collagenase (MMP-8) (also important in cancer)	[13], [14], [15], [16]	
	Collagenase-3 (MMP-13)	[17]	
	Human neutrophil elastase (also important for cystic fibrosis)	[18], [19], [20]	
	Interleukin-1 beta converting enzyme (ICE)	[21], [22]	
	p38 MAP kinase	[23], [24]	
	Phospholipase A2	[25], [26], [27]	
	Renin	[28]	
	Renal and cardiovascular function	None	
		17-Beta-hydroxysteroid dehydrogenase	[29], [30]
		BRCT domain (BRCA1 C-terminus)	[31]
		Bcr-Abl kinase	[32]
		Cathepsin B	[33]
		Cathepsin D	[34], [35]
CDK2		[36]	
CDK6		[37]	
DHFR		[38], [39]	
Acidic fibroblast growth factor (FGF)		[40]	
FGF receptor tyrosine kinase domain		[41]	
Glycinamide ribonucleotide formyl transferase		[42]	
Interferon-beta		[43]	
MMPs: see Inflammation			
p53		[44], [45]	
p60 Src		[46]	
Purine nucleoside phosphorylase		[47]	
ras p21	[48], [49], [50], [51]		
Serine hydroxymethyltransferase	[52]		
S-Adenosylmethionine decarboxylase	[53]		
Thymidylate synthase	[54]		
Topoisomerase I	[55], [56]		
Tumour necrosis factor	[57]		
Interleukin 1-alpha	[58]		
Interleukin 1-beta	[59]		
Interleukin 1-beta receptor	[60], [61]		
Interleukin 8	[62]		
Immunomodulation	Calcineurin	[63]	
	Cathepsin S	[64]	
	Cyclophilin	[65], [66], [67]	
	Immunophilin FKBP12	[68], [69], [70]	
	Inosine monophosphate dehydrogenase	[71]	
	Interferon-gamma	[72], [73]	
	Lymphocyte-specific kinase Lck	[74]	
	PNP	[47]	
	Syk kinase	[75]	
	Tumour necrosis factor	[57]	
	ZAP Tyr kinase	[76]	
	Interleukin 2	[77]	
	Interleukin 5	[78]	
	Erythropoietin receptor	[79], [80]	
	AT III	[81], [82], [83], [84]	
	Factor III	[85], [86]	
	Factor VII	[87]	
Factor IX	[88]		
Factor X	[89]		
Factor XIII	[90]		
Factor XIV	[91]		
Fibrinogen: fragment	[92], [93]		
Plasminogen activator inhibitor (PAI)	[94], [95], [96]		
Thrombin	[97], [98], [99]		
tPA	[100]		
Urokinase-type plasminogen activator	[101]		
von Willebrand factor	[102], [103], [104]		
Haematopoiesis			
Coagulation			

1.3. MACROMOLECULAR CRYSTALLOGRAPHY AND MEDICINE

Table 1.3.4.5 (continued)

Pharmacological category	Protein	Reference
Hormones and hormone receptors	Insulin	[105]
	Insulin receptor	[106], [107]
	Human growth hormone + receptor	[108]
	Oestrogen receptor	[109], [110]
	Progesterone receptor	[111]
	Prolactin receptor	[112]
Ocular function	Carbonic anhydrase	[113]
Genetic diseases	See Table 1.3.3.1	
Drug binding	Human serum albumin	[114], [115]
Drug metabolism	Glutathione S-transferase A-1	[116], [117]
	Glutathione S-transferase A4-4	[118]
	Glutathione S-transferase Mu-1	[119]
	Glutathione S-transferase Mu-2	[120]
Neurodegeneration	Aldose reductase	[121]
	JNK3	[122]
Osteoporosis	Cathepsin K	[123], [64]
	Src SH2	[126]
Various	Interferon-alpha 2b	[124]
	Bcl-xL	[125]

