UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE 2.3Å CRYSTAL STRUCTURE OF THE ANTIBODY FAB HPC-4 IN COMPLEX WITH CALCIUM AND THE EPITOPE PEPTIDE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

PHILIPPE GARTEISER
Norman, Oklahoma
2007

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/1189

UMI Number: 3261112



UMI Microform 3261112

Copyright 2007 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE 2.3Å CRYSTAL STRUCTURE OF THE ANTIBODY FAB HPC-4
IN COMPLEX WITH CALCIUM AND THE EPITOPE PEPTIDE

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF BIOENGINEERING

BY

_____
Dr. M. U. Nollert

_____
Dr. T. Mather

_____
Dr. E. A. O'Rear III

_____
Dr. C. T. Esmon

_____
Dr. D. W. Schmidtke

_____
Dr. A. Striolo

© Copyright by PHILIPPE GARTEISER 2007
All Rights Reserved.

In memory of Dr. F. Farkas

TABLE OF CONTENTS

TABLE OF CONTENTS iv
ABSTRACT vi
1. BACKGROUND 1
1.1. Hemostasis 1
1.1.1. General background 1
1.1.2. The coagulation cascade and the protein C anticoagulant pathway 2
1.1.3. Thrombomodulin influences the properties of thrombin 4
1.2. The calcium dependent antibody HPC-4 7
1.3. Structural biology and biochemical aspects of antibody-antigen interactions 8
1.4. Calcium binding in coagulation factors and immunoglobulins 12
1.4.1. Overview 12
1.4.2. Calcium binding to blood coagulation factors 13
1.4.3. Calcium binding to proteins of the immunoglobulin fold 17
1.5. Generation of antibody fragments by proteolysis 21
1.5.1. Overview 21
1.5.2. Experimental considerations for the use of ficin in HPC-4 hydrolyses 24
2. MATERIAL & METHODS 28
2.1. Generation of HPC-4 fragments 28
2.1.1. Optimizations of HPC-4 hydrolysis and purification 28
2.1.2. Generation of the Fab'$_2$ 29
2.1.3. Generation of the ncFab 31
2.2. Generation of protein C variant 32
2.2.1. General considerations 32
2.2.2. GDPC production methods 33
2.3. Characterization of HPC-4 fragments 35
2.3.1. Fluorimetry 35
2.3.2. Modeling of HPC-4 37
2.4. Crystallization 39
2.4.1. Preparation of the complexes and buffer exchange 39
2.4.2. Concentration 40
2.4.3. Crystallization 42
2.4.4. Crystallization conditions database 43
2.4.5. Macroseeding: theory & methods 45
2.4.6. Cryoprotection methods 46
2.5. Structure determination 48
2.5.1. Initial diffraction studies 48
2.5.2. Handling of crystals at cryogenic temperatures 51
2.5.3. Data collection 52
2.5.4. Structure solving 55
3. RESULTS 57
3.1. HPC-4 ficin hydrolysis 57
3.1.1. Generation of Fab'$_2$ 57

3.1.2. Generation of the Fab fragment 60
3.2. Fluorescence titrations 65
3.3. Crystallization 68
3.3.1. Crystallization 68
3.3.2. Optimization of macroseeding procedure 74
3.3.3. Cryoprotection 78
3.4. X-ray crystal structure 83
3.4.1. Diffraction data collection 83
3.4.2. Structure solving 84
3.4.3. The structure of HPC-4: general observations 88
3.4.4. The antigen and the antigen binding region 91
4. DISCUSSION 99
4.1. Packing environment 99
4.2. On the structure of the protein C zymogen 101
4.3. Correlation with fluorescence results 105
4.4. Hypotheses about the mechanism of the switch 108
4.5. On the affinity of the peptide-Ca^{2+}-HPC-4 interaction 110
4.6. Comparison with Q-425 114
5. CONCLUSION 117
BIBLIOGRAPHY 119
APPENDICES 129
X.1. Characterization of macroseeding-assisted crystal growth 129
X.2. Cryoprotection 135

ABSTRACT

Metal ions can have considerable affinities for proteins, and give rise to geometric constraints that are often taken advantage of in protein-protein interactions. The coordination shell of metal ions can be filled by atoms provided by two different proteins, resulting in a high affinity protein complex. Surprisingly, this highly efficient binding strategy is rarely observed in immunoglobulins, despite the great number of known antibody structures determined in complex with their protein antigens.

The murine monoclonal antibody HPC-4, directed against the activation region of the human anticoagulant zymogen protein C (PC), is one of the few immunoglobulins known to display calcium-dependent antigen binding. Unlike the more common class of antibodies that merely recognize a calcium-bound conformation of their antigen, HPC-4 interacts directly with calcium in the high affinity PC-HPC-4 complex.

To provide a structural understanding of HPC-4 function and of the particular antibody class to which it belongs, we have solved the X-ray crystal structure of the HPC-4 Fab fragment in ternary complex with its epitope peptide in the presence of calcium at a resolution of 2.3Å. Within the crystal, the antigen-binding region is undistorted by crystalline lattice contacts. All complementarity-determining regions and the peptide antigen have well-defined electron density.

The structure reveals a mode of calcium binding which underlies a novel mechanism of metal aided antigen recognition. The ion is located at the antibody-antigen interface, where it functions both as an electrostatic bridge and as a conformational effector of the antibody. The antigen is further stabilized by an extensive and diverse array of interactions spanning a large surface area of contact. Our results provide a structural explanation for many of the observed characteristics of HPC-4, the first member of a unique class of calcium binding antibodies. As such, it represents a significant contribution to the study of interfacial metals and the structural biology of antibodies.

1. BACKGROUND

1.1. Hemostasis

1.1.1. General background

Blood hemostasis results from the combined actions of platelet activation, coagulation and fibrinolysis. The constantly regulated balance between clot formation and degradation required to maintain proper blood circulation is mediated through a network of interactions between these complementary pathways. The plasma protease activated protein C (aPC) is an important member of this system, illustrated by the fact that genetic defects in the protein C gene are associated with increased risks for venous thrombosis[1]. aPC is produced by an ensemble of activation reactions called the protein C pathway[2,3](fig. 1) and functions in both coagulation and inflammation[2]. The present project is primarily focused on its role in preventing the propagation of blood clotting away from a wound site. Indeed, very high concentrations of the procoagulant precursors prothrombin and fibrinogen are necessary to ensure a rapid sealing of any wound and prevent fatal blood losses. An exquisitely organized ensemble of interacting regulatory agents is dedicated to maintain the overall hemostatic balance over a broad range of accidental perturbations.

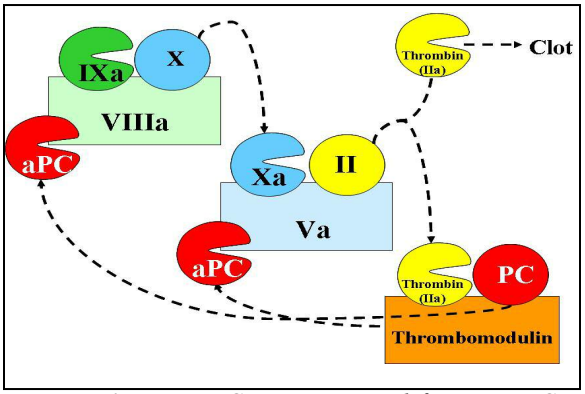


Figure 1: protein C activation and function. PC, aPC: protein C, activated protein C. IXa: activated factor IX. X, Xa: factor X, activated factor X. II: prothrombin.

1.1.2. The coagulation cascade and the protein C anticoagulant pathway

Thrombin is at the endpoint of a series of proteolytic activations defined as the coagulation cascade that is initiated at a site of vascular injury upon exposure of blood to tissue factor within the sub-endothelial tissues. The functional unit of this amplifying sequence is a complex between an obligatory cofactor and a serine protease of the trypsin class. The binding of an enzyme (factor IXa in the example shown in fig. 1) to the cofactor (factor VIIIa) enables the specific substrate cleavage of a downstream serine protease precursor (zymogen, factor X in fig. 1), which in turn activates a different procoagulant zymogen (factor II) in a cofactor-dependent (factor Va) fashion. Because blood coagulation operates at the level of mature circulating proteins, it mostly depends on the concentrations and chemical rates of the procoagulant reactions, and consequently may be very rapid. This benefit is however only obtained by having high levels of circulating procoagulant precursors in the normal situation. In this condition, any uncontrolled procoagulant signal can potentially spread to lethal levels rapidly. The regulatory mechanism of the cascade, triggered simultaneously with the formation of the clot, is the anticoagulant protein C pathway. Its main function is the proteolytic inactivation of two cofactors (factors Va and VIIIa) required for the final amplification steps of the procoagulant pathway[2-4]. The central reaction of this regulatory pathway is the activation of protein C, catalyzed by the complex thrombin-thrombomodulin (T-TM)[5] in a process assisted by the endothelial protein C receptor, EPCR[6,7]. Recruitment of the major anticoagulant enzyme thrombin as the initiator of the anticoagulant pathway is an elegant solution to ensure that any procoagulant event inherently also generates the

enzymatic activities required to stop itself. This process is sometimes referred to as autodownregulation[8], or negative feedback[9]. Two main mechanisms are available to counteract coagulation. At later stages (hours to days after injury), simultaneously with wound healing, the clots are degraded by fibrinolysis, which can be used as a therapeutic treatment[10] as extensively reviewed elsewhere[8].

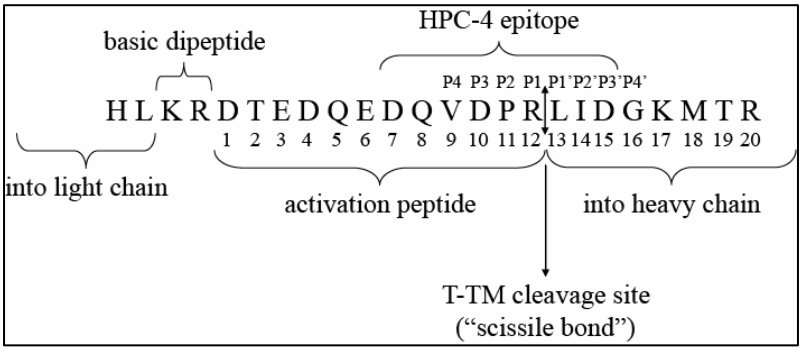


Figure 2: Sequence of protein C (human) in the activation region. The Schechter&Berger numbering is shown atop the sequence. The aPC heavy chain numbering is shown below the sequence. The epitope peptide used throughout this project is shown.

Protein C, the substrate of the T-TM complex, is a plasma serine protease of the trypsin class synthesized in the liver[11]. It is initially secreted as a single polypeptide, but the removal of an internal dipeptide ($K_{156}R_{157}$, see fig. 2) during endoplasmic reticular transit results in the formation of a disulfide bonded heterodimeric species constituted of a light chain and a heavy chain[12,13]. About 20 % of plasma protein C circulates as a single chain isoform where the dipeptide is still present[12,13]. This form has the same reactivity towards the T-TM complex, and yields a functionally identical version of the active protease[14]. The light chain contains a gamma carboxyglutamic acid domain ("Gla domain"), a helical aromatic segment, two successive epidermal growth factor-like domains ("EGF domains") and a C-terminal peptide linking to the interchain disulfide[15]. The heavy chain is a serine protease domain. It contains the canonical catalytic triad of residues (H_{57}, D_{102}, S_{195})

which carry the catalytic activity of the enzyme[16]. The structure of a Gla domainless form of the activated protein C was solved at 2.8Å resolution[17] (fig. 3). The activation reaction takes place at the amino terminus of the heavy chain, where a 12 residue peptide is removed by cleavage of the R_{12}-L_{13} peptide bond by the T-TM complex[3]. Structural evidence demonstrates that the newly formed amino-terminus then folds back in a canonical fashion[18] inside the serine protease domain, where it forms a buried ion pair with D_{194}[17]. The ensuing conformational rearrangement is transmitted to the active site, resulting in a catalytically competent enzyme.



Figure 3: aPC, from pdb code 1AUT. Blue: EGF1, red: EGF2, green: protease domain. The catalytic triad is visible as sticks. The intrachain disulfide is visible as yellow sticks.

1.1.3. Thrombomodulin influences the properties of thrombin

The binding of thrombin to the endothelial receptor thrombomodulin (TM) induces profound changes in the enzyme that enable it to activate the anticoagulant protein C while suppressing its procoagulant activities[5]. The unique cofactor activity of TM resides within a tandem repeat of six epidermal growth factor domains (EGF domain). The binding site of thrombin is located in the last 2 EGF-like folds of the EGF domain (TM56)[19], but all of the last 3 EGF-like folds (TM456) are required for a full cofactor activity[20].

Although the role of TM in blocking the procoagulant activities of thrombin is well established, the precise mechanism by which it confers to thrombin the novel ability to activate protein C remains unclear. The nanomolar affinity of thrombomodulin for thrombin certainly outcompetes other procoagulant substrates that share the same binding site[21], and several lines of evidence show the presence of a protein C binding site on TM4[22-24]. This orienting effect is comparable to other reactions of the coagulation pathway[25], and was confirmed by the structure of thrombin in complex with the TM456 fragment of thrombomodulin[26], in which only modest changes were found in the structure of thrombin. In this study, it was concluded that the binding event only increases the activation rate by providing a binding site for protein C on the activating complex. However, this result must be taken with caution and could be an artifact from the stabilizing presence of the covalent inhibitor d-Phe-Pro-Arg chloromethyl ketone (PPACK) in the structure[27] or related to packing contacts which fill the TM binding site of the apo-thrombin structure used for comparison (manuscript in preparation). Furthermore, a growing body of evidence shows that the most prominent effect of thrombomodulin is the modulation of the catalytic rate of thrombin hydrolysis rather than affinity improvement alone. This type of effect[28] suggests that thrombomodulin could be an allosteric effector of thrombin. Results obtained by differential scanning calorimetry[29], differential susceptibility to proteolysis[30,31], active site-tethered fluorescent probes[32], spin-label agents[33], deuterium exchange-mass spectrometry[29,34] and site-directed mutational analysis[35,36] all converge to the same

hypothetical mechanism in which binding of this cofactor is able to modify the conformation of distal sites including the active center of thrombin.

The structure of the substrate protein C also plays an important role in the activation reaction. This is illustrated by the fact that the conformational changes induced by calcium in protein C[14,37] are able to alter its activation rate by either thrombin or the thrombin-thrombomodulin complex[3,38,39] (Fig. 4). This effect has tentatively been addressed by theoretical modeling approaches[26,40], but important discrepancies appear between the proposed models. By virtue of its experimental nature, our project may provide a more conclusive answer about the influence of calcium on the activation of protein C by the T-TM complex.

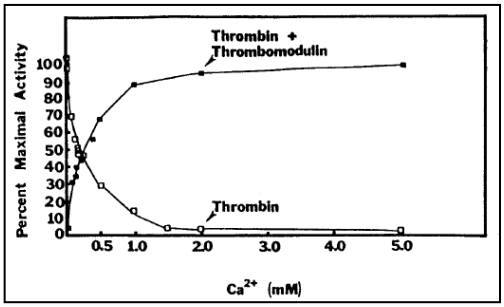


Figure 4: Calcium dependency of protein C activation. Reproduced from ref. (3).

Overall, the precise molecular mechanism of this modulation appears to be that TM provides both a weak binding site for protein C that optimizes its cleavage, and an improved catalytic rate through an allosteric effect. The knowledge of the conformation of the calcium-bound form of the substrate would improve our understanding of the mechanism of thrombomodulin-mediated specificity change of thrombin.

1.2. The calcium dependent antibody HPC-4

The monoclonal antibody HPC-4, directed at the activation region of the human anticoagulant protease precursor protein C (linear epitope represented in fig. 2), has the remarkable characteristic of a high magnitude and reversible increase of affinity for the antigen only in the presence of calcium[41]. Initial biochemical characterizations concluded that this unusual characteristic stems from the presence of a calcium binding site on the antibody, but its exact nature and the resulting mechanism of the calcium dependence of HPC-4 remained under discussion. In particular, the coordination of the calcium ion could be fully provided by the antibody, resulting in a conformational switch that would trigger the conversion of HPC-4 to a high affinity state. Alternatively, the simpler mechanism of strict electrostatic bridging likely to occur in the only other structurally solved calcium dependent antibody Q-425 could also explain some of the biochemical data for HPC-4. These two mechanisms are not exclusive and could likely be combined. The resulting dual effect of conformational rearrangement of the antibody and electrostatic bridging, although unique in the extensively characterized field of antibody-antigen interactions, is not without precedent in other systems.

