Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

MPCD: a new interactive on-line crystallization data bank for screening strategies

Mathieu Charles,‡ Stéphane Veesler and Françoise Bonneté*

CRMCN–CNRS–UPR 7251, Campus de Luminy, Case 913, 13288 Marseille CEDEX 09, France

‡ Present address: URGV-INRAL/CNRS, 2 Rue Gaston Crémieux, CP5708, 91057 Evry CEDEX, France.

Correspondence e-mail: bonnete@crmcn.univ-mrs.fr

Received 3 May 2006
Accepted 17 July 2006

The Marseille Protein Crystallization Database (MPCD) is a new crystallization database, freely accessible *via* http://www.crmcn.univ-mrs.fr/mpcd/, populated with information found in the BMCD and CYCLOP. This new online database includes details about macromolecules (such as name, pI, molecular weight, number of subunits), crystallization conditions, methods and additives used in an easy-to-compare table form. It allows users to choose their own crystallization parameters, to create tables for further analysis and also to enter new proteins and crystallization conditions in order to supplement this database. As an essential tool in structural biology, this crystallization database will be highly relevant to crystal growers, when connected to the Protein Data Bank, for the crystallogenesis of a variety of structurally distinct molecules and assemblies, and to macromolecular and biomaterial researchers designing structures.

1. Introduction

In structural biology and more recently in structural genomics, high-quality resolution of macromolecule structures is required in order to understand biological structure–function relationships. About 85% of molecular structures (proteins, viruses, nucleic acids and complexes) in the Protein Data Bank (PDB; http://www.pdb.org; Berman *et al.*, 2000) are solved by X-ray crystallography. Despite the constant improvement in X-ray sources, in data collection and in data treatment, the resolution of structures is still dependent on production of diffraction-quality crystals. Indeed, although crystal-growth theories have long been developed (Boistelle & Astier, 1988), crystallization is still considered to be more empirical than physico-chemical theory-based, since it depends on biological factors (enzymatic activity, biological or conformational stability, purity *etc.*) and on an extensive number of variables (pH, temperature, additives, ionic strength *etc.*). To crystallize biological macromolecules, three different strategies can be used.

The traditional crystallization strategy used by crystallographers is either factorial screening (Carter & Carter, 1979), in which a finite number of crystallization conditions are tested, or sparse-matrix screening, in which a wide variety of solution conditions, *i.e.* salt type and concentration, pH, temperature, polymer percentage and molecular weight, are varied by trial-and-error methods (Jancarik & Kim, 1991; see, for example, Hampton kits, JBScreen *etc.*). These combination strategies have usually been used with high-throughput screening robots since the advent of structural genomics, allowing crystallographers to screen thousands of crystallization conditions.

Since the 1990s, an alternative strategy to trial-and-error methods has been developed: the colloidal approach, which studies the influence of solvent on weak molecular interaction forces between macromolecules in solution by scattering techniques prior to crystallization (George & Wilson, 1994; Muschol & Rosenberger, 1995). Indeed, a close correlation between a decrease in solubility and an increase in attractive interactions has been shown with various proteins as a function of the addition of salt or PEG (Ducruix *et al.*, 1996; Bonneté *et al.*, 1999; Vivarès & Bonneté, 2002). In addition, the effects of different parameters (salt, type and concentration, pH, temperature, addition of polymer *etc.*) on the interaction potentials between macromolecules in solution have been analyzed and some general trends have been highlighted (Piazza & Pierno, 2000; Tardieu *et al.*, 2002; Finet *et al.*, 2003).

Finally, a bioinformatic strategy can be used, although bioinformatics is dedicated more to the analysis of biological information, genetic sequences and protein structures than to the analysis of crystallization data. Indeed, for the three past decades, the number of solved structures has increased exponentially, reaching more than 37 000 structures in 2006. Even though it represents only a small fraction of the number of already sequenced genes (the SWISS-PROT database contains more than 180 000 entries) and of those available in the future from genomic programs, it makes possible the collection of numerous crystallization data (literature, laboratory or electronic notebooks, crystallization databases) that can be analyzed. In 1988, Gilliland and collaborators constructed a crystallization database from successful trials published in the scientific literature (Gilliland, 1988) and regularly updated it until 1997. As Gilliland wrote in 1994, 'the primary motivation for creating the BMCD was to develop crystallization strategies' (Gilliland, 1994). From this database, several computational strategies were adopted for a better understanding of crystallization. For example, statistical or empirical analyses of crystallization parameters (pH, temperature, molecular weight, macromolecular concentration, precipitant type and crystallization methods) were performed to uncover trends useful in the crystallization of new macromolecules (Samudzi *et al.*, 1992; Hennessy *et al.*, 1994). Other analyses based on knowledge of the macromolecule to be crystallized showed the existence of correlations between families of macromolecules and crystallization conditions (Hennessy *et al.*, 2000). This result suggested that certain regions of the parameter space of crystallization conditions were more likely to lead to successful crystallization than others. Similarly, in another recent article, Kantardjieff & Rupp (2004) have described the protein isoelectric point as a predictor to optimize the efficiency of crystallization screening, the relation between pH of crystallization and pI of the macromolecule being used as prior information. Nevertheless, systematic analysis of the BMCD, as it was conceived, is difficult. To extract general trends for crystallization strategies, we need a database in table form with homogeneous information. To do that, we compile crystallization conditions extracted from two crystallization databases, the IBS Conditions Yielding to Crystallization Of Proteins Database (CYCLOP) and the NASA Biological Macromolecules Crystallization Database (BMCD v.2.0; Gilliland, 1994), in a new database. This new database, the Marseille Protein Crystallization Database (MPCD), which is accessible online at http://www.crmcn.univ-mrs.fr/mpcd/, contains 5376 crystal entries.

2. Methods

2.1. Compilation of crystallization data

The main difficulty in analyzing crystallization databases is obtaining homogeneous and good-quality crystallization data. *i.e.* complete and precise with a maximum of parameters. The two databases CYCLOP (http://www.ibs.fr/ext/cyclop/index.html; unfortunately no longer available online) and BMCD (http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html; not updated since 1997) were designed differently and were not directly usable for our purposes. For instance, CYCLOP used a query constructor, enabling a limited number of queries to be executed. Many parameters could be chosen, such as the name of the molecule, the pH of crystallization, the additives used classified by type or the year of publication. The result of each query was represented in a table, but limited to 100 lines. It took 33 queries to obtain all of the crystallization data and then the tables could be parsed and imported into *Microsoft Excel* for analysis. CYCLOP thus provided 10 581 crystallization entries corresponding to 3157 crystallization conditions of 1726 different proteins.

The Biological Macromolecule Crystallization Database (BMCD) is not an online database. It is a record of successful experimental conditions, which contains 3547 crystal data entries and crystallization conditions compiled from 2526 biological macromolecules from scientific literature. It was more difficult to obtain data in table form from BMCD. Unlike CYCLOP, it was not possible to execute queries and generate tables with the BMCD database directly from the website. The website pages which show the data are script-generated, which means that the html code of the page is not accessible. We could not use, as we did for CYCLOP, a parser to scan pages and to obtain directly all data in a table form. Fortunately, free software (*XtalGrow*, available from http://www.xtal.pitt.edu) gave us access to a 'light' version of the BMCD v.2.0 database in which some data were missing, but which could be exported in table form. Missing data were manually completed from the original BMCD database. Finally, we created a database close to the original BMCD v.2.0, which contains 5892 crystallization entries corresponding to 2220 crystallization conditions of 1485 different proteins.

Neither the CYCLOP nor BMCD data are homogenous. In the CYCLOP database, one entry corresponds to one crystallization trial, whereas in the BMCD database, one entry corresponded to one additive used in a crystallization trial. This means that one entry in CYCLOP can be equal to four entries in BMCD. Since the BMCD classification is more flexible to use, even though the database is larger, we converted all CYCLOP entries into BMCD-type entries. Thus,

our new crystallization database, a combination of CYCLOP and BMCD v.2.0, contains 5376 crystallization conditions (14 872 crystallization entries) for 2699 different molecules.

2.2. Reorganization of the database

In order to facilitate further analysis of our database, standardization of the data format was necessary (standard data labelling, uniform data units). Most of the changes were to additives. For example, no unit of measurement was given for many CYCLOP concentration entries, so they had to be recovered from articles linked to each entry, and not all units were homogeneous (*e.g.* ammonium sulfate expressed either in % or in molarity). Most (83%) of both the CYCLOP and BMCD concentrations expressed in molarity were lower than 1 *M*. For a better distribution of concentrations and to avoid very small values of concentration (10^{-6}), all molar concentrations were converted into millimolar concentrations. Moreover, some important information about protein buffer was contained in a text field in the CYCLOP database. This text data had to be transformed into additive entries (1607 entries) and protein buffer pH entries (669 entries) by using a parser program, which transformed HTML pages into integrated data in MySQL format (Fig. 1).

Other inconsistencies were found in additive names, which made database analysis difficult. For example for the same additive different names (*e.g.* NaCl/sodium chloride; dimethylsulfoxide/dimethyl sulfoxide/dimethylsulfoxide) or syntax errors were found. After correction, the 956 different additive names were reduced to 662 different additive names. Moreover, in CYCLOP all numeric data (concentration, pH, temperature) were expressed as value1, value2, whereas in BMCD all numeric data were expressed as maximum value, minimum value. In each database, '0' meant that the value was unavailable and must not be taken as real information.

We have standardized all numeric data to the BMCD format (maximum value and minimum value). When only one value was available, we fixed maximum value = minimum value. Hence, our new database contains both information on molecules (name, molecular weight, number of subunits, pI and any references available) and information on crystallization conditions (method, temperature, pH, nature and concentration of additives).

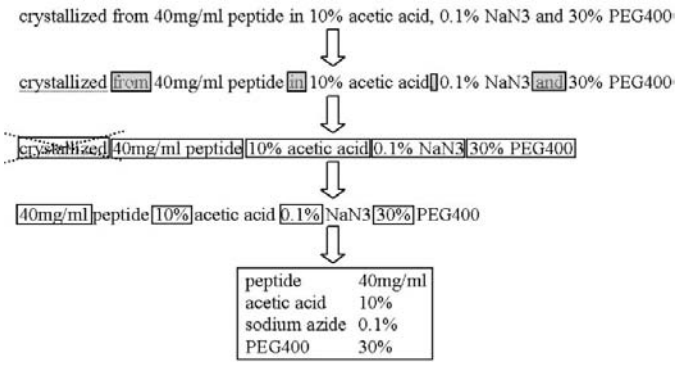

Figure 1
Keyword analysis to extract crystallization information for integration into the crystallization database.

2.3. Missing key data

Both databases contained very few experimental isoelectric points, a parameter which is rarely measured. Its theoretical value can be computed from the PDB (Protein Data Bank) or the ExPaSy server, if the sequence is known. CYCLOP and BMCD initially had only 121 and 23 experimental pI values, respectively. We supplemented our database with the experimental pI from SWISS-PROT and the theoretical pI calculated from the sequence to obtain 373 macromolecule pI values, after having verified that, for the same molecule, experimental pI values and theoretical pI values were in agreement. We limited our comparison because making the CYCLOP entries correspond with the SWISS-PROT entries was long and arduous.

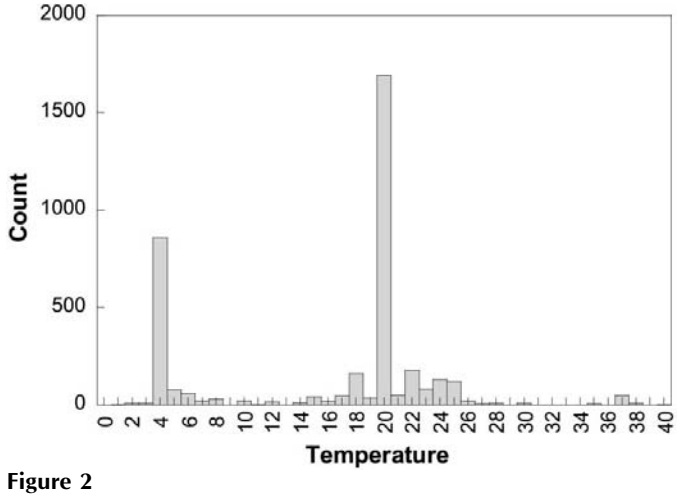

Figure 2
Distribution of temperature used for successfully crystallized macromolecules.

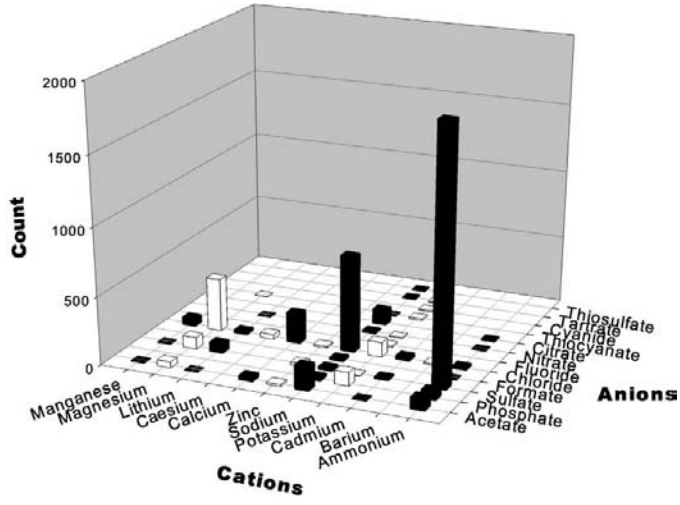

Figure 3
Distribution of anions and cations used as crystallizing agents for successfully crystallized macromolecules.

2.4. Creating a new open crystallization database: MPCD

We have therefore created a new database, the Marseille Protein Crystallization Database. In this paper, we have examined some relevant physico-chemical parameters (temperature, pH, nature and concentration of salt, nature and percentage of polymer) to show the usefulness and the capabilities of our new database. In a future paper, we will analyze the influence of different physico-chemical effects such as pH effect and salt efficiency, in order to correlate these observations with the different results obtained on potential interaction studies and crystallization mechanisms, thus establishing simple guidelines.

3. Results and discussion

The MPCD consists of 5376 successful crystallization conditions for 2699 different molecules. Various parameters are reported, such as the name of the macromolecule, its molecular weight, its isoelectric point, the concentration and the pH of the macromolecule solution, the pH of the crystallization buffer, the temperature of crystallization, the additives used (salt, polymer) and the crystallization methods. It is accessible online (http://www.crmcn.univ-mrs.fr/mpcd) for crystal growers and crystallographers interested in consulting crystallization conditions for particular macromolecules (option CONSULT) or in adding new crystallization condi-

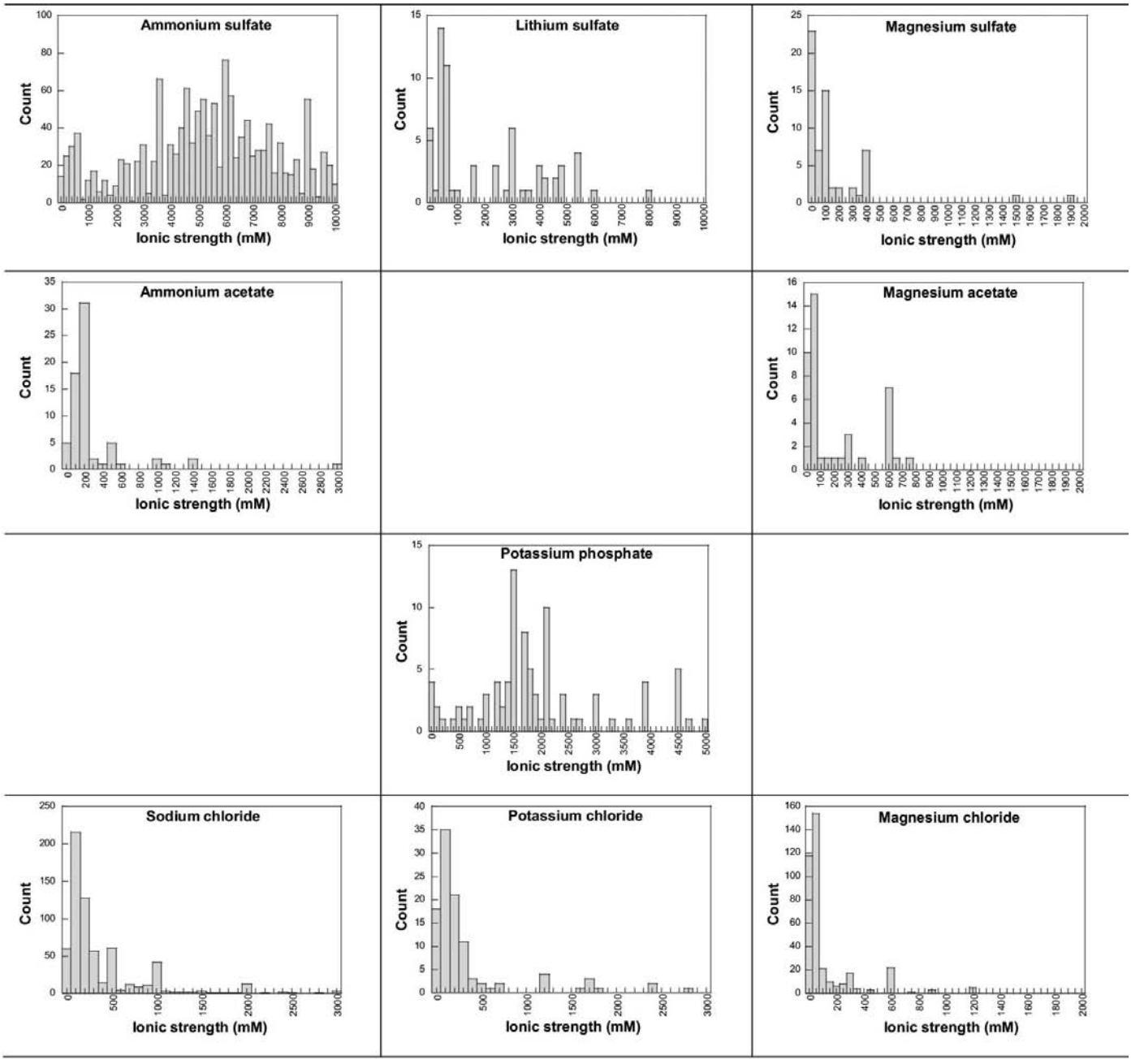

Figure 4
Histograms of ionic strengths used for successfully crystallized macromolecules as function of type of salt used.

tions for new molecules in order to complement the existing data (option CONNECT after registration). Protein and crystallization criteria are chosen by users *via* a menu-driven interface which can be partially or entirely completed. Statistical analysis can also be performed online by users. The result of their request appears in table form, which can easily be exported into *Microsoft Excel* for analysis.

3.1. Temperature effect

Of the 70% of all entries where temperature is indicated, two values clearly dominate: around 277 K (cold room) and around 293 K (room temperature) (Fig. 2). Although this tendency to crystallize only at these two temperatures, being based on experimental practices or constraints, is not the true behaviour of proteins, it reminds us that temperature is an important but under-utilized variable in biological macromolecular crystallization. Indeed, temperature influences nucleation and crystal growth by changing solubility and supersaturation of the sample (Boistelle *et al.*, 1992; Lorber & Gie;gé, 1992). It has also been shown that protein solubility is more sensitive to temperature at low ionic strength (McPherson, 1982; Riès-Kautt & Ducruix, 1992; Lafont *et al.*, 1994). The advantage of such temperature sensitivity is the ability to move around the solubility curve during nucleation and crystal growth through small changes in temperature (Jones *et al.*, 2001; Veesler *et al.*, 2004).

3.2. Precipitants

Three different types of crystallizing agents are generally used, alone or mixed: salts, polymers and alcohols. A rather limited set of salts is generally used to produce crystals, as shown in Fig. 3, because they are available in the laboratory. Approximately 45% of macromolecules in the MPCD are crystallized with salt alone (Table 1). Salt-induced precipitation (salting-out effect; Arakawa & Timasheff, 1985) is an important method frequently used as an initial step to purify proteins or to grow crystals. Ammonium sulfate is the most popular and the most widely used salt in protein purification and crystallization owing to its high solubility and its low cost. Sodium chloride is also very widely used either at low ionic strength to increase solubility (salting-in effect) or at high ionic strength to promote crystallization. Divalent metals are often used for the crystallization of various enzymes which require such metals for their activity. They can have pronounced solubility effects at very low concentrations generally up to 100 m*M* (Fig. 4). Although they are very well represented, it is difficult to affirm that these salts are the most effective salts for crystallization. They can also act as bridging species or structure-stabilizing cofactors (Trakhanov *et al.*, 1998) and facilitate protein crystallization. Moreover, the fact that crystallization occurs with one type of salt does not mean that it will not occur with another salt. Indeed, the efficiency of salts in inducing a decrease in solubility and therefore to favour crystallization has been shown to vary as a function of salt type, following the Hofmeister series (Hofmeister, 1888). This effect is well illustrated, for example, in the case of BPTI (bovine pancreatic trypsin inhibitor; Fig. 5), where it appears clear that although a high ionic strength of ammonium sulfate (about 5 *M*) is necessary to induce crystallization of BPTI, only 2 *M* sodium chloride or 0.5 *M* potassium thiocyanate is sufficient (Lafont *et al.*, 1997).

Table 1
Frequently used precipitant agents, polymers and alcohols.

(*a*) Percentage of precipitant agents used mixed or alone for the three main categories.

Non-mixed precipitant			Mixed precipitant		
Salt (%)	Polymer (%)	Alcohol (%)	Salt /polymer (%)	Salt/ alcohol (%)	Polymer/ alcohol (%)
44.7	15	8.8	20	7.6	2

(*b*) Percentage of frequently used polymers and alcohols, used either alone or mixed.

PEG 4000 (%)	PEG 6000 (%)	PEG 8000 (%)	Ethanol (%)	Glycerol (%)	2-Propanol (%)	Methanol (%)	MPD (%)
9.6	9.7	7.3	3.7	1.5	1.6	1.5	8.5

Nevertheless, in numerous cases salt alone is not able to promote crystallization and it is necessary to use polymers or organic solvents. Polyethylene glycol (PEG) is the most widely used polymer (alone at 15% or mixed with salt at 20%). PEG standards for crystallization exist in a variety of molecular weights, ranging from 200 to approximately 20 000 Da, as it is difficult to create monodisperse polymers with higher molecular weights from ethylene glycol. As Table 1 shows, most successful crystallization trials are performed with medium molecular-weight PEGs (MMW PEG) between 4000 and 8000 Da without predicting the percentage used. This is probably a consequence of the fact that in most commercial crystallization kits, MMW PEGs are the most widely used. Crystallographers know that adding water-soluble poly(ethylene glycol) is the most common way of producing protein crystals, but can we predict that these MMW PEGs will be the most effective for crystallization?

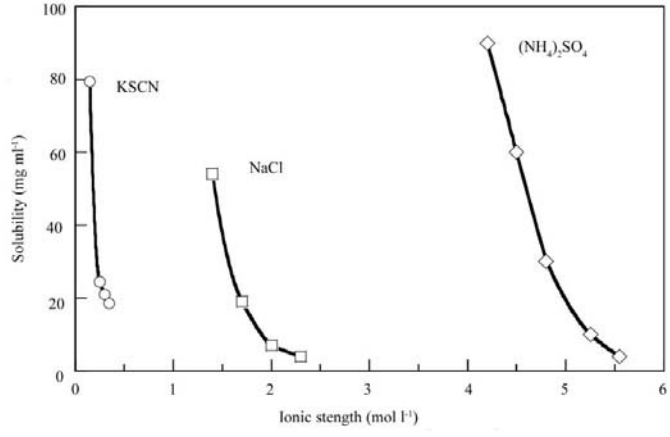

Figure 5
Comparison of BPTI solubility curves with KSCN, NaCl and $(NH_4)_2SO_4$ at pH 4.5 and 293 K.

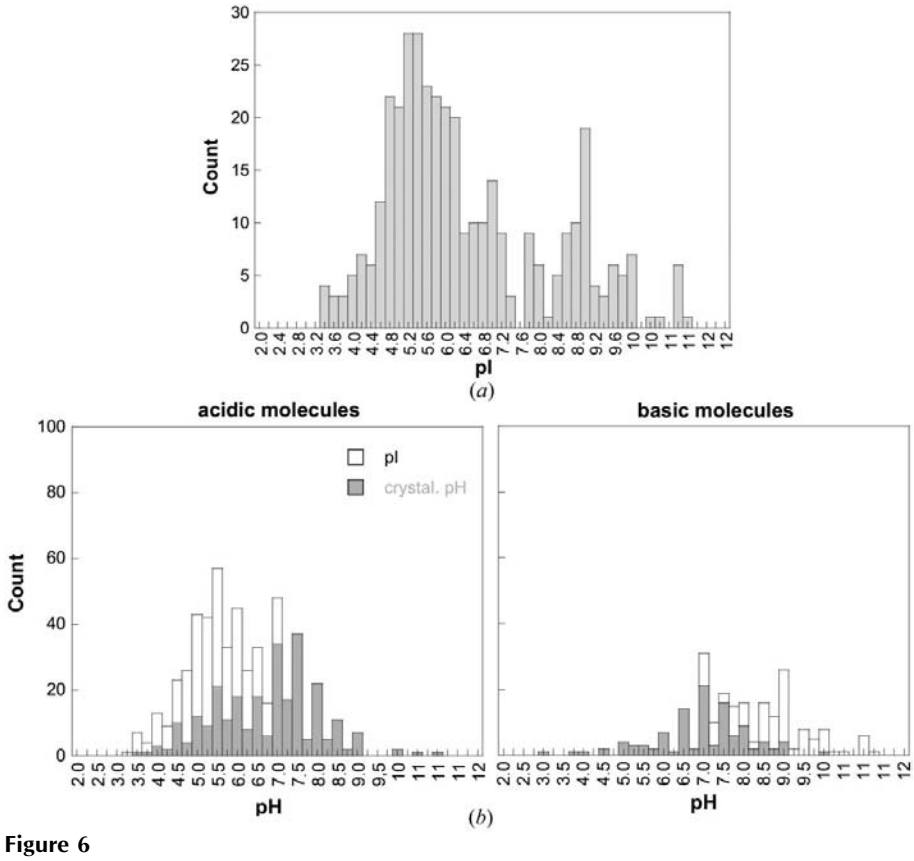

Figure 6
(*a*) Distribution of isoelectric point (pI) for successfully crystallized macromolecules; (*b*) superimposition of pH of crystallization solutions with successfully crystallized macromolecule pI for acidic and basic molecules.

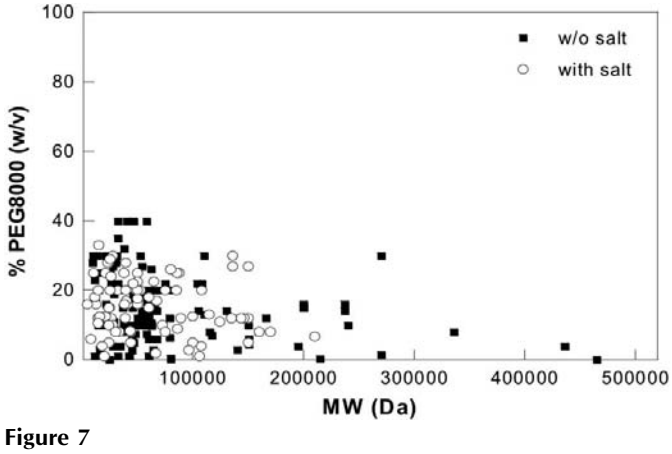

Figure 7
Counts of crystallization conditions using PEG 8000 with or without salt, as a function of crystallized macromolecule molecular weight.

In some cases, salts and polymers are not efficient for crystallization. Organic compounds such as ethanol, 2-propanol, methanol and 2-methyl 2,4-pentanediol (MPD) are other crystallizing agents, as seen in Table 1, which are less used than other agents. There have been few fundamental studies, to our knowledge, of the effect of organic solvents on the interactions in solution and on the crystallization of proteins (Anand *et al.*, 2002). Nevertheless, Boyer *et al.* (1999) showed that for ribonuclease A in ethanol attractive interactions were slightly enhanced when ethanol is added by decreasing the dielectric constant of the solvent, which directly affects electrostatic interactions. In the study of halophilic malate dehydrogenase (Costenaro *et al.*, 2001), a correlation between crystal formation and evolution of protein–protein interactions from repulsive to attractive has been shown when MPD is added, whereas glycerol induces an increase in solubility of BPTI (Farnum & Zukoski, 1999).

Some additives described in crystallization conditions, although they are not considered as crystallizing agents, are necessary for biological activity or stability (cofactors or substrates, which stabilize the quaternary structure of the protein and promote lattice packing) or necessary to avoid the oxidation of macromolecules, to chelate metal ions if metal cations prevent crystal formation *etc*. The efficiency of such additives in the crystallization process of biological macromolecules has not been demonstrated but their absence may prevent crystallization. Although a study centred on these particular additives could be considered, we will focus the present study on ordinary physico-chemical parameters.

3.3. pH effect

As a general rule, the solubility of protein is lowest at the isoelectric point (pI), since at pI the protein carries a zero net charge and therefore Coulombic repulsion is minimized (note that this is mainly true in pure water, but proteins and biological macromolecules do not like pure water). This generally leads to precipitation or aggregation of proteins rather than to a well controlled crystallization. To nucleate and grow crystals, it is necessary to reach sufficient supersaturation β ($\beta = c/s$, where c is the real protein concentration and s is the protein solubility) beyond the metastable zone. To favour nucleation events (*i.e.* nucleation frequency), it is better to increase not only supersaturation but also protein solubility (Boistelle, 1986), which can be obtained by moving the pH away from the isoelectric point. To verify this assumption, we analyzed crystallization conditions as a function of pH − pI. To do this, we needed sufficient crystallization data including isoelectric points, but we only had 369 pseudo-experimental pI values available in the MPCD. We found a bimodal frequency distribution for experimental pI values (Fig. 6*a*) with a maximum distribution for acidic proteins, which is in agree-

ment with the results of Kantardjieff & Rupp (2004) on calculated protein isoelectric points. The pH of crystallization solutions seems to be correlated with the molecule's pI: for acidic (basic) molecules, the crystallization pH is higher (lower) than the molecule's pI (Fig. 6*b*). It appears that basic molecules are more likely to crystallize for pH below their pI (from 0 to 3 pH units), whereas acidic molecules are more likely to crystallize for pH above their pI (from 0 to 3 pH units).

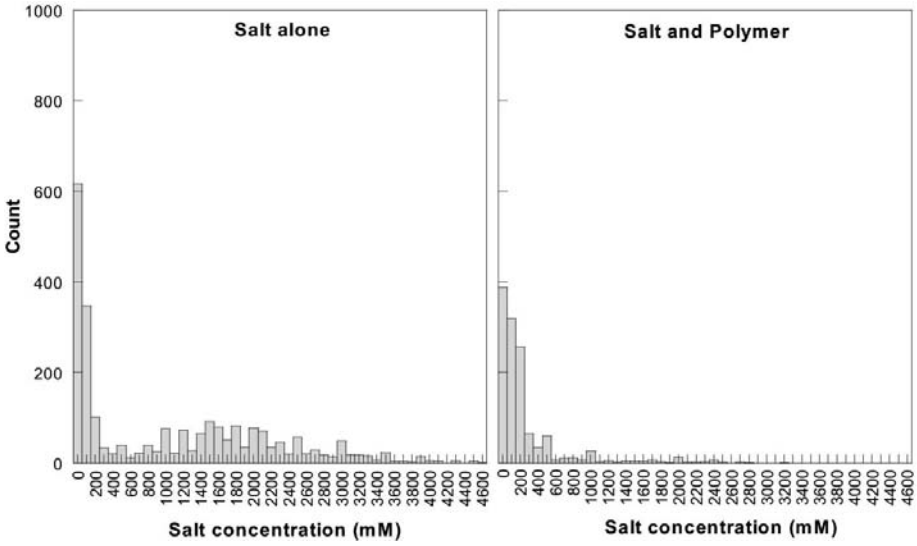

Figure 8
Distribution of salt concentrations for successfully crystallized macromolecules when salt alone is used (left) and when both polymer and salt are used (right).