References: [1] Borkakoti *et al.* (1994); [2] Lovejoy, Cleasby *et al.* (1994); [3] Lovejoy, Hassell *et al.* (1994); [4] Spurlino *et al.* (1994); [5] Libson *et al.* (1995); [6] Gohlke *et al.* (1996); [7] Becker *et al.* (1995); [8] Dhanaraj *et al.* (1996); [9] Esser *et al.* (1997); [10] Gomis-Ruth *et al.* (1997); [11] Finzel *et al.* (1998); [12] Browner *et al.* (1995); [13] Bode *et al.* (1994); [14] Reinemer *et al.* (1994); [15] Stams *et al.* (1994); [16] Betz *et al.* (1997); [17] Gomis-Ruth *et al.* (1996); [18] Bode *et al.* (1986); [19] Wei *et al.* (1988); [20] Navia, McKeever *et al.* (1989); [21] Walker *et al.* (1994); [22] Rano *et al.* (1997); [23] Wilson *et al.* (1996); [24] Tong *et al.* (1997); [25] Scott *et al.* (1991); [26] Wery *et al.* (1991); [27] Kitadokoro *et al.* (1998); [28] Sielecki *et al.* (1989); [29] Ghosh *et al.* (1995); [30] Breton *et al.* (1996); [31] Zhang *et al.* (1998); [32] Nam *et al.* (1996); [33] Musil *et al.* (1991); [34] Baldwin *et al.* (1993); [35] Metcalf & Fusek (1993); [36] De Bondt *et al.* (1993); [37] Russo *et al.* (1998); [38] Oefner *et al.* (1988); [39] Davies *et al.* (1990); [40] Blaber *et al.* (1996); [41] McTigue *et al.* (1999); [42] Varney *et al.* (1997); [43] Karpusas *et al.* (1997); [44] Cho *et al.* (1994); [45] Gorina & Pavletich (1996); [46] Xu *et al.* (1997); [47] Ealick *et al.* (1990); [48] DeVos *et al.* (1988); [49] Pai *et al.* (1989); [50] Kregel *et al.* (1990); [51] Scheffzek *et al.* (1997); [52] Renwick *et al.* (1998); [53] Ekstrom *et al.* (1999); [54] Schiffer *et al.* (1995); [55] Redinbo *et al.* (1998); [56] Stewart *et al.* (1998); [57] Banner *et al.* (1993); [58] Graves *et al.* (1990); [59] Priestle *et al.* (1988); [60] Schreuder *et al.* (1997); [61] Vigers *et al.* (1997); [62] Baldwin *et al.* (1991); [63] Kissinger *et al.* (1995); [64] McGrath *et al.* (1998); [65] Kallen *et al.* (1991); [66] Ke *et al.* (1991); [67] Pflugl *et al.* (1993); [68] Van Duyne, Standaert, Karplus *et al.* (1991); [69] Van Duyne, Standaert, Schreiber & Clardy (1991); [70] Van Duyne *et al.* (1993); [71] Colby *et al.* (1999); [72] Ealick *et al.* (1991); [73] Walter *et al.* (1995); [74] Zhu *et al.* (1999); [75] Futterer *et al.* (1998); [76] Meng *et al.* (1999); [77] Brandhuber *et al.* (1987); [78] Milburn *et al.* (1993); [79] Livnah *et al.* (1996); [80] Livnah *et al.* (1998); [81] Carrell *et al.* (1994); [82] Schreuder *et al.* (1994); [83] Skinner *et al.* (1997); [84] Skinner *et al.* (1998); [85] Muller *et al.* (1994); [86] Muller *et al.* (1996); [87] Banner *et al.* (1996); [88] Rao *et al.* (1995); [89] Padmanabhan *et al.* (1993); [90] Yee *et al.* (1994); [91] Mather *et al.* (1996); [92] Pratt *et al.* (1997); [93] Spraggon *et al.* (1997); [94] Mottonen *et al.* (1992); [95] Aertgeerts *et al.* (1995); [96] Xue *et al.* (1998); [97] Bode *et al.* (1989); [98] Rydel *et al.* (1990); [99] Rydel *et al.* (1994); [100] Laba *et al.* (1996); [101] Spraggon *et al.* (1995); [102] Bienkowska *et al.* (1997); [103] Huizinga *et al.* (1997); [104] Emsley *et al.* (1998); [105] Ciszak & Smith (1994); [106] Hubbard (1997); [107] Hubbard *et al.* (1994); [108] DeVos *et al.* (1992); [109] Schwabe *et al.* (1993); [110] Brzozowski *et al.* (1997); [111] Williams & Sigler (1998); [112] Somers *et al.* (1994); [113] Kannan *et al.* (1975); [114] He & Carter (1992); [115] Curry *et al.* (1998); [116] Sinning *et al.* (1993); [117] Cameron *et al.* (1995); [118] Bruns *et al.* (1999); [119] Tskovskiy *et al.* (1999); [120] Raghunathan *et al.* (1994); [121] Wilson *et al.* (1992); [122] Xie *et al.* (1998); [123] Thompson *et al.* (1997); [124] Radhakrishnan *et al.* (1996); [125] Muchmore *et al.* (1996); [126] Waksman *et al.* (1993).

molecules will shed light on some of the deepest mysteries of humanity, including memory, cognition, desire, sleep *etc.* At the same time, such structures will provide opportunities for treating those suffering from neurodegenerative diseases due to age, genetic disposition, allergies, infections, traumas and combinations thereof. Such ‘CNS protein structures’ are one of the major challenges of biomacromolecular crystallography in the 21st century.

