Table 1.3.3.1
Crystal structures and genetic diseases

Crystal structure	Disease	Reference
Acidic fibroblast growth factor receptor	Familial Pfeiffer syndrome	[1]
Alpha-1-antitrypsin	Alpha-1-antitrypsin deficiency	[2]
Antithrombin III	Hereditary thrombophilia	[3], [4]
Arylsulfatase A	Leukodystrophy	[5]
Aspartylglucosaminidase	Aspartylglusominuria	[6]
Beta-glucuronidase	Sly syndrome	[7]
Branched-chain alpha-keto acid dehydrogenase	Maple syrup urine syndrome, type Ia	[39]
Carbonic anhydrase II	Guibaud–Vainsel syndrome, Marble brain disease	[8]
p53	Cancer	[9], [10]
Ceruloplasmin	Hypoceruloplasminemia	[11]
Complement C3	C3 complement component 3 deficiency	[12]
Cystatin B	Progressive myoclonus epilepsy	[13]
Factor VII	Factor VII deficiency	[14]
Factor VIII	Factor VIII deficiency	[40]
Factor X	Factor X deficiency (Stuart–Prower factor deficiency)	[15]
Factor XIII	Factor XIII deficiency	[16]
Fructose-1,6-bisphosphate aldolase	Fructose intolerance (fructosemia)	[41]
Gelsolin	Amyloidosis V	[17]
Growth hormone	Growth hormone deficiency	[18]
Haemochromatosis protein HFE	Hereditary haemochromatosis	[19]
Haemoglobin	Beta-thalassemia, sickle-cell anaemia	[20]
Tyrosine hydroxylase	Hereditary Parkinsonism	[21]
Hypoxanthine–guanine phosphoribosyltransferase	Lesch–Nyhan syndrome	[22]
Insulin	Hyperproinsulinemia, diabetes	[42]
Isovaleryl–coenzyme A dehydrogenase	Isovaleric acid CoA dehydrogenase deficiency	[23]
Lysosomal protective protein	Galactosialidosis	[24]
Ornithine aminotransferase	Ornithine aminotransferase deficiency	[25]
Ornithine transcarbamoylase	Ornithine transcarbamoylase deficiency	[43]
p16INK4a tumour suppressor	Cancer	[26]
Phenylalanine hydroxylase	Phenylketonuria	[27]
Plasminogen	Plasminogen deficiency	[28], [29], [30]
Protein C	Protein C deficiency	[31]
Purine nucleotide phosphorylase	Purine nucleotide phosphorylase deficiency	[32]
Serum albumin	Dysalbuminemic hyperthyroxinemia	[33]
Superoxide dismutase (Cu, Zn-dependent)	Familial amyotrophical lateral sclerosis	[34]
Thrombin	Hypoprothrombinemia, dysprothrombinemia	[35]
Transthyretrin	Amyloidosis I	[36]
Triosephosphate isomerase	Triosephosphate isomerase deficiency	[37]
Trypsinogen	Hereditary pancreatitis	[38]

References: [1] Blaber *et al.* (1996); [2] Loebermann *et al.* (1984); [3] Carrell *et al.* (1994); [4] Schreuder *et al.* (1994); [5] Lukatela *et al.* (1998); [6] Oinonen *et al.* (1995); [7] Jain *et al.* (1996); [8] Liljas *et al.* (1972); [9] Cho *et al.* (1994); [10] Gorina & Pavletich (1996); [11] Zaitseva *et al.* (1996); [12] Nagar *et al.* (1998); [13] Stubbs *et al.* (1990); [14] Banner *et al.* (1996); [15] Padmanabhan *et al.* (1993); [16] Yee *et al.* (1994); [17] McLaughlin *et al.* (1993); [18] DeVos *et al.* (1992); [19] Lebron *et al.* (1998); [20] Harrington *et al.* (1997); [21] Goodwill *et al.* (1997); [22] Eads *et al.* (1994); [23] Tiffany *et al.* (1997); [24] Rudenko *et al.* (1995); [25] Shah *et al.* (1997); [26] Russo *et al.* (1998); [27] Erlandsen *et al.* (1997); [28] Mulichak *et al.* (1991); [29] Mathews *et al.* (1996); [30] Chang, Mochalkin *et al.* (1998); [31] Mather *et al.* (1996); [32] Ealick *et al.* (1990); [33] He & Carter (1992); [34] Parge *et al.* (1992); [35] Bode *et al.* (1989); [36] Blake *et al.* (1978); [37] Mande *et al.* (1994); [38] Gaboriaud *et al.* (1996); [39] Ævarsson *et al.* (2000); [40] Pratt *et al.* (1999); [41] Gamblin *et al.* (1990); [42] Bentley *et al.* (1976); [43] Shi *et al.* (1998).

major infectious agents will be completed five years hence, if not sooner. This is likely to be followed up by '*selected pathogen structural genomics*', which will provide a wealth of pathogen protein structures for the design of new pharmaceuticals and probably also for vaccines.

This overview, written in late 1999, aims to convey some feel of the current explosion of 'crystallography in medicine'. Ten, perhaps even five, years ago it might have been feasible to make an almost comprehensive list of all protein structures of potentially direct medical relevance. Today, this is virtually impossible. Here we mention only selected examples in the text with apologies to the crystallographers whose projects should also have been mentioned, and to the NMR spectroscopists and electron microscopists whose work falls outside the scope of this review. Tables 1.3.3.1 and 1.3.4.1 to 1.3.4.5 provide more information, yet do not claim to cover comprehensively this exploding field. Also, not all of the structures listed were determined with medical applications in mind, though they might be exploited for drug design one day. These tables show at the same time tremendous achievements as well as great gaps in our structural knowledge of proteins from humans and human pathogens.

1.3.3. Crystallography and genetic diseases

Presently, an immense number of genetic diseases have been characterized at the genetic level and archived in OMIM [On-line Mendelian Inheritance in Man. Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and the National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 1999. URL: http://www.ncbi.nlm.nih.gov/omim/], with many more discoveries to occur in the next decades. Biomolecular crystallography has been very successful in explaining the cause of numerous genetic diseases at the atomic level. The stories of sickle cell anaemia, thalassemias and other deficiencies of haemoglobin set the stage (Dickerson & Geis, 1983), followed by numerous other examples (Table 1.3.3.1). Given the frequent occurrence of mutations in humans, it is likely that for virtually every structure of a human protein, a number of genetic diseases can be rationalized at the atomic level. Two investigations from the authors' laboratory may serve as examples:

(i) The severity of various cases of galactosialidosis – a lysosomal storage disease – could be related to the predicted

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

(ii) The modification of Tyr393α 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 Glu6βVal 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 α1-antitrypsin deficiency, which leads to cirrhosis and emphysema. Normal α1-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 α1-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 antihypertensive 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