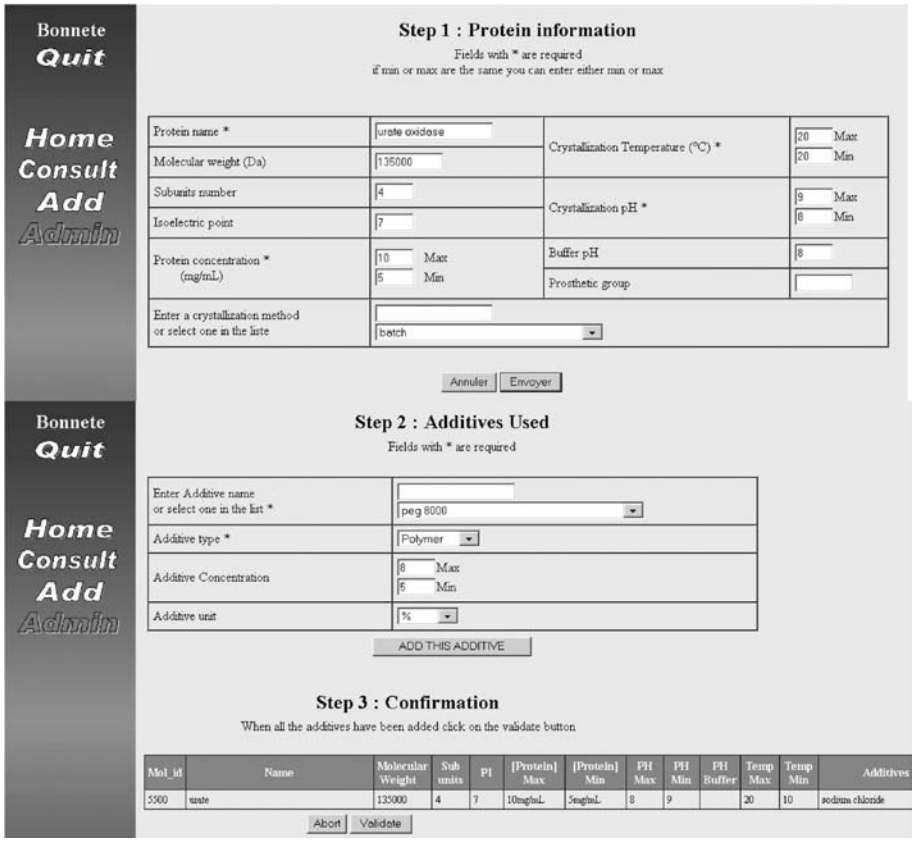

Figure 9
Registration procedure in three steps for a new crystallized macromolecule and its crystallization conditions.

Predicting crystallization-condition rules for biological macromolecules is the ultimate goal of researchers dealing with crystallization studies. Protein isoelectric points seem to be an interesting starting point, but alone do not indicate the precipitant components that should be used. Nevertheless, phase-diagram studies on basic and acidic protein solubility (Riès-Kautt & Ducruix, 1989; Carbonnaux *et al.*, 1995; Lafont *et al.*, 1997) have shown that the sign of pH − pI can be a useful guide in choosing salt type.

Although it is difficult to conclude whether the percentage of PEG used depends on the pH − pI variation, the percentage of PEG necessary to crystallize macromolecules decreases as the molecular weight increases (Fig. 7). This result has also been observed when ammonium sulfate is used as a crystallizing agent (Peat *et al.*, 2005). Moreover, when salt and polymer are both used, the concentration of salt is limited to roughly 200 m*M*, whereas when salt is used alone high concentrations of salt are often necessary to induce crystallization (Fig. 8). This means that low salt concentrations are necessary to screen the macromolecule charges when the pH is different from the pI and allows users to decrease the percentage of polymer used.

Crystallization in solution is a process which depends both on chemical properties of the macromolecules to be crystallized and on the chemical properties of the solvent in which they are dissolved. Because of successive purification and preparation stages, macromolecules are frequently stored in complex solutions whose component effects on solubility are poorly controlled. Fundamental studies on interactions in solution and crystal-growth mechanisms have generally been performed with model systems in relatively simple solutions, allowing us to understand the effect of each para-

meter on crystallization. In order to initiate a crystallization screening, it would be advisable to simplify the solvent of the purified macromolecule as far as possible (*i.e.* by maintaining biological activity and stability, macromolecule integrity *etc.*) and to concentrate the macromolecule as highly as possible (*i.e.* up to 10 mg ml^{-1}). Moreover, a basic knowledge of the solution of macromolecules (MW, pI, macromolecule stability with temperature, pH *etc*.) should make it possible to generate favourable crystallization conditions.

3.4. Future of the MPCD

The MPCD analysis has shown some interesting physicochemical parameters for the crystallization of macromolecules. Nevertheless, this analysis has also shown the lack of crystallization data for thorough analysis. BMCD was the primary source of crystallization information, but is no longer updated. Although the PDB is an obvious source for the collection of crystallographic information (Peat *et al.*, 2005), it does not contain well described crystallization data either in the 'Remark 280' field in the PDB file nor in the 'Materials and Methods Report' on the PDB website, whereas crystallization requires precise and complete information. Populating a database from the literature can rapidly be a source of mistakes. To avoid inconsistencies, data typing mistakes *etc.*, users must do it themselves. This is why we propose this new online database, which can be supplemented by users. Three steps are necessary (Fig. 9): the first step describes the macromolecule solution and the second step allows the user to add as many crystallizing agents as necessary; the third step is validation. After verification by the website administrator (syntax errors or request for complementary information, for example), the new entry is validated and appears in the database to all users.