1.3.4.4. Drug metabolism and crystallography

As soon as a drug enters the body, an elaborate machinery comes into action to eliminate this foreign and potentially harmful molecule as quickly as possible. Two steps are usually distinguished in this process: phase I metabolism, in which the drug is functionalized, and phase II metabolism, in which further conjugation with endogenous hydrophilic molecules takes place, so that excretion *via* the kidneys can occur. Whereas this ‘detoxification’ process is essential for survival, it often renders promising inhibitors useless as drug candidates. Hence, structural knowledge of the proteins involved in metabolism could have a significant impact on the drug development process.

Thus far, only the structures of a few proteins crucial for drug distribution and metabolism have been elucidated. Human serum albumin binds hundreds of different drugs with micromolar dissociation constants, thereby altering drug levels in the blood dramatically. The structure of this important carrier molecule has been solved in complex with several drug molecules and should one day allow the prediction of the affinity of new chemical entities for this carrier protein, and thereby deepen our understanding of the serum concentrations of new candidate drugs

(Carter & Ho, 1994; Curry *et al.*, 1998; Sugio *et al.*, 1999). Human oxidoreductases and hydrolases of importance in drug metabolism with known structure are: alcohol dehydrogenase (EC 1.1.1.1) (Hurley *et al.*, 1991), aldose reductase (EC 1.1.1.21) (Wilson *et al.*, 1992), glutathione reductase (NADPH) (EC 1.6.4.2) (Thieme *et al.*, 1981), catalase (EC 1.11.1.6) (Ko *et al.*, 2000), myeloperoxidase (EC 1.11.1.7) (Choi *et al.*, 1998) and beta-glucuronidase (EC 3.2.1.31) (Jain *et al.*, 1996). Recently, the first crystal structure of a mammalian cytochrome P-450, the most important class of xenobiotic metabolizing enzymes, has been reported (Williams *et al.*, 2000).

Of the conjugation enzymes, only glutathione S-transferases (EC 2.5.1.18) have been characterized structurally: A1 (Sinning *et al.*, 1993), A4-4 (Bruns *et al.*, 1999), MU-1 (Patskovsky *et al.*, 1999), MU-2 (Raghunathan *et al.*, 1994), P (Reinemer *et al.*, 1992) and THETA-2 (Rossjohn, McKinstry *et al.*, 1998). Tens of structures await elucidation in this area (Testa, 1994).

1.3.4.5. Drug manufacturing and crystallography

The development of drugs is a major undertaking and one of the hallmarks of modern societies. However, once a safe and effective therapeutic agent has been fully tested and approved, manufacturing the compound on a large scale is often the next major challenge. Truly massive quantities of penicillin and cephalosporin are produced worldwide, ranging from 2000 to 7000 tons annually (Conlon *et al.*, 1995). In the production of semi-synthetic penicillins, the enzyme penicillin acylase plays a very significant role. This enzyme catalyses the hydrolysis of penicillin into 6-aminopenicillanic acid. Its crystal structure has been elucidated (Duggleby *et al.*, 1995) and may now be used for

1. INTRODUCTION

protein-engineering studies to improve its properties for the biotechnology industry. The production of cephalosporins could benefit in a similar way from knowing the structure of cephalosporin acylase (CA), since the properties of this enzyme are not optimal for use in production plants. Therefore, the crystal structure determination of CA could provide a basis for improving the substrate specificity of CA by subsequent protein-engineering techniques. Fortunately, a first CA structure has been solved recently (Kim *et al.*, 2000), with many other structures expected to be solved essentially simultaneously. Clearly, crystallography can be not only a major player in the design and optimization of therapeutic drugs, but also in their manufacture.

1.3.5. Vaccines, immunology and crystallography

Vaccines are probably the most effective way of preventing disease. An impressive number of vaccines have been developed and many more are under development (National Institute of Allergy and Infectious Diseases, 1998). Smallpox has been eradicated thanks to a vaccine, and polio is being targeted for eradication in a worldwide effort, again using vaccination strategies. To the best of our knowledge, crystal structures of viruses, viral capsids or viral proteins have not been used in developing the currently available vaccines. However, there are projects underway that may change this.

For instance, the crystal structure of rhinovirus has resulted in the development of compounds that have potential as antiviral agents, since they stabilize the viral capsid and block, or at least delay, the uncoating step in viral cell entry (Fox *et al.*, 1986). These rhinovirus capsid-stabilizing compounds are, in a different project, being used to stabilize poliovirus particles against heat-induced denaturation in vaccines (Grant *et al.*, 1994). This approach may be applicable to other cases, although it has not yet resulted in commercially available vaccine-plus-stabilizer cocktails. However, it is fascinating to see how a drug-design project may be able to assist vaccine development in a rather unexpected manner.

