International Tables for Crystallography (2006). Vol. F, Section 1.3.2, pp. 10–11.

1.3. Macromolecular crystallography and medicine

BY W. G. J. HOL AND C. L. M. J. VERLINDE

1.3.1. Introduction

In the last hundred years, crystallography has contributed immensely to the expansion of our understanding of the atomic structure of matter as it extends into the three spatial dimensions in which we describe the world around us. At the beginning of this century, the first atomic arrangements in salts, minerals and low-molecular-weight organic and metallo-organic compounds were unravelled. Then, initially one by one, but presently as an avalanche, the molecules of life were revealed in full glory at the atomic level with often astonishing accuracy, beginning in the 1950s when fibre diffraction first helped to resolve the structure of DNA, later the structures of polysaccharides, fibrous proteins, muscle and filamentous viruses. Subsequently, single-crystal methods became predominant and structures solved in the 1960s included myoglobin, haemoglobin and lysozyme, all of which were heroic achievements by teams of scientists, often building their own X-ray instruments, pioneering computational methods, and improving protein purification and crystallization procedures. Quite soon thereafter, in 1978, the three-dimensional structures of the first viruses were determined at atomic resolution. Less than ten years later, the mechanisms and structures of membrane proteins started to be unravelled. Presently, somewhere between five and ten structures of proteins are solved each day, about 85% by crystallographic procedures and about 15% by NMR methods. It is quite possible that within a decade the Protein Data Bank (PDB; Bernstein *et al.*, 1977) will receive a new coordinate set for a protein, RNA or DNA crystal structure every half hour. The resolution of protein crystal structures is improving dramatically and the size of the structures tackled is sometimes enormous: a virus with over a thousand subunits has been solved at atomic resolution (Grimes *et al.*, 1995) and the structure of the ribosome is on its way (Ban *et al.*, 1999; Cate *et al.*, 1999; Clemons *et al.*, 1999).

Macromolecular crystallography, discussed here in terms of its impact on medicine, is clearly making immense strides owing to a synergism of progress in many scientific disciplines including:

(*a*) *Computer hardware and software*: providing unprecedented computer power as well as instant access to information anywhere on the planet *via* the internet.

(*b*) *Physics*: making synchrotron radiation available with a wide range of wavelengths, very narrow bandwidths and very high intensities.

(*c*) *Materials science and instrumentation*: revolutionizing X-ray intensity measurements, with currently available charge-coupled-device detectors allowing protein-data collection at synchrotrons in tens of minutes, and with pixel array detectors on the horizon which are expected to collect a complete data set from a typical protein within a few seconds.

(*d*) *Molecular biology*: allowing the cloning, overexpression and modification of genes, with almost miraculous ease in many cases, resulting in a wide variety of protein variants, thereby enabling crystallization of 'impossible' proteins.

(*e*) *Genome sequencing*: determining complete bacterial genomes in a matter of months. With several eukaryote genomes and the first animal genome already completed, and with the human genome expected to be completed to a considerable degree by 2000, protein crystallographers suddenly have an unprecedented choice of proteins to study, giving rise to the new field of structural genomics.

(*f*) *Biochemistry and biophysics*: providing a range of tools for rapid protein and nucleic acid purification by size, charge and affinity, and for characterization of samples by microsequencing, fluorescence, mass spectrometry, circular dichroism and dynamic light scattering procedures.

(*g*) *Chemistry, in particular combinatorial chemistry*: discovering by more and more sophisticated procedures high affinity inhibitors or binders to drug target proteins which are of great interest by themselves, while in addition such compounds tend to improve co-crystallization results quite significantly.

(*h*) *Crystallography itself*: constantly developing new tools including direct methods, multi-wavelength anomalous-dispersion phasing techniques, maximum-likelihood procedures in phase calculation and coordinate refinement, interactive graphics and automatic model-building programs, density-modification methods, and the extremely important cryo-cooling techniques for protein and nucleic acid crystals, to mention only some of the major achievements in the last decade.

Numerous aspects of these developments are treated in great detail in this volume of *International Tables.*

1.3.2. Crystallography and medicine

Knowledge of accurate atomic structures of small molecules, such as vitamin B_{12}, steroids, folates and many others, has assisted medicinal chemists in their endeavours to modify many of these molecules for the combat of disease. The early protein crystallographers were well aware of the potential medical implication of the proteins they studied. Examples are the studies of the oxygen-carrying haemoglobin, the messenger insulin, the defending antibodies and the bacterial-cell-wall-lysing lysozyme. Yet, even by the mid-1980s, there were very few crystallographic projects which had the explicit goal of arriving at pharmaceutically active compounds (Hol, 1986). Since then, however, we have witnessed an incredible increase in the number of projects in this area with essentially every major pharmaceutical company having a protein crystallography unit, while in academia and research institutions the potential usefulness of a protein structure is often combined with the novelty of the system under investigation. In one case, the HIV protease, it might well be that, worldwide, the structure has been solved over one thousand times – in complex with hundreds of different inhibitory compounds (Vondrasek *et al.*, 1997).

Impressive as these achievements are, this seems to be only the beginning of medicinal macromolecular crystallography. The completion of the human genome project will provide an irresistible impetus for '*human structural genomics*': the determination, as rapidly and systematically as possible, of as many human protein structures as possible. The genome sequences of most 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

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).