Contrary to what is frequently stated in certain articles, there are few complete crystallization data banks available online and regularly updated. This is why we decided to create the Marseille Protein Crystallization Database, a compilation of CYCLOP (IBS source) and BMCD (from the *NIST*), which is accessible to all scientists interested in enriching the database or simply interested in the particular crystallization conditions of an already crystallized macromolecule, of a modified or mutated macromolecule whose wild-type macromolecule was already crystallized or of homologous sequences *etc*. The advantage of having such an interactive database would be enhanced if it were accessible *via* the PDB to give crystallographers more detailed crystallization conditions.

CRMCN is an associated laboratory of Universities Aix-Marseille II and III. We gratefully thank Majorie Sweetko for reading the manuscript. We thank F. Augier and F. Rosier for their help in the building of the database.

References

Anand, K., Pal, D. & Hilgenfeld, R. (2002). *Acta Cryst.* D**58**, 1722–1728.
Arakawa, T. & Timasheff, S. (1985). *Methods Enzymol.* **114**, 49–77.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Boistelle, R. (1986). *Adv. Nephrol.* **15**, 173–217.
Boistelle, R. & Astier, J.-P. (1988). *J. Cryst. Growth*, **90**, 14–30.
Boistelle, R., Astier, J.-P., Marchis-Mouren, G., Desseaux, V. & Haser, R. (1992). *J. Cryst. Growth*, **123**, 109–120.
Bonneté, F., Finet, S. & Tardieu, A. (1999). *J. Cryst. Growth*, **196**, 403–414.
Boyer, M., Roy, M.-O., Jullien, M., Bonnete, F. & Tardieu, A. (1999). *J. Cryst. Growth*, **196**, 185–192.
Carbonnaux, C., Riès-Kautt, M. & Ducruix, A. (1995). *Protein Sci.* **4**, 2123–2128.
Carter, C. W. Jr & Carter, C. W. (1979). *J. Biol. Chem.* **254**, 12219–12223.
Costenaro, L., Zaccai, G. & Ebel, C. (2001). *J. Cryst. Growth*, **232**, 102–113.
Ducruix, A., Guilloteau, J. P., Riès-Kautt, M. & Tardieu, A. (1996). *J. Cryst. Growth*, **168**, 28–39.
Farnum, M. & Zukoski, C. (1999). *Biophys. J.* **76**, 2716–2726.
Finet, S., Vivarès, D., Bonneté, F. & Tardieu, A. (2003). *Methods Enzymol.* **368**, 105–129.
George, A. & Wilson, W. W. (1994). *Acta Cryst.* D**50**, 361–365.
Gilliland, G. L. (1988). *J. Cryst. Growth*, **90**, 51–59.
Gilliland, G. L. (1994). *Acta Cryst.* D**50**, 408–413.
Hennessy, D., Buchanan, B., Subramanian, D., Wilkosz, P. A. & Rosenberg, J. M. (2000). *Acta Cryst.* D**56**, 817–827.
Hennessy, D., Gopalakrishnan, V., Buchanan, B. G., Rosenberg, J. M. & Subramanian, D. (1994). *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 179–187.
Hofmeister, F. (1888). *Arch. Exp. Pathol. Pharmakol.* **24**, 247–260.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, W. F., Wiencek, J. M. & Darcy, P. A. (2001). *J. Cryst. Growth*, **232**, 221–228.
Kantardjieff, K. A. & Rupp, B. (2004). *Bioinformatics*, **20**, 2162–2168.
Lafont, S., Veesler, S., Astier, J. P. & Boistelle, R. (1994). *J. Cryst. Growth*, **143**, 249–255.
Lafont, S., Veesler, S., Astier, J. P. & Boistelle, R. (1997). *J. Cryst. Growth*, **173**, 132–140.
Lorber, B. & Giegé, R. (1992). *J. Cryst. Growth*, **122**, 168–175.
McPherson, A. (1982). *Preparation and Analysis of Protein Crystals.* New York: Wiley.
Muschol, M. & Rosenberger, F. (1995). *J. Chem. Phys.* **103**, 10424–10432.
Peat, T. S., Christopher, J. A. & Newman, J. (2005). *Acta Cryst.* D**61**, 1662–1669.
Piazza, R. & Pierno, M. (2000). *J. Phys. Condens. Matter*, **12**, A443–A449.
Riès-Kautt, M. & Ducruix, A. (1989). *J. Biol. Chem.* **264**, 745–748.
Riès-Kautt, M. & Ducruix, A. (1992). *Crystallization of Nucleic Acids and Proteins*, pp. 195–218. New York: IRL Press.
Samudzi, C. T., Fivash, M. J. & Rosenberg, J. M. (1992). *J. Cryst. Growth*, **123**, 47–58.
Tardieu, A., Bonneté, F., Finet, S. & Vivarès, D. (2002). *Acta Cryst.* D**58**, 1549–1553.
Trakhanov, S., Kreimer, D. I., Parkin, S., Ames, G. F. L. & Rupp, B. (1998). *Protein Sci.* **7**, 600–604.
Veesler, S., Ferté, N., Costes, M.-S., Czjzek, M. & Astier, J.-P. (2004). *Cryst. Growth Des.* **4**, 1137–1141.
Vivarès, D. & Bonneté, F. (2002). *Acta Cryst.* D**58**, 472–479.