Three-dimensional structural information about viruses is also being used to aid in the development of vaccines. Knowledge of the architecture of and biological functions of coat proteins has been used to select loops at viral surfaces that can be replaced with antigenic loops from other pathogens for vaccine-engineering purposes (*e.g.* Burke *et al.*, 1988; Kohara *et al.*, 1988; Martin *et al.*, 1988; Murray *et al.*, 1998; Arnold *et al.*, 1994; Resnick *et al.*, 1995; Smith *et al.*, 1998; Arnold & Arnold, 1999; Zhang, Geisler *et al.*, 1999). The design of human rhinovirus (HRV) and poliovirus chimeras has been aided by knowing the atomic structure of the viruses (Hogle *et al.*, 1985; Rossmann *et al.*, 1985; Arnold & Rossmann, 1988; Arnold & Rossmann, 1990) and detailed features of the neutralizing immunogenic sites on the virion surfaces (Sherry & Rueckert, 1985; Sherry *et al.*, 1986). In this way, one can imagine that in cases where the atomic structures of antigenic loops in 'donor' immunogens are known as well as the structure of the 'recipient' loop in the virus capsid protein, optimal loop transplantation might become possible. It is not yet known how to engineer precisely the desired three-dimensional structures and properties into macromolecules. However, libraries of macromolecules or viruses constructed using combinatorial mutagenesis can be searched to increase the likelihood of including structures with desired architecture and properties such as immunogenicity. With appropriate selection methods, the rare constructs with desired properties can be

identified and 'fished out'. Research of this type has yielded some potentially immunogenic presentations of sequences transplanted on the surface of HRV (reviewed in Arnold & Arnold, 1999). For reasons not quite fully understood, presenting multiple copies of antigens to the immune system leads to an enhanced immune response (Malik & Perham, 1997). It is conceivable that, eventually, it might even be possible for conformational epitopes consisting of multiple 'donor' loops to be grafted onto 'recipient capsids' while maintaining the integrity of the original structure. Certainly, such feats are difficult to achieve with present-day protein-engineering skills, but recent successes in protein design offer hope that this will be feasible in the not too distant future (Gordon *et al.*, 1999).

Immense efforts have been made by numerous crystallographers to unravel the structures of molecules involved in the unbelievably complex, powerful and fascinating immune system. Many of the human proteins studied are listed in Table 1.3.4.5 with, as specific highlights, the structures of immunoglobulins (Poljak *et al.*, 1973), major histocompatibility complex (MHC) molecules (Bjorkman *et al.*, 1987; Brown *et al.*, 1993; Fremont *et al.*, 1992; Bjorkman & Burmeister, 1994), T-cell receptors (TCR) and MHC:TCR complexes (Garboczi *et al.*, 1996; Garcia *et al.*, 1996), an array of cytokines and chemokines, and immune cell-specific kinases such as lck (Zhu *et al.*, 1999). This knowledge is being converted into practical applications, for instance by humanising non-human antibodies with desirable properties (Reichmann *et al.*, 1988) and by creating immunotoxins.

The interactions between chemokines and receptors, and the complicated signalling pathways within each immune cell, make it next to impossible to predict the effect of small compounds interfering with a specific protein-protein interaction in the immune system (Deller & Jones, 2000). However, great encouragement has been obtained from the discovery of the remarkable manner by which the immunosuppressor FK506 functions: this small molecule brings two proteins, FKB12 and calcineurin, together, thereby preventing T-cell activation by calcineurin. The structure of this remarkable ternary complex is known (Kissinger *et al.*, 1995). Such discoveries of unusual modes of action of therapeutic compounds are the foundation for new concepts such as 'chemical dimerizers' to activate signalling events in cells such as apoptosis (Clackson *et al.*, 1998).

In spite of the gargantuan task ahead aimed at unravelling the cell-to-cell communication in immune action, it is unavoidable that the next decades will bring us unprecedented insight into the many carefully controlled processes of the immune system. In turn, it is expected that this will lead to new therapeutics for manipulating a truly wonderful defence system in order to assist vaccines, to decrease graft rejection processes in organ transplants and to control auto-immune diseases that are likely to be playing a major role in cruelly debilitating diseases such as rheumatoid arthritis and type I diabetes.

1.3.6. Outlook and dreams

At the beginning of the 1990s, Max Perutz inspired many researchers with a passion for structure and a heart for the suffering of mankind with a fascinating book entitled *Protein Structure – New Approaches to Disease and Therapy* (Perutz, 1992). The explosion of medicinal macromolecular crystallography since then has been truly remarkable. What should we expect for the next decades?

In the realm of safe predictions we can expect the following: