24.4. The Biological Macromolecule Crystallization Database

BY G. L. GILLILAND, M. TUNG AND J. E. LADNER

24.4.1. Introduction

The crystallization of a biological macromolecule is the first step in determining its three-dimensional structure by X-ray crystallographic techniques. In crystallizing macromolecules, empirical procedures are used that take advantage of the knowledge gained from past successes. The solution properties of a macromolecule, determined by factors such as shape, size, conformational stability and surface complexity, directly relate to if and how it will crystallize. Usually, when a crystallization study is initiated, limited information is available on the properties of the macromolecule. Thus, a series of experiments are carried out that vary parameters such as pH, temperature, ionic strength and macromolecule concentration. The number of experiments required for success is variable. In many cases the search ends quickly either because the right choices were made early or because crystallization occurs over a broad range of conditions. Unfortunately, in other cases, a large number of experiments are required before the discovery of crystallization conditions, and in some cases no crystallization conditions are found regardless of how many experiments are performed.

After more than 50 years of experience in the production of diffraction-quality crystals, there is still no generally accepted strategy for searching for the crystal-growth parameters for a biological macromolecule. However, a number of systematic procedures and strategy suggestions have been put forth (*e.g.* McPherson, 1976; Blundell & Johnson, 1976; Carter & Carter, 1979; McPherson, 1982; Gilliland & Davies, 1984; Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996; McPherson, 1999). These and other strategies are all based on the successful experiences of the authors and of other investigators in the production of suitable crystals for diffraction studies. Most current strategies employ a version of the *fast screen* first popularized by Jancarik & Kim (1991). Fast screens are sets of experiments that use premixed solutions that have frequently produced crystals. Crystals are often found quickly in such experiments, but failure results in the need for a more general approach.

The motivation for the creation of the Biological Macromolecule Crystallization Database (BMCD) was to provide comprehensive information to facilitate the development of crystallization strategies to produce large single crystals suitable for X-ray structural investigations (Gilliland & Davies, 1984). The earlier and current versions of the BMCD (Gilliland, 1988; Gilliland & Bickham, 1990; Gilliland *et al.*, 1994, 1996) include entries for all classes of biological macromolecules for which diffraction-quality crystals have been obtained. These include proteins, protein–protein complexes, nucleic acids, nucleic acid–nucleic acid complexes, protein–nucleic acid complexes and viruses.

24.4.2. History of the BMCD

The BMCD has its roots in work that was initiated in Dr David Davies' laboratory at NIH in the late 1970s and early 1980s (Gilliland & Davies, 1984). Working on a variety of frustrating protein-crystallization problems, a large body of crystallization information was extracted from the literature. This eventually led to a systematic search of the literature and a compilation of data that included almost all of the crystallization reports of biological macromolecules available at the time. In 1983 the data, as an ASCII file, were submitted to the Protein Data Bank (Chapter 24.5) for public distribution. The data included the crystallization conditions for 1025 crystal forms of more than 616 biological macromolecules.

In 1987, with assistance from the National Institute of Standards and Technology (NIST) Standard Reference Data Program, the data were incorporated into a true database and distributed with software that made it accessible using a personal computer. The database was released to the public in 1989 as the NIST/CARB (Center for Advanced Research in Biotechnology) Biological Macromolecule Crystallization Database, version 1.0 (Gilliland, 1988). In 1990, a second version of the software and data for the PC database was released (Gilliland & Bickham, 1990), and in 1994 the BMCD began including data from crystal-growth studies carried out in microgravity (Gilliland *et al.*, 1994). Recently, the BMCD has been ported to a UNIX platform to take advantage of the development of network capabilities that give the user community access to the most recent updates and allow rapid implementation of new features and capabilities of the software (Gilliland *et al.*, 1996).

24.4.3. BMCD data

The BMCD contains both data extracted from the literature defining the macromolecules and data describing the crystallization and crystal form. Macromolecule data are included for biological macromolecules for which crystals have been obtained that are suitable for diffraction studies. Crystal entries must have unique unit-cell constants. Both macromolecule and crystal entries are assigned a four-character alphanumeric identifier, beginning with M or C for macromolecule and crystal, respectively.

Macromolecule data. Each macromolecule entry includes the name of the macromolecule and other aliases. Each entry includes biological source information that includes the common name, genus, species, tissue, cell and organelle from which the macromolecule was isolated. Attempts have also been made to include this information for recombinant proteins expressed in a foreign host. The subunit composition and molecular weight are also included. This information consists of the total number of subunits, the number of each type of distinct subunit, the total molecular weight and the molecular weight for each type of individual subunit. (A subunit of a biological macromolecule entity is defined as a part of the assembly that is associated with another part by non-covalent interactions. For example, haemoglobin has four subunits, two α-globins and two β-globins, and the two oligomeric nucleic acid strands of a double-stranded nucleic acid fragment are considered as two subunits.) A representative macromolecule entry is illustrated in Fig. 24.4.3.1.

Crystallization and crystal data. The data in each crystal entry include the crystal data, crystal morphology, the experimental details of the crystallization procedure and complete references. The crystal data include the unit-cell dimensions (a, b, c, α, β, γ), the number of molecules in the unit cell (Z), the space group and the crystal density. The crystal size and shape are given along with the diffraction quality. If crystal photographs or diffraction pictures are published, the appropriate references are indicated. The experimental details include the macromolecule concentration, the temperature, the pH, the chemical additives to the growth medium, the crystallization method and the length of time required to produce crystals of a size suitable for diffraction experiments. A description of the procedure is provided if the crystallization protocol deviates from methods that are in general use. Cross-references to two other structural biology databases, the Protein Data Bank (Chapter 24.5) and the Nucleic Acid Database (Berman *et al.*, 1992), are given if the identifiers are known. One of the crystal entries for the macromolecule entry illustrated in Fig. 24.4.3.1 is shown in Fig. 24.4.3.2.

references