Scaffolds for protein crystallisation

Enrico A. Stura,[a*] Michael J. Taussig,[b] Brian J. Sutton,[c] Stéphane Duquerroy,[d] Stéphane Bressanelli,[d] Anthony C. Minson[e] and Felix A. Rey[d]

[a]CEA, Département d'Ingénierie des Protéines (DIEP), C.E. Saclay, 91191 Gif-sur-Yvette, France, [b]Technology Research Group, Babraham, Cambridge CB2 4AT, UK, [c]The Randall Centre, King's College London, London SE1 1UL, UK, [d]Unité de Virologie Moléculaire Structurale, Génétique des Virus, UMR 2472, CNRS-INRA, 91198 Gif-sur-Yvette, France, and [e]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK. E-mail: estura@cea.fr

In the absence of a method to ensure that crystals can be obtained for any given protein, the possibility of developing scaffolds for protein crystallisation becomes attractive. Among several approaches that could yield scaffolds, two are particularly promising: the first is based on immunoglobulin Fab fragments and immunoglobulin binding proteins while the second is based on fusion proteins. In the Fab based scaffold, the protein of interest is the antigen recognised by the antibody. In the second case, it is a protein fused to one of the scaffold components. The operational difference between the two methods is the existence of a flexible covalent tether compared to a highly specific interaction. The relative merits and disadvantages of each approach are discussed here. We also describe a lattice obtained through a combinatorial approach which appears to have the required properties to be considered a scaffold. The system makes use of an Fab derived from a rheumatoid factor and an Fc-fusion protein. The Fc-fusion system is ideal for enhanced expression of glycoproteins in mammalian cells and provides a useful tag for their purification. The molecular replacement shows a mode of binding for this rheumatoid factor that is not competitive with bacterial Fc-binding proteins. Hence it may be possible to generalise the method to include bacterial expression of fusion proteins with either protein A or protein G as the fusion partner.

Keywords: crystallisation scaffold; combinatorial crystallisation; immunoglobulin binding proteins; antibody complex crystallisation.

1. Introduction

Any form of screening, including a robotic one, is likely to yield crystals only for proteins that crystallise easily. Such proteins readily form the protein-protein interactions that are needed to stabilise a crystalline lattice. Unfortunately, a large proportion of proteins do not fall in this category. By increasing the number of trials, we reap diminishing returns so that the added effort is difficult to justify. For this reason we favour a scaffold approach to crystallisation. However, careful consideration must be given to its design, to avoid introducing a method which is no better than the one it proposes to replace. This can be avoided by maintaining a cleavage site within the linker so that the crystallisation can be carried out as always, without the fusion protein attached, but also with the fusion-protein within the scaffold system. Both approaches can be carried out in parallel.

With the availability of antibodies against the protein of interest, crystallisation can also be carried out in complex with antibodies and immunoglobulin binding proteins (Stura *et al.*, 2001*a*). The advantage of the fusion-protein scaffold method is that it is not necessary to raise and purify monoclonal antibodies for each target protein. Its disadvantage is the flexibility of the linker. With such flexibility it is not possible to ensure that the protein of interest will connect strongly to the lattice formed by scaffold proteins unless several different scaffolds systems are tried to increase the chances that among the various scaffolds one will have the desired properties. Therefore, we need to generate several scaffold systems. To reach this objective we make use of crystallisation combinatorial techniques such as stoichiometry variation screening (SVS), where the ratio of the proteins that form a multi-component system is varied and combinatorial complex crystallisation (CCC) (Stura *et al.*, 2001*a*,*b*).

The methods that have been used to obtain the potential scaffold system are described in greater detail below. In summary, the fusion protein comprises amino acids 1-318 from viral glycoprotein D (gD) from Herpes Simplex Virus 1 (HSV1). A six residue linker connects it to the natural hinge between the CH1 domain to CH2. The construct continues with the two immunoglobulin constant domains CH2-CH3 of which constitute the Fc portion of IgG (Fig. 1). gD is one of the Herpes glycoproteins responsible for receptor recognition, binding and subsequent fusion of the viral membrane with that of the target cell, a process that leads to infection. The fusion construct includes the glycosylation sites from gD and from the Fc, which explains the need for mammalian cell expression. The other component of the scaffold system is an Fab derived from a human rheumatoid factor. Rheumatoid factors are auto-antibodies that recognise the constant domain, Fc, of immunoglobulin G and hence also recognise proteins fused to an Fc. Fab derived from such auto-antibodies can potentially be used to help crystallise Fc complexes with and without a fusion protein.

```
HSV1 gD GKYALADASLKMADPNRFRGKDLPVLDQLTDPPGVRRVYHIQAGLPDPFQ  50
        PPSLPITVYYAVLERACRSVLLNAPSEAPQIVRGASEDVRKQPYNLTIAW 100
        FRMGGNCAIPITVMEYTECSYNKSLGACPIRTQPRWNYYDSFSAVSEDNL 150
        GFLMHAPAFETAGTYLRLVKINDWTEITQFILEHRAKGSCKYALPLRIPP 200
        SACLSPQAYQQGVTVDSIGMLPRFIPENQRTVAVYSLKIAGWHGPKAPYT 250
        STLLPPELSETPNATQPELAPEDPEDSALLEDPVGTVAPQIPPNWHIPSI 300
        QDAATPYHPPATPNNMGL                                 318
added linker              RLSGLA
natural Fc linker                216 EPKSCDKTHTCPPCPAPELLG 236
Igg1 H                                    GPSVFLFPPKPKD 250
        TLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNST 300
        YRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVY 350
        TLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD 400
        SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK   448
```

Figure 1 Sequence of the fusion protein gDFc.

