Systematic Use of the Incomplete Factorial Approach in the Design of Protein Crystallization Experiments*

(Received for publication, February 8, 1991)

Chantal Abergel§, Maxime Moulard§, Hervé Moreau§, Erwann Loret§, Christian Cambillau§, and Juan C. Fontecilla-Camps**

From the Laboratoire de Cristallographie et de Cristallisation des Macromolécules Biologiques, URA 1296-Centre National de la Recherche Scientifique, Faculté de Médecine Secteur-Nord, Bd Pierre Dramard, 13326 Marseille Cedex 15, France, the Centre de Biochimie et de Biologie Moléculaire, Centre National de la Recherche Scientifique, 31 chemin Joseph Aiguier, 13402 Marseille Cedex, France, the Laboratoire de Membranes Biologiques et Toxines Animales, URA 1179-Centre National de la Recherche Scientifique, U 172-Institut National de la Santé et de la Recherche Médicale, Faculté de Médecine Secteur-Nord, Bd Pierre Dramard, 13326 Marseille Cedex 15, France, and the **Laboratoire de Cristallogénèse et de Cristallographie des Protéines, Direction des Etudes du Vivant/Laboratoire d’Ingenierie des Protéines, Centre d’Etudes Nucleaires de Grenoble, 85X 38041 Grenoble Cedex, France.

We have used the Incomplete Factorial Approach (Carter, C. W., and Carter, C. W., Jr. (1976) J. Biol. Chem. 251, 5193–5195) in conjunction with the program Cristal (Roussel, A., Serre, L., Frey, M., and Fontecilla-Camps, J. (1990) J. Crystal Growth 106, 405–409) to crystallize six different proteins. We were able to obtain crystals and to identify the critical factors for crystallization for each of these six proteins. In some of the cases, we succeeded on the first try while using only minute amounts of protein. This study proves that the Incomplete Factorial Approach is a powerful tool in identifying the factors that need to be varied to achieve crystallization. Single crystals of adequate size were obtained for all the proteins reported here, although some did not diffract well enough to be studied by x-ray diffraction methods; the symmetric units of these latter crystals contain a large number of molecules, which is most likely due to the presence of significant amounts of carbohydrate in the proteins.

In 1979, Carter and Carter (1979) demonstrated that the application of Incomplete Factorial (IF) Approach to protein crystallization could reduce substantially the number of crystallization trials. Their method, which is based on a factorial approach to experimental design (Fisher, 1942), permits the assembly of a large number of crystallization conditions with as few experiments as possible. This is accomplished by varying more than one factor at a time in a given experiment; this saves material, and from the analysis of the results, it is possible to readily determine the factors that are critical for crystallization. The study can be viewed as being conducted in a multidimensional space in which each dimension is associated with a given factor. Each factor is then a variable which can assume more than one value. These values are then the states which define the coordinates for each experiment involving the given factor. The experimental space is then filled randomly such that the individual crystallization setups are as different from each other as possible and that the various factors and their states are represented in a balanced fashion, i.e. with similar frequencies.

Protein solubility depends on many different factors; the result of a solubility analysis represents the combined effects of these different factors. Consequently, the effect of each individual factor may be masked by the presence of the others. To obtain information on the effect of each factor, it may be necessary to perform a very large number of experiments where each parameter is varied in turn while the others are kept constant. The factorial approach is a means to reduce this number, while still allowing access to this information.

Carter and co-workers (Carter and Carter, 1979; Betts et al., 1989; Carter et al., 1988; Carter, 1990; Bell et al., 1991), in their application of the factorial approach to the crystallization of a number of proteins, used as variables the pH of the solution, the temperature, the presence of cofactors (e.g. ions, detergents, substrates, ligands, etc.) and the presence of various precipitating agents (e.g. 2-methyl-2,4-pentanediol, PEG, salts, alcohols, etc.), and they succeeded in identifying the factors critical to the production of crystals suitable for x-ray analysis. Furthermore, Carter and co-workers (1988) have

*The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

§To whom correspondence should be addressed: Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892. Tel.: 301-402-1780; Fax: 301-496-0201.

**The abbreviations used are: IF, incomplete factorial; PEG, polyethylene glycol (sometimes followed by the average molecular weight, e.g. PEG 6000); MES, morpholinethanesulfonic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazineneethanesulfonic acid; HGL, human gastric lipase; DGL, dog gastric lipase; RGL, rabbit gastric lipase; SU III, subunit III; As HIT5, A. australis Hector insect-specific toxin 2; CMV, cumulative merit value.
designed a general approach to the scoring of experimental results and the characterization of significant variables and of their levels (values) by multiple regression analysis. In spite of the positive results reported by these authors, little attention has been paid to their method (Carter and Carter, 1979; Betta et al., 1989; Carter et al., 1988; Carter, 1990; Bell et al., 1991; Lewit-Bentley et al., 1989; Jančarik and Kim, 1991; Abergel et al., 1990).

Here we describe the crystallization of a variety of proteins where the IF is extensively used. We employed this experimental design to crystallize three lipases, two elastase-like proteins, and one scorpion toxin. The experiments thus encompass a variety of proteins with different characteristics. In some of the cases that we describe, the use of available information on previous crystallization experiments (Dijkstra et al., 1978; Pasek et al., 1975; Wang and Yang, 1981; Dijkstra et al., 1982; Drenth et al., 1976; Hough et al., 1978; Tomoo et al., 1989) helped in determining the variables to be tried. This was made possible by the existence of a protein crystallization data base, originally compiled by Gilliland (1988) and recently updated by us, which can be accessed rapidly by the use of a computer program Cristal written in our laboratory (Roussel et al., 1990). By using the factorial approach in association with the program Cristal, we have been able to obtain crystals in all the cases and to identify the critical factors for each protein. In some of the cases, we succeeded on the first try while using only minute amounts of protein. This study proves that the IF method is a very powerful tool in identifying the factors that need to be varied in order to achieve crystallization.

Our use of the IF method differs from that of Carter and Carter (1979) in three respects. (a) We used the data available on the crystallization of other proteins to determine which variables to try. (b) The scoring scheme that we employed to analyze the results is much simpler than the one used by Carter and Carter, in that we use only three score values instead of their six; we give a score of 0 to the trials in which protein remains soluble, a 1 to those where precipitates are formed without regard to the type of precipitate produced, and a 2 to those where crystals are obtained. (c) In our case, the variables could be continuous, whereas Carter and Carter (1979) used discontinuous ones which allowed them to perform a proper statistical analysis.

**MATERIALS AND METHODS**

The IF protocols were established using the program INFAC kindly made available to us by C. W. Carter Jr. (University of North Carolina, Chapel Hill). The crystallization variables were defined on the basis of the known properties of the protein, e.g. its isoelectric point, optimum pH for activity, interactions with ions, solubility properties, etc., and by using the information on the crystallization conditions for related proteins, in terms of activity and sequence similarity. The information for the other proteins was accessed through the use of the program Cristal which provided the precipitating agents, additives, buffers, ions, method of crystallization, pH, temperature, protein concentration, crystal parameters, and other data for those proteins. The crystallographic information on serine proteases, for example, revealed that they have all been crystallized from ammonium sulfate; this information was used in the case of human subunit III (see below), which is an inactive elastase-like serine protease. Furthermore, the use of Cristal revealed that all phospholipases were crystallized in the presence of Mg$^2+$; accordingly, the lipase studies (see below) were conducted in the presence of this ion.

The experimental results were evaluated in the manner of Carter and Carter (1979), but the characterization of the results was simplified so that we employed only three score values instead of six: 0 for a negative result (no precipitate, no crystals), 1 for a neutral result (precipitate, no crystals), and 2 for a positive result (crystals). The effect of each variable was isolated by studying its influence on the score value throughout the whole experimental design without regard to the other variables. Then, we calculated a cumulative merit value (CMV, defined as the sum of the score values) for each level of the variable of interest, and by comparing all the CMVs of each variable, we were able to extract those critical for crystallization.

Further experiments were initiated on the basis of the evaluation of each parameter, and in many instances a new IF protocol was generated. The strategies followed are described in detail below. The experimental results have been compiled as tables depicting both the combination of levels of each variable in every case and the score given to that particular experiment. In certain cases it was necessary to eliminate combinations that resulted in insoluble mixtures of precipitating agents. All experiments were performed on Linbro plates using hanging drops equilibrated against 1-ml reservoir solutions (Wlodawer and Hodgson, 1975). The various proteins studied are described in Table I. The strategies for the individual proteins are included under “Results.”

**RESULTS**

**Elastase-like Proteases**

**Crystallographic properties**

| Protein | Biochemical characteristics |
|---------|-----------------------------|
| SU III bovine (Kerfelec et al., 1984) | MW: 25 800 Da, pl: 4.5 |
| Human (Moutard et al., 1989) | MW: 32 000 Da, pl: 5.5 |
| Gastric lipase | MW: 42 000 to 50 000 Da, Contains Carbohydrates of Molecular Weight: 8000 to 10 000 Da, pl: 6.8 |
| Rabbit (Moreau et al., 1988) | MW: 7 869 Da, pl: 7.0 |
| Human (Tirupathy et al., 1982) carbohydrate | 20%: pl 7.2 - 7.6, 10%: pl 7.2 - 7.5, 0%: pl 7.3 |
| As HIT2 (Loret et al., 1991) | MW: 7 869 Da, pl: 10.5 |
Table II. All the drops consisted of 2 μl of an aqueous protein solution mixed with an equal amount of the appropriate reservoir solution. The results from this initial crystallization trial, shown in Table II, demonstrated that: 1) since the only experiments where the CMV is greater than to zero are the ones in which PEG 6000 is the nonionic precipitating agent, we can conclude that this component is essential for crystallization. 2) The CMV when the salt used is NaCl is much higher than when (NH₄)₂SO₄ is used; moreover, we get crystals with NaCl, whereas, with (NH₄)₂SO₄, we get only precipitate. We can think of (NH₄)₂SO₄ as not favoring crystallization. 2) The CMV when the salt used is NaCl is much higher than when (NH₄)₂SO₄ is used; moreover, we get crystals with NaCl, whereas, with (NH₄)₂SO₄, we get only precipitate. We can think of (NH₄)₂SO₄ as not favoring crystallization. 3) The CMV when the salt used is NaCl is much higher than when (NH₄)₂SO₄ is used; moreover, we get crystals with NaCl, whereas, with (NH₄)₂SO₄, we get only precipitate. We can think of (NH₄)₂SO₄ as not favoring crystallization.

Initially search conditions were tested between pH 4.0 and 4.5, 15 and 20% (w/v) PEG 6000, 5-10% (w/v) NaCl, and 0.006 m CaCl₂. Optimal crystallization conditions for bovine SU III were obtained by mixing 5 μl of a reservoir solution, containing 0.1 m sodium acetate buffer, pH 4.0, 20% (w/v) PEG 6000, 7.5% w/v NaCl, and 0.006 m CaCl₂, with 5 μl of an aqueous protein solution at 10 mg/ml. The crystals grew to typical sizes of 0.7 x 0.3 x 0.3 mm² in a few days. They belong to the monoclinic space group P2₁, with a = 49.66 Å, b = 83.45 Å, c = 59.1 Å, and β = 97.9°. The calculated Vₐ (the ratio of the unit cell volume to the total protein content) (Matthews, 1968) is 4.5 Å³/Da for one and 2.25 Å³/Da for two molecules in the asymmetric unit. Since the latter value is within the range normally found in protein crystals (Matthews, 1968), there are very likely two molecules in the asymmetric unit. These crystals diffract to better than 1.8 Å resolution.

Crystallization of Human Subunit III—A protein similar to the bovine SU III has been isolated from human pancreatic secretions (Pascual et al., 1989). Although it is highly homologous to the bovine enzyme, it has a molecular weight of 32,000 Da and contains 6.8% carbohydrate (Moulard et al., 1989; Guy-Crotte et al., 1988). Initially, a set of 24 different conditions was used in the IF experiment (Table III). However, the combination of PEG, NaCl, and (NH₄)₂SO₄ resulted in insoluble mixtures even in the absence of the protein. Consequently, six experiments that contained this combination were eliminated leaving 18 conditions to test. The hanging drops had an initial volume of 4 μl. Due to the availability of only 1 mg of this protein, the pH search was restricted to three values that were close to the isoelectric point and similar to those used in the crystallization of both forms of the bovine subunit III. The results shown in Table III led to the following conclusions: 1) (NH₄)₂SO₄ is essential for crystallization, since when this salt is absent the score of the experiment is 0. 2) NaCl seems to have a negative effect, the CMV being higher when this salt is absent. PEG was rather ineffective, since the CMV is quite the same in the absence of PEG or in presence of a high percentage of this precipitating agent. 3) Ca²⁺ does not favor crystallization since the CMV is the same with and without this ion.

Subsequent trials were carried out in 0.1 M MES buffer, pH 5.5, in the range of 25-35% (NH₄)₂SO₄ saturated at 20 °C. The largest crystals were obtained when the reservoir solution was composed of 0.1 M MES buffer, pH 5.5, 28%-30% (NH₄)₂SO₄ saturated at 20 °C and 0.005 M CaCl₂, and when a ratio of 2

| pH | PEG 6000 | NaCl | (NH₄)₂SO₄ | CaCl₂ | Score |
|----|----------|------|-----------|-------|-------|
| 4.0 | 40% | 15% | 7.5% | 0.006 | 1 |
| 4.5 | 30% | 10% | 5% | 0.003 | 2 |

Table III: Initial search conditions for the crystallization of the human subunit III and for the human gastric lipase

Variable 1, pH 5.3, 5.5, 6.0 (0.1 M MES buffer); variable 2, precipitant 1: none, PEG 6000 15% (w/v), 20% (w/v); variable 3, precipitant 2: none, NaCl 5% (w/v), 7.5% (w/v); variable 4, precipitant 3: none, (NH₄)₂SO₄ 10% (saturated at 20 °C), (NH₄)₂SO₄ 25% (saturated at 20 °C); variable 5, final protein concentration: 10 mg/ml - 20 mg/ml; variable 6, cation²⁺: Ca²⁺ 5 mM, 0 mM, * experiments eliminated in the case of HGL.

Human SU III; † experiments eliminated in the case of HGL.
Protein Crystallization by the Incomplete Factorial Approach

µl of aqueous protein solution to 4 µl of reservoir solution was used for the initial drop. The crystals did not diffract sufficiently well to permit a crystallographic study.

Gastric Lipases

Gastric secretions contain lipases which are highly stable under acidic pH conditions; their activity is optimal between pH values of 4.0–6.0 (Volhard, 1901; Hamosh, 1984; Gargouri et al., 1989). This is in contrast with pancreatic lipases which operate most effectively between pH 8.0 and 9.0 (Verger, 1984) and lose their activity at pH <5.0. Gastric lipases are glycoproteins of 42,000–50,000 Da, which typically contain about 20% carbohydrate. A free sulphydryl group is essential for catalytic activity (Gargouri et al., 1988; Moreau et al., 1988a, 1988b). Three lipases purified from rabbit (Moreau et al., 1988b), dog, and human gastric tissues (Tipurapathy and Balasubramanian, 1982) were crystallized using the IF approach.

Crystallization of Rabbit Gastric Lipase (RGL)—Preliminary titration using various precipitating agents allowed us to determine that PEG 6000 was most effective in precipitating this protein. A set of 24 different conditions was chosen (Table IV) using the IF approach. Different ratios of protein solution to reservoir solution were used. They were 3 µl:2 µl and 4 µl:2 µl for expected final protein concentrations of 7.5 and 10 mg/ml, respectively. Evaluation of the drops led to the following conclusions: 1) crystals are obtained when using PEG, at any of the tested concentrations, and 0.6% (w/v) NaCl at pH ≥ 6.5. 2) By increasing the PEG concentration it is possible to get crystals at pH <8.5; however, although the solubility of RGL seems to decrease at acidic pH values, no crystals are obtained below pH 5.0.

The initial conditions were refined by performing experiments between pH 6.5 and 7.5, and 8–12% (w/v) PEG 6000, 0.1 M NaCl, and 0.01 M MgCl₂. The best crystals, in terms of size and state of aggregation, were obtained at pH 6.9–7.2, 0.1 M HEPES, in the presence of 11% (w/v) PEG 6000, 0.6% (w/v) NaCl, and 0.010 M MgCl₂. Crystals first appeared after 5 days and they grew to a typical size of 0.3 × 0.75 × 0.15 mm³. They belong to the trigonal class as revealed by their morphology, with a = 164 Å and c = 510 Å. The calculation of the Vₐ value indicates that there are many molecules in the asymmetric unit. The crystals diffract to 8 Å resolution.

Crystallization of Dog Gastric Lipase (DGL)—The protocol used in the crystallization of RGL was also applied to DGL (see Table IV). In fact, the results were quite similar to those obtained previously for RGL and the refined conditions for the crystallization of this lipase are also very close to those of the rabbit enzyme: 0.1 M HEPES buffer, pH 6.8, 12% (w/v) PEG 6000, 0.6% (w/v) NaCl and 0.010 M MgCl₂. The drops were formed by adding 4 µl of an aqueous protein solution at 4 mg/ml to 2 µl of the reservoir solution. The theoretical final protein concentration was 8 mg/ml. Crystals appeared after 1 week and they grew to a typical size of 1.5 × 0.3 × 0.1 mm³. They belong to the orthorhombic space group P2₁2₁2₁, with a = 182.8 Å, b = 211.2 Å, c = 97.9 Å. According to the Vₐ calculation, there should be 8 molecules in the asymmetric unit. The crystals diffract to about 4 Å resolution.

Crystallization of Human Gastric Lipase (HGL)—The first experiments were done on the fully glycosylated form of human gastric lipase. The protocol used in the case of RGL and DGL was also applied to the human enzyme (Table IV). However, the best results obtained in these experiments were showers of very small needles. Furthermore, these crystals were obtained within a very narrow range of conditions and all efforts to refine them were unsuccessful. We decided then to perform two additional IF experiments: the first contained 12 new conditions and the second contained 10 conditions which had been previously used in the crystallization of human SU III (see Table III). In fact, the results were quite similar to those contained the combination of PEG, NaCl, and (NH₄)₂SO₄, which resulted in insoluble mixtures even in the absence of the protein (Table III), were discarded. The protein concentration was 3.8 mg/ml. The drops were composed of 4 µl of aqueous protein solution and 2 µl of the appropriate reservoir solution. An additional series of experiments repeated conditions 1–15 in Table III and 8–12 in Table IV, but with equal amounts, 4 µl each, of protein and reservoir solutions. From these experiments the following conclusions were reached: 1) PEG is essential for crystallization since the CMV was always equal to zero when this precipitating agent is absent. 2) NaCl has a rather negative effect since the CMV is quite the same for the 2 concentrations of this salt and better when there is no NaCl at all. 3) (NH₄)₂SO₄ in combination with PEG, seems to give very positive results, since the only crystals observed were obtained when these two precipitating agents were combined 4) The use of β-octyl glucoside does not seem to play a role since crystals are obtained in the absence of the detergent.
A new IF protocol was devised taking these conclusions into consideration. A total of 48 different conditions were set up (Table VI). From the results of this experiment, a refined set of conditions was obtained: 0.1 M MES buffer, pH 5.6, 12% (w/v) PEG 6000, 12.5% (NH₄)₂SO₄ saturated at 20 °C, and 0.005 M KCl. The drops were formed by adding 6 µl of the protein solution to 3 µl of the reservoir solution. The expected final protein concentration was 7.6 mg/ml. Crystals appeared after 1 week. Subsequently, the addition of β-octyl glucoside led to larger and less aggregated crystals. These crystals grew to a typical size of 0.6 × 0.2 × 0.2 mm³, but were not good diffractors.

Because the crystals obtained from the fully glycosylated HGL were not useful for crystallographic studies, we decided to investigate the deglycosylated form of the enzyme. A new sample of HGL with 50% of the normal carbohydrate content was used in a new series of experiments. An IF protocol with 12 conditions was designed (Table VII). In all cases, 25 mM β-octyl glucoside and 5 mM KCl were added to the reservoir solution. The protein concentration was 4 mg/ml, and the drops were made of 4 µl of the protein solution and 2 µl of the appropriate reservoir solution. The following conclusions were reached upon examination of the different drops (Table VII): 1) when the concentration of (NH₄)₂SO₄ is >5%, no crystals are observed regardless of the levels of the other variables used. 2) When the concentration of (NH₄)₂SO₄ is equal to 5%, crystals are obtained for pH values ≥4.0 regardless of the concentration of PEG used.

Given the rather reduced number of conditions in the initial IF experiment, the refinement of the initial crystallization conditions was extended in the pH range of 5.0–8.0, 10–15% (w/v) PEG 6000, 5–15% (NH₄)₂SO₄ saturated at 20 °C, and 0.025 M β-octyl glucoside. The optimum condition was obtained when using 0.1 M HEPES, pH 7.5, 12.5% (w/v) PEG 6000, 10% (NH₄)₂SO₄ saturated at 20 °C, 0.025 M β-octyl glucoside as reservoir solution, and a final protein concentration of 8 mg/ml. The first crystals appeared after 2 weeks, and they reached typical sizes of 0.2 × 0.2 × 0.1 mm³.

Many crystals were obtained from these experiments, belonging to several different space groups and usually having many molecules in the asymmetric unit. None of these crystals were useful for crystallographic studies.

Crystrallography of Androctonus australis Hector Insect-specific Toxin 2 (AaH IT₂). In addition to α toxins, specific to mammalian sodium channels (Rochat et al., 1979), the venom of Androctonus australis Hector contains molecules which are neurotoxic to insects (Lester et al., 1982) and crustacea (Zlotkin et al., 1975). The insect-specific toxins are of potential interest in the design of insecticides devoid of side effects. AaH IT, the first toxin reported to be active in insects, was initially described by Zlotkin et al. (1971), and its amino acid sequence and disulfide cross-linking pattern are known (Darbon et al., 1982). More recently, two geographical variants of AaH IT, AaH IT₁ and AaH IT₂, have been characterized (Loret et al., 1991). A common feature among these toxins is that one of their four disulfide bridges is located differently relative to its position in the mammal-specific toxins from scorpion.

A set of 18 different conditions were chosen using the IF scheme (Table VIII). Two additional conditions were added where 0.005 M Ca²⁺ was replaced by 0.010 M Mg²⁺ at pH 5.0 buffered with 0.1 M MES, 10% (w/v) PEG 6000, 0.6% (w/v) NaCl at two protein concentrations, 7 and 14 mg/ml. In each case, 2 µl of the aqueous protein solution were mixed with an equal amount of the appropriate reservoir solution.

The crystallization scheme, shown in Table VIII, resulted in two general ranges of conditions where single AaH IT₂ crystals were obtained within 1 week. The first range, which yielded small tetragonal plates, was restricted to pH < 5.0, 20% (w/v) to 40% (w/v) PEG 6000, and 5% (w/v) to 10% (w/v) NaCl. These plates diffracted very weakly and their space group could not be determined; no further refinement of the crystallization conditions was done on this crystal form.

The second range, at pH ≥ 5.0, 10% (w/v) to 20% (w/v) PEG 6000, and 0.6–10% (w/v) NaCl, with or without divalent cation, produced rod-like crystals. The presence or absence of

| Variable | pH Conditions | PEG 6000 Conditions | (NH₄)₂SO₄ Conditions | Score |
|----------|---------------|---------------------|---------------------|-------|
| 1: PEG 6000 | 4 states | 2 states | 3 states | Protein conc 2 states | CaCl₂ 2 states | Score |
| 2 | 3 | 1 | 2 | 1 | 3 | 4 |
| 3 | 4 | 1 | 2 | 1 | 3 | 4 |
| 4 | 3 | 2 | 1 | 2 | 1 | 3 |
| 5 | 2 | 1 | 3 | 4 | 1 | 2 |
| 6 | 4 | 1 | 2 | 1 | 3 | 4 |
| 7 | 3 | 2 | 1 | 2 | 1 | 3 |
| 8 | 2 | 1 | 3 | 4 | 1 | 2 |
| 9 | 4 | 1 | 2 | 1 | 3 | 4 |
| 10 | 3 | 2 | 1 | 2 | 1 | 3 |
| 11 | 2 | 1 | 3 | 4 | 1 | 2 |
| 12 | 4 | 1 | 2 | 1 | 3 | 4 |
| 13 | 3 | 2 | 1 | 2 | 1 | 3 |
| 14 | 2 | 1 | 3 | 4 | 1 | 2 |
| 15 | 4 | 1 | 2 | 1 | 3 | 4 |
| 16 | 3 | 2 | 1 | 2 | 1 | 3 |
| 17 | 2 | 1 | 3 | 4 | 1 | 2 |
| 18 | 4 | 1 | 2 | 1 | 3 | 4 |
| 19 | 3 | 2 | 1 | 2 | 1 | 3 |
| 20 | 2 | 1 | 3 | 4 | 1 | 2 |
| 21 | 4 | 1 | 2 | 1 | 3 | 4 |
| 22 | 3 | 2 | 1 | 2 | 1 | 3 |
| 23 | 2 | 1 | 3 | 4 | 1 | 2 |
| 24 | 4 | 1 | 2 | 1 | 3 | 4 |
| 25 | 3 | 2 | 1 | 2 | 1 | 3 |
| 26 | 2 | 1 | 3 | 4 | 1 | 2 |
| 27 | 4 | 1 | 2 | 1 | 3 | 4 |
| 28 | 3 | 2 | 1 | 2 | 1 | 3 |
| 29 | 2 | 1 | 3 | 4 | 1 | 2 |
| 30 | 4 | 1 | 2 | 1 | 3 | 4 |
| 31 | 3 | 2 | 1 | 2 | 1 | 3 |
| 32 | 2 | 1 | 3 | 4 | 1 | 2 |
| 33 | 4 | 1 | 2 | 1 | 3 | 4 |
| 34 | 3 | 2 | 1 | 2 | 1 | 3 |
| 35 | 2 | 1 | 3 | 4 | 1 | 2 |
| 36 | 4 | 1 | 2 | 1 | 3 | 4 |
| 37 | 3 | 2 | 1 | 2 | 1 | 3 |
| 38 | 2 | 1 | 3 | 4 | 1 | 2 |
| 39 | 4 | 1 | 2 | 1 | 3 | 4 |
| 40 | 3 | 2 | 1 | 2 | 1 | 3 |
| 41 | 2 | 1 | 3 | 4 | 1 | 2 |
| 42 | 4 | 1 | 2 | 1 | 3 | 4 |
| 43 | 3 | 2 | 1 | 2 | 1 | 3 |
| 44 | 2 | 1 | 3 | 4 | 1 | 2 |
| 45 | 4 | 1 | 2 | 1 | 3 | 4 |
| 46 | 3 | 2 | 1 | 2 | 1 | 3 |
| 47 | 2 | 1 | 3 | 4 | 1 | 2 |
| 48 | 4 | 1 | 2 | 1 | 3 | 4 |
K+ did not have any apparent effect on the crystallization. Better crystals of this form were obtained by refining these conditions and the best crystals of this protein (0.4 mm × 0.15 mm × 0.15 mm in dimension) were produced when the drops initially contained 5 µl of each of the following solutions: (a) 0.1 M MES, pH 5.0, 20% (w/v) PEG 6000, 0.6% (w/v) NaCl, 0.010 M MgCl2 and (b) 13 mg/ml toxin in water. Precession photographs of centric zones showed that these crystals, which diffract to at least 2.5-Å resolution, belong to the orthorhombic space group P212121, with a = 66.4 Å, b = 52.5 Å, and c = 36.1 Å. By using a calculated molecular weight of 7,869, the Vm (Matthews, 1968) values are 4.0 and 2.0 Å3/Da for one and two molecules in the asymmetric unit, respectively. Since the second value falls within the Vm values commonly found in protein crystals (Matthews, 1968), there seem to be two molecules in the asymmetric unit. An account of these results has already been published (Abergel et al., 1990).

**DISCUSSION AND CONCLUSION**

From the results reported above it is clear that the use of the incomplete factorials approach is a very powerful tool for the crystallization of proteins. Because it samples the experimental space very efficiently, in the great majority of the cases, some kind of crystals are obtained in the initial trials. Another important aspect of the IF design is that all variables are eventually combined with each other. This leads to combinations of precipitating agents that seem to be more effective that when they are tested alone. An important consideration is the correct choice of conditions. In our hands, the best protocols have been those that were balanced in such a way as to produce results ranging from the two extremes in terms of protein solubility: clear solutions and precipitate. Indeed, since the combination of all the variables leads to cases in which the protein remains soluble in one extreme and where the supersaturation is too high and where we get a precipitate in the other, we can conclude that these experiments will reveal all the conditions for the crystallization to take place, at least in terms of solubility. This is why we combine the use of continuous variables, which allows us to extract information about the protein solubility with the precipitating agent of interest, with the discontinuous ones, which gives us the information about the best precipitating agent to be used for the particular protein.

An IF experiment can be designed more efficiently if the values of each variable are chosen judiciously. Several criteria can be used to achieve this: 1) the results of a preliminary titration of proteins can narrow the range of precipitant concentrations to use, 2) the consideration of factors, such as stability, optimum pH, and isoelectric point, will help in choosing the pH range to be tested, and 3) the use of previous crystallization data on related proteins can be instrumental in choosing the proper precipitants and cofactors (e.g. ions, detergents, etc). This latter has been facilitated in our case by the use of a computer program written in our laboratory which allows rapid access to an updated crystallization data base (Roussel et al., 1990). As an example, the use of Ca2+ in the crystallization of the gastric lipases was deduced from previous phospholipase crystallization protocols.

The evaluation of the IF experiment has been already discussed by Carter and Carter (Carter et al., 1988). They developed a scoring system distinguishing between six levels of results ranging from cloudy precipitates to prismatic crystals. A powerful tool to make these fine distinctions would be the dynamic light scattering method (Kam et al., 1978; Carter et al., 1988), although the required equipment is not normally available in a standard crystallographic laboratory.

We felt it would be better to use fewer levels of distinction than to make assumptions on the quality of precipitates and crystals obtained, and therefore, we set up a three-value scoring system that has proven quite effective in the evaluation of the outcome of our IF experiments. Here, a score of 2 was given to experiments that led to crystallization without regard to the quality of the crystals, and a score of 1 was given to experiments that contained a precipitate regardless of the nature of the precipitate. Our aim was not to try to discriminate between amorphous and crystalline precipitates, but rather to obtain information about the solubility of the protein and to determine the agents that would allow precipitation.

For the drops that remained clear, it was not possible either to extract any information on the precipitating agent's ability to insolubilize the protein or to determine how far the experimental conditions were from protein precipitation or crystallization. Therefore, those experiments which resulted in clear solutions were not given a weight and were scored as 0. Once the scoring was done, we calculated the CMV at each

**TABLE VIII**

*Initial search conditions for the crystallization of insect toxin AaH Tt.*

| Variable | Value | Protein conc | pH | PEG 6000 | NaCl | Ca2+ | Score |
|----------|-------|--------------|----|----------|------|------|-------|
| 1        | 1     | 7.0 mg/ml    | 5.0 | 32       | 0.0  | 0.0  | 1     |
| 2        | 1     | 14 mg/ml     | 5.0 | 32       | 0.0  | 0.0  | 0     |

**TABLE IX**

*Incomplete Factorial Approach*

| Variable | Value | Protein conc | pH | PEG 6000 | NaCl | Ca2+ | Score |
|----------|-------|--------------|----|----------|------|------|-------|
| 1        | 1     | 7.0 mg/ml    | 5.0 | 32       | 0.0  | 0.0  | 1     |
| 2        | 1     | 14 mg/ml     | 5.0 | 32       | 0.0  | 0.0  | 0     |

**TABLE IX**

*Determination of the critical factors for the crystallization of the human subunit III by looking factor by factor at the score of each experiment for each state.*

On this incomplete example from the Table III, the variation of the cumulative merit values for ammonium sulfate strongly suggests that this factor is critical for crystallization. In contrast, the pH is not critical in the range tested.

**TABLE IX**

| SCORE VALUES BY EXPERIMENT | pH Variable | Ammonium sulfate Variable |
|----------------------------|-------------|---------------------------|
| Level 1 | Level 2 | Level 3 | Level 1 | Level 2 | Level 3 |
| 5.5 | 5.5 | 6 | 0% | 19% | 25% |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 1 | 2 | 2 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 2 |
| 0 | 0 | 0 | 0 | 1 | 3 |
| 0 | 0 | 0 | 0 | 1 | 4 |
| 0 | 0 | 0 | 0 | 1 | 5 |

Cumulative merit values

| 4 | 3 | 2 | 0 | 4 | 6 |
level or value of the variables (Table IX). The levels which show the higher CMVs are then expected to lead to more positive results. Using this simple method, we were able to identify the factors critical for crystallization and design new crystallization experiments in our study of the six proteins described above.

There may be instances where the highest CMV is displayed by a level which only results in precipitates. An example will be column 1 in Table IX which should be contrasted with columns 2 and 3, which have lower CMVs but which produced crystals. The interpretation of the results presented in the first three columns of Table IX leads us to conclude that the pH values in columns 2 and 3 definitely lead to crystallization. In contrast, for the results shown in column 1, we can not make a conclusive statement for this pH value. In fact, we have two possible interpretations. First, this value will never allow crystallization. Second, levels of the other variables have resulted in supersaturation, and it may be possible by lowering the concentration of these other variables to obtain crystals. Usually, we will have two kinds of conditions: ones that are clear enough from the first IF protocol and ones that are less obvious and must be taken into account in the design of the next IF protocol to determine if they can lead to crystallization or if they will only give an amorphous precipitate.

The strategy described here can aid others in designing experiments to crystallize their own proteins. They could apply their knowledge of the properties of the protein, together with information obtained from the crystallization data base using Cristal, to decide on the variables and levels to try. Using this information as input to INFAC, they would be able to screen a large range of crystallization conditions in a minimum number of experiments. Then, by using the three-value scoring system and the CMVs, they would be able to design more focused crystallization experiments.

In many cases, the results of initial IF experiments may be clear enough to make any scoring unnecessary. Three different strategies may be followed after obtaining crystals in an initial trial depending on the crystal quality: 1) if the crystals are good enough for diffraction studies, the initial conditions can be used without further modification. 2) If the crystals are already useful but further improvement is desired (e.g., bigger size, a more suitable morphology, etc.), then a series of experiments could then be performed using the classical systematic approach, searching in the vicinity of the conditions found in the IF experiment. 3) If the crystals are not adequate for a diffraction study, a new IF experiment could be carried out, using more restricted values based on the first results. If this does not result in better crystals, more IF experiments could be performed using new variables and/or different levels for the old variables.

Our study contains examples of the general strategy outlined above. In most cases, our initial IF experiments produced crystals that could be improved by a simple systematic search around the successful experimental conditions, specifically the case of bovine and human SU III, RGL, DGL, and AaH IT

The case of the bovine and human SU III, RGL, DGL, and AaH IT, in the case of HGL, on the other hand, new IF experiments were designed in an effort to obtain better crystals; these new experiments led to large single crystals.

We have been able to produce single crystals of adequate size for all the proteins reported here. However, those of the lipases and of the human SU III did not diffract well enough to be suitable for X-ray diffraction studies. This is most likely due to the presence of significant amounts of carbohydrate in these proteins. Typically, they yield crystals where the asymmetric unit is occupied by a large number of molecules. New developments in molecular biology may help solve this problem by site-directed mutagenesis of potential glycosylation sites. Alternatively, an effort similar to that of Michel and Oesterheld, 1980 for the crystallization of membrane proteins may provide the right crystallization conditions for glycoproteins so that fewer molecules are found in the asymmetric unit.

Acknowledgments—We thank Dr. C. W. Carter Jr. for making his program INFAC available to us and Dr. E. A. Padlan for comments.

REFERENCES

Abercrombie, C., Loeve, E., Rochat, H., and Fontecilla-Camps, J. C. (1990) J. Mol. Biol. 214, 637-638
Bell, J. B., Jones, M. E., and Carter, C. W., Jr. (1991) Protein 9, 143-151
Betts, L., Frick, L., Wolfenden, R., and Carter, C. W., Jr. (1989) J. Biol. Chem. 264, 6737-6740
Cambillau, C., Kerfelec, B., Fogliozzo, E., and Chapus, C. (1986) J. Mol. Biol. 189, 709-710
Carter, C. W., Jr., and Carter, C. W., Jr. (1979) J. Biol. Chem. 254, 12219-12223
Carter, C. W., Jr., Baldwin, E. T., and Frick, L. (1988) J. Crystal Growth 90, 60-70
Carth, C. W., Jr. (1990) Methods 1, 12-24
Cox, M. J., and Weber, P. C. (1987) J. Appl. Crystallogr. 20, 366-373
Darbon, H., Zlotkin, E., Kopeyan, F., Van Rietschoten, J., and Rochat, H. (1982) Int. J. Pept. Protein Res. 20, 320-330
Dijkstra, B. W., Drenth, J., Kalk, K. H., and Vandermaelen, P. (1978) J. Mol. Biol. 124, 53-60
Dijkstra, B. W., Van Nes, G. J. H., Kalk, K. H., N. Brandenburg, P., Hol, W. G. J., and Drenth, J. (1982) Acta Crystallogr. B38, 793-795
Drenth, J., Enzing, C. M., Kalk, K. H., and Vessies, J. C. A. (1976) Nature 264, 373-374
Feher, G. (1942) The Design of Experiments, 3rd Ed., Olivier and Boyd, London
Gargouri, Y., Moreau, H., Piéroni, G., and Verger, R. (1988) J. Biol. Chem. 263, 2159-2162
Gargouri, Y., Moreau, H., and Verger, R. (1989) Biochim. Biophys. Acta 1006, 255-271
Guillard, G. (1988) J. Crystal Growth 90, 51-57
Guy-Crotte, O., Barthe, C., Basso, D., Fournet, B., and Figarella, C. (1988) Biochem. Biophys. Res. Commun. 156, 318-325
Hamosh, M. (1984) in Lipases (Borgstrom, B., and Brockman, H. L., eds) Elsevier Science Publishers B.V., Amsterdam
Hough, E., Little, C., and Jynge, K. (1978) J. Mol. Biol. 121, 567-569
Jancarik, J., and Kim, S. H. (1991) J. Appl. Crystallogr., in press
Kam, Z., Shore, H. B., and Feher, G. (1978) J. Mol. Biol. 123, 539-555
Keller, P. J., Cohen, E., and Neurath, H. (1956) J. Biol. Chem. 233, 457-463
Kerfelec, B., Chapus, C., and Puigserver, A. (1984) Biochem. Biophys. Res. Commun. 121, 162-168
Kerfelec, B., Cambillau, C., Puigserver, A., and Chapus, C. (1986), Eur. J. Biochem. 157, 531-538
Lester, D., Lazzarovi, P., Pelhate, M., and Zlotkin, E. (1982) Biochim. Biophys. Acta 701, 370-381
Lewit-Bentley, A., Dublié, S., Fourme, R., and Bodo, G. (1989) J. Mol. Biol. 210, 875-876
Loret, E., Mansuelle, F., Rochat, H., and Granier, C. (1990) Biochim. Biophys. Acta 1012, 633-640
Matthews, B. W. (1968) J. Mol. Biol. 33, 491-497
Michel, H., and Oesterheld, D. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1283-1285
Moreau, H., Gargouri, Y., Piéroni, G., and Verger, R. (1989a) FEBS Lett. 236, 383-387
Moreau, H., Gargouri, Y., Lecat, D., Junien, J. L., and Verger, R. (1989b) Biochim. Biophys. Acta 960, 286-293
Moulaud, M., Kerfelec, B., Mallet, B., and Chapus, C. (1989) FEBS Lett. 250, 166-170
Pascual, R., Burgos, F. J., Salva, M. F., Soriano, Mendez, E., and Villas, P. X. (1989) Eur. J. Biochem. 179, 609-616
Pasek, M., Keith, C., Feldman, D., and Sigler, P. B. (1975) J. Mol. Biol. 97, 395-397
Protein Crystallization by the Incomplete Factorial Approach

Puigserver, A., and Desnuelle, P. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 2442-2445
Rochat, H., Bernard, P., and Couraud, F. (1979) Adv. Cytopharmacol. 3, 325-334
Rousset, A., Serre, L., Frey, M., and Fontecilla-Camps, J. (1990) J. Crystal Growth 106, 405-409
Tirupathy, C., and Balasubramanian, K. A. (1982) Biochim. Biophys. Acta 712, 692-698
Tomoo, K., Ohishi, H., Ishida, T., Inoue, M., Ikeda, K., Aoki, Y., and Samejima, Y. (1989) J. Biol. Chem. 264, 3636-3638
Venot, N., Sciaky, M., Puigserver, A., Desnuelle, P., and Laurent, G. (1986) Eur. J. Biochem. 157, 91-99
Verger, R. (1984) in Lipases (Borgström, B., and Brockman, H. L., eds) Elsevier Science Publishers B. V., Amsterdam
Volhard, F. (1901) Z. Klin. Med. 42, 414-418
Wang, A. H-J., and Yang, C. C. (1981) J. Biol. Chem. 256, 9279-9282
Wicker, C., and Puigserver, A. (1981) FEBS Lett. 128, 13-16
Wlodawer, A. M., and Hodgson, K. O. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 398-399
Yamasaki, M., Brown, J. R., Cox, D. J., Greenahelds, R. N., Wade, R. D., and Neurath, H. (1963) Biochemistry 2, 859-866
Zlotkin, E., Martine, G., Rochat, H., and Miranda, F. (1975) Insect. Biochem. 5, 243-250
Zlotkin, E., Miranda, F., Kopeyan, G., and Lissitzky, S. (1971) Toxicon 9, 9-13