2. Theoretical and practical considerations

The system that is envisaged consists of three stages: first the screening for a scaffold system, second understanding which components of the system can be replaced and third utilisation of the system. Each of these stages will require screening, and as the system becomes more complicated, a robotic approach will become desirable to cope with the large number of combinations of proteins that can complex together to give rise to an usable scaffold system. Once several scaffold systems have been developed, each will need to be screened with a fusion protein whose properties match those of the scaffold. The strength with which the "guest" protein can attach itself to the lattice formed by the scaffold proteins will determine how well defined the electron density for the "guest" will be. Taking into consideration these needs, we can formulate some guidelines: *(i) It is desirable that the scaffold proteins should crystallise alone to yield a lattice that can accommodate a guest protein but that do not need any contribution from such guest. (ii) The guest protein should not perturb the lattice formed by the scaffold proteins but contribute to its stability.* These conditions appear to be fulfilled to some extent

by the system presented here. The system needs improving and can be improved. One of the advantages of a scaffold system is that it can always be improved: stronger scaffold interactions, sites for heavy metals, ability to withstand detergents and most important a large degree of flexibility since "guests" will have different needs.

2.1. The problem with scaffolds

Theoretically, even in a one-dimensional case the size of the protein that can be accommodated on the scaffold is limited. In addition it must be such as not to perturb the packing of the scaffold. For example, a protein that forms a dimer could upset the correct alignment of the scaffold. In some cases such problems can be overcome through stoichiometry variation screening (Stura *et al.* 2001*b*). By allowing extra copies of the scaffold proteins, we can accommodate larger "guest" proteins. The example of the crystallisation of Fab 2A2 in complex with protein A (SpA) shows that the addition of extra scaffold proteins helps to overcome problems that arise when the added protein has a tendency to dimerise. The crystals of the Fab 2A2-SpA complex show a 3:2 stoichiometry with two Fab complexed and one free (Graille *et al.* 2000). The two SpA molecules form a dimer and force the dimerisation of two of the three Fabs in the asymmetric unit. Without the extra copy of Fab 2A2, it would not be possible to maintain the "scaffold" direction which in the complex replicates the packing found in crystals of the uncomplexed Fab (Stura *et al.* 2001*b*). Some precise molecular design would be needed to obtain the same result without the extra free Fab. The design would have to take into account the crystallographic and non-crystallographic two-fold axes. It is unfortunate that the dimer of SpA formed by two SpA molecules is not related by 180^{o} (Graille *et al.* 2000) as this non-180^{o} angle cannot be matched by any crystallographic or non-crystallographic dyad. The non-180^{o} angle is unusual, but the example illustrates clearly the problem introduced by a dimeric "guest" since the SpA dimerisation forces the Fabs complexed to it to follow the same non-180^{o} relationship. Based on a single,perhaps unusual example, one cannot state categorically that it is impossible to design a scaffold system without SVS, but SVS can definitely alleviate design failures and allow crystallisation with minimal or no specific design.

2.2. The problem with 2- and 3-dimensional scaffolds

Changing the stoichiometry alleviates the size and polymerisation problems: as the complex gets larger, the concentration of scaffold protein is increased. In the Fab 2A2-SpA example (Graille *et al.* 2000), the free Fab in the lattice of the 2A2-SpA complex packs along one lattice direction as it does in the crystal of the uncomplexed 2A2 suggesting that 1-dimensional scaffolds could be generated as the stoichiometry is varied. Unfortunately a 1-dimensional scaffold is not a "real" scaffold. To pass from a 1-dimensional scaffold to a 3-dimensional one, there must be some propensity in the system to create the other remaining lattice interactions. With the increase in the number of dimensions, the number of restrictions also increases. In a 2-dimensional case translational and rotational symmetries must be respected in both dimensions. A "guest" that does not respect such constraints is likely to disrupt the integrity of the scaffold. Given the fact that explicit design is difficult, it is best to opt for a flexible system based on screening: the "guest" protein will have to match the size of the cavity and screening will have to be carried out within a large combinatorial set of interacting proteins to find a suitable "host" scaffold lattice. For disruptive guests, the larger the size of the host lattice with respect to that of the guest, the less likely is the guest to be able to cause irremediable lattice damage. In the future it might be possible to predict accurately a number of important features regarding the structure of the guest protein and hence design a fusion protein to take advantage of such predicted properties. For example a trimeric protein could be fused to a scaffold protein that crystallises in a trigonal space group, a dimeric protein could take advantage of a 2-fold axis and a pentameric protein could be expressed in the interior of a icosahedral particle to match the 5-fold symmetry. The design must be accurate and given that at present such accuracy in prediction is beyond reach, a combinatorial approach where we screen stoichiometry, various complexes and various crystallisation conditions is more suited to the present situation. The results from the robotic screening could be used to make the method more efficient by allowing the selection of relatively small subset of "useful" crystallisation conditions. These conditions could then be combined with SVS and a large repertoire of scaffold components such as antibodies and immunoglobulin binding proteins (IBPs), to create a system that will have a high rate of success. By necessity, the system we currently employ is based on a limited precipitant subset in the initial search. Such conditions are selected on the basis of the crystallisation of the scaffold proteins used, followed by a wider search consistent with the principles highlighted in reverse screening.

3. The scaffold systems

The approach to scaffold building consists of two stages to avoid the problem of imperfect design. In the first stage we screen for crystals that may create a scaffold and in a second stage we use seeds from these crystals and look for a streak seeding response with a truncated protein or in the absence of a scaffold component. This will select for "holes" that will then to be filled with a variable "guest" protein. By including "guest" proteins which have some intrinsic heterogeneity during the first stage, there will be a tendency to select a set of scaffold proteins that can form a lattice that does not rely on crystal contacts being made by the "guests". However, the presence of "guests" during the screening ensures that sufficient space is reserved for them and that their presence is likely to strengthen the overall lattice. The heterogeneity introduced in the screening is likely to pose some problems, but at present this may be unavoidable until an alternative solution has been found. However, the chances of success in creating a scaffold system can be increased by using a large set of interacting proteins can be used in a combinatorial manner. Two potential scaffold systems are currently being considered. The first is an extension of the antibody complex crystallisation method with IBPs (Stura *et al.* 2001*a;* 2002*a*) and the second is based on fusion proteins.

3.1. The antibody system

The antibody system is essentially unchanged in principle from that described in (Stura *et al.* 2001*a*). The method has been substantially improved by the addition of several protein L (PpL) mutants with special lattice forming characteristics (Stura *et al.* 2002*a*). While the wild type binds the V_L region of immunoglobulin light chains at two distinct sites (Graille *et al.* 2001), two of the mutants bind at only one site. The mutants and the wild type with two binding sites find themselves sandwiched between two antibodies and therefore cannot contribute to the formation of lattice contacts, although they can promote crystallisation by mediating Fab dimerisation. On the other hand, the one-binding site mutants can participate in lattice contacts. For one of these mutants we have obtained a complex lattice containing four Fab fragments, four protein L molecules and four antigen binding sites. Two of these are highly accessible while the other two are usable only for a peptide antigen as the 45 residue peptide used in that crystallisation. In this lattice PpL forms a dimer. Further refinement and model building is needed to elucidate

Table 1 Data collection and refinement statistics for data from crystals of gD-Fc RF61 complexes.

Data Set	FcRF	gD-1	Hg-Eu‡	Cd-Sm‡	Sr-Pt‡
Crystallisation	*45% WS1*¶	*WS0*§	*50% WS1*¶	*50% WS1*¶	*40% WS1*¶
gDFc : RF Fab	1 : 2.4	1 : 2	1 : 2.4	1 : 2.4	1 : 2.4
SpA-domD	no	no	yes	yes	no
MPD concentration	5%	none	none	none	5%
Growth period	8 months	4 months	2 months	2 months	5 months
Data Collection	ESRF-ID14EH1	ESRF-ID14EH1	ESRF-ID14EH1	ESRF-ID14EH1	ESRF-ID14EH1
Space group	C2	C2	C2	C2	C2
Unit cell: a, b, c (Å)	242.0, 75.6, 102.4	243.6, 76.4, 102.8	256.2, 75.9, 103.0	225.4, 75.3, 102.0	242.8, 76.4, 102.8
α, β, γ (°)	90.0, 91.1, 90.0	90.0, 91.5, 90.0	90.0, 101.4, 90.0	90.0, 92.0, 90.0	90.0, 91.5, 90.0
resolution range(Å)	25.0 - 3.0	25.0 - 3.0	25.0 - 3.1	25.0 - 3.5	25.0 - 3.7
completeness (%)	97.9	100	99.9	97.1	98.9
<I/σI>	14.1	7.4	13.4	9.2	8.9
Rsym (%)	8.1	12.9	9.7	15.0	17.2
Refinement	*CNS**	*not done*	*CNS**	*CNS**	*not done*
R-free (%)	40.4†	-	40.8	47.5	-
R-work (%)	35.7†	-	35.5	38.3	-

§ *WS0:* 10% w/w MPEG 5000, 3 mM zinc acetate, 3mM $CdCl_2$, 100 mM sodium cacodylate, pH 6.5.
¶ *WS1:* 19% w/w PEG 3350, 10% v/v isopropanol, 4 mM zinc acetate, 2 mM calcium acetate, 100 mM sodium cacodylate, pH 6.5.
‡ Soak solution: 10% w/w PEG 3350, 5% v/v isopropanol, 50 mM sodium cacodylate, pH 6.5
* Model used in refinement consists of Fc, Fab and Fv.
† R-free = 41.9% ; R-work 36.2 with model composed of Fc, 1 gD, Fab and Fv.

the details of the binding at the two remaining sites (Stura *et al.* 2002*a*). This system might be useful for the crystallisation of 30 residue-tagged peptides and through screening it could accommodate a larger fusion protein. The combinatorial system based on PpL is of great interest, since about 50% of human and murine antibodies are avidly bound by this IBP. In summary, the advantage of the antibody based system is that the guest protein is the antigen against which the antibody was raised and hence is held the "guest" will be held in place with high affinity. It has the disadvantage that new antibodies, recognised by IBPs, must be generated for each guest protein. To avoid generating antibodies each time an anti-peptide antibody could be used and the tag recognised by the antibody could be fused via a short linker (or tether) to a protein of interest. This merely transforms the antibody system into a fusion protein system similar to the one described below, but the guest protein must still find suitable attachment locations within a cavity in the lattice generated by the scaffold proteins.

3.2. Using low affinity binding

A general method to create scaffolds, that could be used to study a series of mutants of the same protein, could make use of a low affinity mutant of the protein of interest. In the crystallisation of an antibody with a κ2 light chain we have added PpL, although κ2 light chains are poorly recognised by PpL. The resulting crystal lattice has space for PpL, although there is no electron density for PpL at the appropriate position (Stura *et al.* 2002*b*). The result shows that a low affinity mutant can be used to reserve space in the lattice which could later be used by a series of higher affinity mutants of the same protein. The space is created because the low affinity protein cannot be present in the lattice dependably. Because of the inconsistent presence of the low affinity protein, no scaffold crystal contacts will therefore depend on it. However, the higher affinity mutants will then be able to make use of the reserved space when co-crystallised with the other elements of the system. The example, detailed here, was carried out as a pilot study to establish a proof of principle, but is of no practical utility. A low affinity PpL mutant is currently being designed to be used with an antibody capable of binding strongly wild type PpL. This could results in a scaffold for PpL mutants. This approach can be combined both with the antibody based scaffold system and the fusion protein system.

3.3. The Fc-fusion system

In the Fc-fusion system that we are using, the fusion protein comprises amino acids 1-318 from viral glycoprotein D (gD) from Herpes Simplex Virus 1 (HSV1) fused at the N-*terminus* of the dimeric immunoglobulin constant CH2-CH3 domain fragment (Fig. 1; Fig. 2*d*) in Chinese hamster ovary cells (CHO). The presence of Fc enhances expression and protein production in this mammalian system. Expression in CHO cells is compatible with the crystallisation of glycoproteins (Stura *et al.* 1992). Since crystals could not be obtained for the fusion protein in isolation, the crystallisation was pursued with the entire fusion protein in complex with rheumatoid factors.

4. Materials and methods

4.1. Crystallisation and data collection

The Fabs used in the screening were 2A2 (Graille *et al.* 2000), RF61 (Harindranath *et al.* 1991) , SJ2 (Randen *et al.* 1993) and B20 (Zhang *et al.* 1998) from IgM rheumatoid factors. B20 is from EBV-transformed peripheral B cells from RA patient; 2A2 is a hybridoma from synovial B cells. The Fabs were produced by trypsin cleavage of the IgM secreted by a hybridoma created from synovial B cells of rheumatoid arthritis patients. gDFc was expressed in CHO cells. By gDFc we denote the molecule composed of an Fc dimer, where each monomer has a fused viral gD glycoprotein (i.e. gD-Fc-Fc-gD). The protein was de-glycosylated with PNGase F from *Flavobacterium meningosepticum* (Boerhinger), 1 U per 0.4 mg protein, 24 h at 37°C in 150 mM NaCl, 50 mM Tris-HCl pH 7.0, and then concentrated to 9 mg/ml exchanging the buffer to 50 mM ammonium acetate, pH 6.5. The four rheumatoid factors Fabs were screened simultaneously in complex with gDFc. gDFc alone was also set up as a control. The screening was done in parallel as described in (Stura, 2001). Only four precipitant conditions were set up in the initial screening. All four conditions included high molecular weight polyethylene glycol (PEG) and zinc, consistent with Fc crystallisation conditions obtained from prior screening with Fab 2A2 with an excess of Fc. Crystals were obtained from one of the four wells set up and with two of the RF Fabs. Only one of these could be pursued, since only a limited amount of protein was available for the other monoclonal antibody. Different stoichiometry of gDFc and RF Fabs were screened in a second

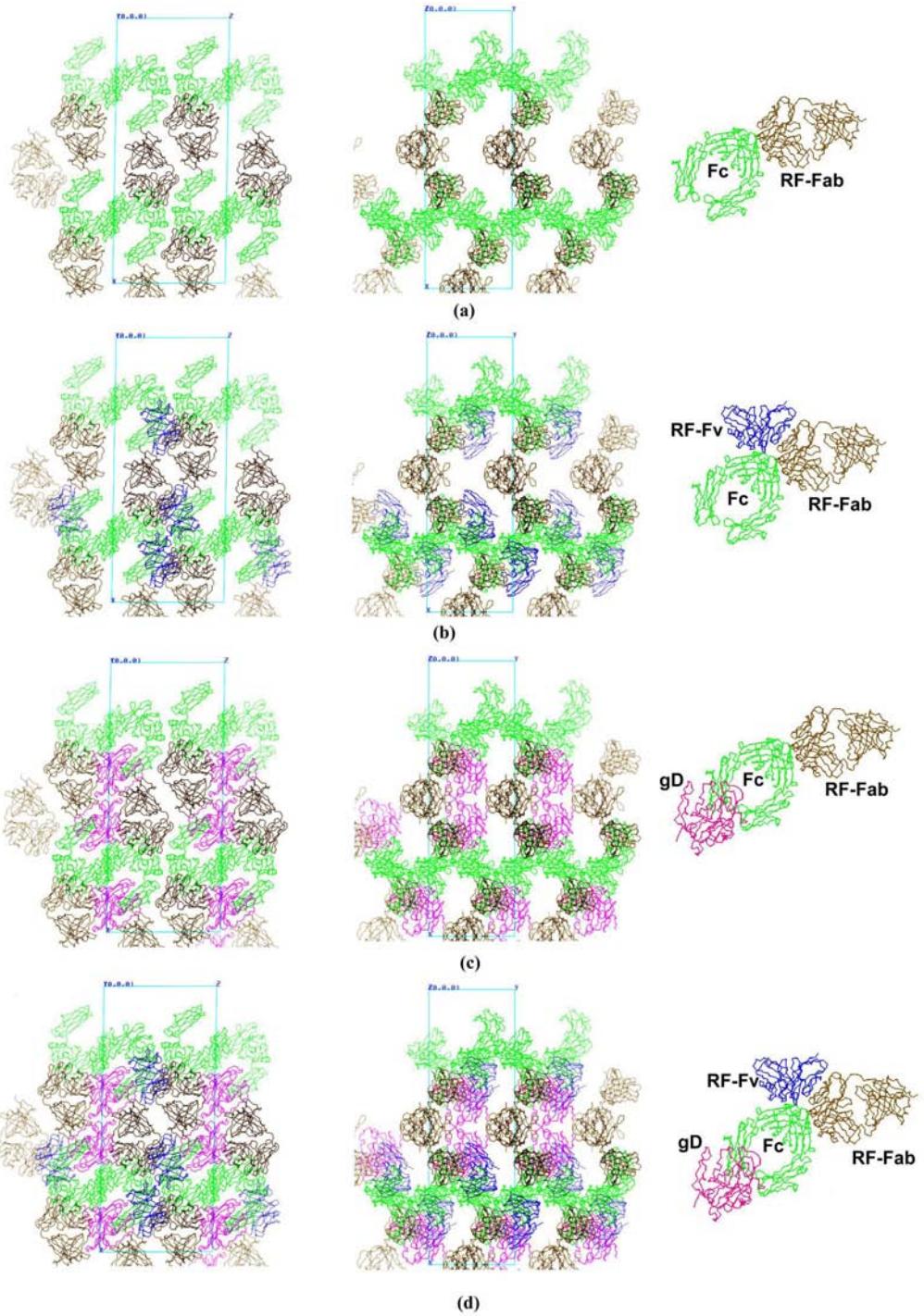

Figure 2

Backbone traces showing orthogonal views of the packing (column 1 down the z-axis; column 2 down the y-axis) of the different building blocks: RF61 Fab & Fv, Fc and gD in the unit cell (column 3) for data set FcRF (Table 1): (a) The lattice with just one Fc (green) and one Fab (brown). The Fab molecule connected to the Fc molecule forms a three dimensional continuous reticulum that builds up the basic crystalline lattice. Without further components this is the skeleton on the crystal scaffold. The solvent content of this lattice is greater than 72% leaving plenty of space for other components. Binding mode of the two building blocks of the basic lattice, the Fc and the Fab. The CH2-CH3 junction where the bacterial proteins bind is not blocked by the rheumatoid factor Fab recognition of the Fc. This could permit the use of SpA and SpG based fusion proteins together with this Fab and Fc. (b) The addition of an RF-Fv (blue) gives rise to a symmetric binding of two antibody fragments to the dimeric Fc. This reduces the solvent content of the cell 65.2% and increases the area buried in crystal contacts by 1257 Å^2 distributed exclusively on the Fc and the Fv showing that the Fv does not participate substantially to building the three dimensional lattice. (c) The basic lattice with a gD molecule (pink) added. The added gD packs against the two N terminal domains of the Fc though it is connected via it linker to just one of then. The contribution of this fusion protein to the crystal contacts is 1298 Å^2 making crystal contacts with either the Fc and the Fab. This addition reinforces the contacts between the different rows of the scaffold molecules (Fc and Fab). (d) The basic lattice with both the Fv portion of a second Fab and the gD fusion protein. This cell can accommodate either two complete Fab, two gD molecules or one Fab, one Fv one Fc and one gD fusion. With these components the solvent content will range from 50 to 60%.

Table 2 Protein content of the unit cell and molecular replacement results for the various models obtained using the data set of FcRI.

Model	Fab Fc	Fab Fc Fv	2XFab Fc	Fab gD Fc	Fab $(gD)_2Fc$	Fab gD Fc Fv
Protein content (%)	27.6	34.8	41.3	38.2	48.9	45.5
Area buried in crystal contacts ($Å^2$)	2105	3362	5048	3398	5387	5685
Molecular Replacement	*AMoRe*	*AMoRe*	*AMoRe*	*AMoRe*	*AMoRe*	*AMoRe*
Correlation coefficient (%)	46.1	51.5	47.1	44.4	42.8	49.5
R-cryst (%)	44.7	43.2	45.0	45.5	46.3	43.9

round of parallel screening with subsequent optimisation of ratios to improve crystal size. Crystals were grown by vapour diffusion at 17°C in sitting drops by mixing 1.5 µl of a reservoir solution consisting of 10% (w/w) monomethyl polyethylene glycol (MPEG) 5000, 3 mM zinc acetate, 3 mM $CdCl_2$, 100 mM sodium cacodylate, pH 6.5 with 1.6 µl of RF61 Fab at 12 mg/ml and 0.8 µl gDFc at 10 mg/ml in 50 mM ammonium acetate, pH 6.0. Crystals for data collection were enlarged by using streak seeding followed by macroseeding (Stura, 1999). The crystals were transferred to a cryoprotectant consisting of 14% (w/w) MPEG 5000, 27% ethylene glycol, 1.7 mM zinc acetate, 1.7 mM $CdCl_2$, 55 mM sodium cacodylate, pH 6.5 and then cryo-cooled in liquid ethane and stored in solid ethane in a liquid nitrogen storage tank until defrosted in the cryostream at the data collection facility. Heavy atom soaks were carried out on sitting drop microbridges. Crystals were soaked in a solution of 10% PEG 3350, 5% isopropanol, 50 mM sodium cacodylate, pH 6.5 and 2 mM each of two heavy metals: $HgCl_2$ and $EuCl_3$ (Hg-Eu); $CdCl_2$ and $SmCl_3$ (Cd-Sm) for two days. The (Sr-Pt) data set (Table 1) was collected from crystals soaked in 2 mM $SrCl_2$ and 1 mM $K_2Pt(CN)_4$ for 18 hours before vitrification in liquid ethane. The two native data sets were collected from crystals that were grown over different time spans (Table 1). Data set "FcRF" was obtained from crystals grown over a period of time twice as long as those used for the data set "gD-1" (Table 1), giving the residual trypsin in the RF Fab sample twice as long to digest the gD attached to the gDFc.

Data were recorded at ESRF beamline ID14-EH1 at cryo-temperature on a Quantum4 CCD detector and processed by using the HKL package (Otwinowski & Minor 1997). These crystals belong to the monoclinic space group C2 and diffract from 3.7-3.0 Å resolution. A total of five data sets were collected from crystals grown from varying Fab:gDFc ratios. The various crystals show some variation in cell parameters from data set to data set (Table 1).

4.2. Structure determination and refinement

The structure was solved by molecular replacement by using the program AMoRe (Navaza, 1994). One Fc and 2 Fab were searched for. An Fc molecule is a dimer of the C_H2-C_H3 domains, and therefore is expected to have two symmetrical binding sites for rheumatoid factor. Several Fab with elbow angles differing by 5° from each other were used in a sequential search. The model for the Fc was derived from PDB-ID: 1DN2 (DeLano *et al.*, 2000). The Fc was easily located, one Fab appeared in several solutions, but none gave a 2 Fab molecular replacement solution. The search was repeated with antibodies with the closest heavy and light chain sequences obtained from a FASTA search within the Protein Data Bank. The light chain was matched by 2FB4 and the heavy chain by 1QLR. Both gave good results but again, not a 2 Fab solution. 1QLR is an IgM Fab (Cauerhff, *et al.* 2000), and was used for further refinement with the light chain replaced by that of 2FB4 using the least square fit routine in *Xfit* (McRee, 1999). The amino acids that differed were replaced or were truncated to alanine. Further refinement was carried out using *CNS* (Brünger *et al.* 1998).

5. Results and discussion

5.1. Crystallisation results

Each gDFc molecule, composed of an Fc dimer and two gD glycoproteins has two binding sites for the RF Fab. Hence it is not surprising that the best crystals were obtained from a 1 : 2 stoichiometry of gDFc and RF61 Fab or with a slight excess of Fab. What is surprising is that these crystals grow relatively slowly while drops set up with a 1 : 1 stoichiometry of gDFc and RF61 Fab give crystals within a few hours. These crystals nucleate very rapidly producing a shower of very small crystals that are unsuitable for X-ray diffraction experiments. To date we have been able to obtain large crystals only from solutions consisting of approximately 1.5 - 2 Fab for each gDFc but the best growth conditions have 2 RF Fab for each gDFc. The crystals are flat plates and their size is 0.2 % 0.2 % 0.02 mm on average.

5.2. Molecular replacement

In the best of cases, data obtained from any of the crystals described here extends to 3 Å. The Fc portion of gDFc is easily found by molecular replacement, as was one of the two possible RF Fab fragments that could be present in the asymmetric unit (Table 2). The solvent content obtained with only one Fc and one Fab is about 72% of the unit cell. There is room in the crystal for a second Fab symmetrically placed in the structure with respect to the first. This does not lead to steric clashes with the other components of the unit cell. Such placement does not improve the statistics even after refinement, unless the C_H1-C_L domain of the Fab is removed. A MR solution for the second Fab can also be found using AMoRe2002 (Navaza unpublished) . However only the variable portion of the RF Fab contributes to the solution as seen by the statistics shown in Table 2. AMoRe can also locate two gD molecules using the coordinates from the published gD fragment (Carfí *et al.*, 2001). The two solutions superpose spatially with the second Fab (bound to an adjacent Fc in the crystal lattice). One gD solution falls on top of the variable region, the other on top of the constant region of the second Fab. Table 2 shows that the best MR statistics obtained correspond to a model in which gD is absent and the variable portion of the second Fab (Fv) is present. However, one of the two gD solutions (the one overlapping the constant portion of the second Fab) together with the Fv portion of the second Fab gives reasonable statistics, although it is clear that the gD molecule must be better positioned in the density, or else it is partially disordered. To explain a possible cleavage of the fusion protein and /or the second Fab, we have analysed the mixture of RF Fab and gDFc by SDS-PAGE. The results are consistent with residual trypsin being present in the RF Fab protein sample. This could explain the presence of some RF Fv in the RF Fab sample and the slow cleavage of gD during the crystallisation of the gDFc. The presence of uncleaved gDFc would limit the packing to a single Fab bound to the Fc portion. Cleavage of gD by residual trypsin allows the presence of two Fabs bound, and the fact that the MR solutions are less clear may mean that a mixture of the two is present in the crystals. Another alternative is that both cleavage of one gD (out of the gDFc-FcgD dimer) and the constant region of a Fab is needed for crystal growth.

5.3. Interpretation

The molecular replacement results and the presence of residual trypsin explain some of the observations from the crystallisation. The ability of the asymmetric unit to accommodate one undigested gDFc and just one Fab or two Fab with gD completely cleaved off or an Fc with just one gD cleaved off, one whole Fab and a second one cleaved to give rise to an Fv, explain in well the crystallisation observations. First, crystals grow rapidly, with a 1 : 1 gDFc : RF ratio, as the action of trypsin is not needed and there is space in the lattice for both gD attached to the dimeric Fc. This occurs rapidly also because of the relative solubility of the free RF compared to gDFc. The RF is the more soluble, and at a 1 : 2 gDFc : RF, the drop remains clear under the crystallisation conditions. The 1 : 1 gDFc : RF precipitates and nucleates rapidly, while gDFc alone precipitates without giving crystals. However, even by increasing the precipitant concentration it is not possible to grow crystals of 1 : 2 gDFc : RF much faster. The crystals can be grown in as little as two months, but the best diffracting crystal was grown over an eight month period, if the precipitant is increased the complex even under 1 : 2 gDFc : RF or greater ratio will readily precipitate. This is consistent with the need of the action of trypsin, which must cleave off at least one of the gD molecules from the dimeric Fc. Whether by making changes in the crystallisation procedure, such as the use of a trypsin inhibitor, or a non-cleavable construct of gDFc it will be possible to solve the structure of the full form of gD. This does not change the fact that the Fc-Fab based lattice can accommodate two gD or alternatively two whole Fab RF. The ability of the lattice composed of just 1 Fc and 1 RF Fab to make all the needed crystal contacts may allow rather large "guest" proteins to take the place of gD or the second RF Fab. The lattice appears also to have a certain degree of flexibility, as shown by considerable variation in the cell parameters (Table 1). This is an advantage for a scaffold system, since the added flexibility could permit the system to accommodate "guests" of slightly different size and shape. The basic scaffold lattice (Fig. 2*a*) makes extensive crystal contacts making strong Fc-Fc interactions along the y and z directions and strong Fc-Fab-Fc interactions along the x direction, and the weakest point appears to be at the Fab elbow. This is consistent with the fact that the x direction is the most variable with the ***a***-axis varying from 242 Å to 256 Å (5.8% variation) while the other axis remain substantially unchanged (1.4%). The change in the ***a***-axis is matched by a 2^{o} change in the β-angle.

5.4. Expression of scaffold proteins in a bacterial system

An important result that is derived from the molecular replacement results reported here is the novel binding mode of RF61 when compared to that of another rheumatoid factor, RFAN (Corper *et al.* 1997). While the binding of RFAN is competitive in its mode of binding to Fc with SpA and SpG, RF61 is not. This is supported by inhibition assays (Taussig, unpublished results). A combinatorial system consisting of RF61, Fc, and a guest protein fused to either SpA or SpG could be screened to select a scaffold that would share many of the characteristics of the current gDFc-RF61 system. While proteins fused to an Fc must be expressed in an eukaryotic cell system capable of glycosylating the Fc fusion protein, a SpG or SpA fusion protein could be expressed in bacteria including *Escherichia coli*. A vector for expression of SpA fusion proteins is commercially available although some modifications may be needed to exploit it in the system proposed here.

5.5. Future developments

Glycoprotein gD is essential for Herpes Simplex Virus entry into target cells. This type I membrane protein is one of several glycoprotein present at the viral surface. Although a crystal structure for a fragment of the gD ectodomain is currently available, both alone and in complex with one of its receptors (Carfí *et al.*, 2001), several biological problems still remain unanswered. These structures show that the contacts with the receptor are made exclusively by the 36 N-terminal amino acids of gD. In the absence of receptor, the N-terminal 15 amino acids of gD are disordered and the remaining amino acids of the contact region (16 through 36) adopt a different conformation. The complete ectodomain of gD has an affinity for its receptor that is reduced by a 100-fold with respect to the gD fragment used in the structure determination. A hypothesis that could explain this observation is that the complete gD ectodomain might be stabilised by interactions of certain N-*terminus* and C-*terminus* (just before the transmembrane sequence) segments. These segments are absent in the construct from which the structure was determined. In order to interact with the receptor, this intramolecular interaction would have to be destabilised, which could explain the lower affinity of the intact ectodomain. The gDFc construction encompasses the complete ectodomain of gD and so it is important to determine its structure to answer the underlying biological questions. To this end we are designing a construct without the trypsin cleavage sites and with linkers of variable lengths.

6. Conclusions

We have determined that one RF61 Fab and one dimeric Fc can form a stand-alone lattice without further contribution from the guest protein. When trypsin cleaves off the guest protein, crystallisation is slowed down but not disrupted. We have determined that the mode of binding of RF61 is compatible with scaffolds based on Fc-fusion or SpA/SpG-fusion proteins. This allows both glycosylated and non-glycosylated proteins to use the same screening system. Although, scaffold systems can be used in all situations where crystals cannot be obtained easily by other means, the system is likely to be particularly useful for the crystallisation of glycoproteins that have heterogeneous glycosylation and which is a class of proteins that pose particular problems for crystallisation. Within the confines of a scaffold, such heterogeneous glycosylation is likely to have a smaller effect on the resolution to which crystals diffract, as this property is determined principally by the strong periodic interactions that build up the scaffold lattice. The Fc-fusion system is also particularly adapted to glycoproteins as the presence of Fc may enhance expression.

This and other similar scaffold systems will still require screening to obtain the best possible crystals, but they can rescue important projects that do not yield crystals in the classical approach.

We thank the staff at the ESRF synchrotron facility in Grenoble on beamlines ID14-EH2 and ID14-EH1. B.J.S. and M.J.T. thank the Arthritis Research Campaign (UK) and the Biotechnology and Biological Sciences Research Council (BBSRC; UK) for support. We are grateful to A. Carfí and D. Wiley for the coordinates of the gD/HveA complex prior to publication. We are indebted to J. Navaza for help with molecular replacement and M. Graille for reading the manuscript. Part of this work was funded by a EU BIOMED grant, for the period 1997-2000, to ACM and FAR.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve R. W., Jiang, J. S, Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905-921.

Carfí, A., Willis, S. H., Whitbeck, J. C., Krummenacher, C., Cohen, G. H., Eisenberg, R. L. & Wiley, D.C. (2001). *Mol.Cell*, **8**, 169-179.
Cauerhff, A., Braden, B. C., Carvalho, J. G., Aparicio, R., Polikarpov, I., Leoni, J. & Goldbaum, F. A. (2000). *J. Immunol.* **165**, 6422-6428.
Corper, A. L., Sohi, M. K., Bonagura, V.R., Steinitz, M., Jefferies, R., Feinstein, A., Beale, D., Taussig, M. J. & Sutton, B. J. (1997). *Nature Struct. Biol.* **4**, 374-381.
DeLano, W. L., Ultsch, M. H., De Vos, A. M. & Wells, J. A. (2000). *Science* ,**287**, 1279-1283.
Graille, M., Stura, E. A., Taussig, M. J., Corper, A., Sutton, B. J., Charbonnier, J.-B. & Silverman, G. J. (2000). *Proc. Natl. Acad. Sci. USA*,. **97**, 5399-5404.
Graille, M., Stura, E. A., Housden, N. G., Beckingham, J. A., Bottomley, S. P., Beale, D., Taussig, M. J., Sutton, B. J., Gore, M. G. &Charbonnier, J.-B. (2001). *Structure*, **9**, 679-687.
Harindranath, N., Goldfarb, I. S., Ikematsu, H., Burastero, S. E., Wilder, R. L., Notkins, A. L. & Casali, P. (1991). *Int. Immunol.* **3**, 865-875.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Navaza, J. (1994). *Acta Cryst.* A **50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Randen, I., Potter, K. N., Li, Y., Thompson, K. M., Pascual, V., Forre, O., Natvig, J. B. & Capra, J. D. (1993). *Eur. J. Immunol.* **23**, 2682-2686.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273-285.
Stura, E. A. (1999). *Crystallization of Nucleic Acids and Proteins: a practical approach*, edited by A. Ducruix & R. Giegé, 2nd ed., pp. 177-208. Oxford University Press.
Stura, E. A. (2001). *J. Cryst. Growth*, **232**, 545-552.
Stura, E. A., Graille, M. J. & Charbonnier, J.-B. (2001a). *J. Cryst. Growth*, **232**, 573-579.
Stura, E. A., Graille, M., Taussig, M. J., Sutton, B. J. Gore, M. G., Silverman, G. J. & Charbonnier, J.-B. (2001b). *J. Cryst. Growth*, **232**, 580-590.
Stura, E. A., Graille, M., Houdsen, N. G. & Gore, M. G. (2002a). *Acta Cryst.* D**58**, 1744-1748.
Stura, E. A. , Tête-Favier, F., Delforge, D., Muller, S. & Aubry, A. (2002b). *Acta Cryst.* D**58**, 1740-1743.
Zhang, M., Majid, A., Bardwell, P., Vee, C. & Davidson, A. (1998). *J. Immunol.* **161**, 2284-2289.

Received 7 May 2002 • Accepted 16 July 2002