The region of protein C recognized by HPC-4 is centered on the peptide bond cleaved during the activation reaction by the thrombin-thrombomodulin complex (T-TM)[41]. The activation peptide released during activation is not recognized by HPC-4, nor is the activated protein C (aPC). The crystal structure of aPC[17] clearly shows that the portion of HPC-4 epitope which remains after activation is recruited away

from the solvent to provide a buried ion pair essential to the catalytically active conformation of the active site triad. Importantly, protein C activation by T-TM only occurs when the region of the scissile bond undergoes a conformational change induced by the binding of calcium to the calcium binding loop in the protease domain of protein C[14,17,39,42]. Structural modeling, homology modeling and molecular dynamics simulation results suggest that this effect is mediated through direct contact between the two regions[17,40]. Importantly, the calcium dependence of HPC-4 binding was observed even for the small peptide of protein C that spans the scissile bond, thus devoid of the structural influence of the calcium-binding loop of the serine protease domain.

To decipher the effects of calcium on HPC-4 and the resulting increase of affinity for even a structurally relaxed antigen, we have solved the X-ray crystal structure of an antigen binding fragment ("Fab") variant of HPC-4 in complex with the 9 residues protein C epitope peptide in the presence of calcium. Our results provide important insights into the mostly unexplored field of interfacial metal aided antigen recognition.

1.3. Structural biology and biochemical aspects of antibody-antigen interactions

Antibodies are heterotetramers of 2 heavy and two light chains linked by disulfide bonds. Heavy and light chains are built with repetitions of a common all-β structural domain (the Ig fold) belonging to the immunoglobulin structural superfamily, as defined in the SCOP (structural classification of proteins[43]). Heavy

chains have 4 Ig folds, while light chains only have 2. Heavy chains are bound together between the second and third Ig domains by several disulfide bonds of the hinge region. Each heavy chain is bound to a light chain by one disulfide bond. There are an additional 12 intrachain disulfide bonds, one in each Ig domain. The first Ig domains of both the heavy and light chains are grouped in the tertiary structure, and together form the antigen binding site. The Ig folds participating in antigen binding have an additional two strands that are important for their function.

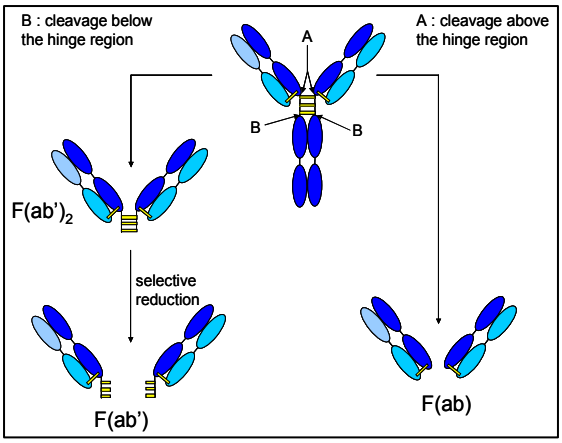


Figure 5: Antibody hydrolysis fragmentation routes. Dark blue : heavy chains ; light blue: light chains. Interchain disulfides are depicted as yellow boxes.

Antibodies can be fragmented by proteolysis, yielding different fragments depending on the place of cleavage. In the case of a cleavage N-terminal to the hinge region, two monovalent "Fab" fragments are generated along with the constant fragment ("Fc"). In the case of a cleavage C-terminal to the hinge region, the bivalent Fab'$_2$ fragment is obtained together with the Fc. These different possibilities are presented on figure 5.

Antigen specificity arises from the high sequence variability encountered in the loops that connect the strands 2-3, 3b-3c and 6-7 (fig. 6). The molecular basis for antigen recognition is reasonably well understood, mostly because many high resolution crystal structures of antibody fragments in presence or absence of their antigen have been solved. First, the structural variability of antibody molecules is

almost exclusively located in hypervariable regions, the rest of the fold being essentially identical in structure between different antibodies. Protein-binding antibodies share the additional feature of having an antigen binding site resembling a rather shallow groove. Their amino acid composition is believed to be biased with an unusually high percentage of asparagine, histidine[44,45], tyrosine[45,46] and serine[47], although some variability exist in the literature. These analyses were performed when antibody structures were still relatively uncommon, and the datasets that were used may not have been widely representative. The conformational flexibility of antigen binding sites is also subject to debate, and examples of both key-lock and induced fit mechanisms have been documented. In the cases where conformational changes were detected, they involved sidechain rearrangements[48], movements of CDR backbone atoms, modifications to the hydration layer involved in antigen binding[49] and even shifts in the relative positions of the heavy and light chains[50,51].

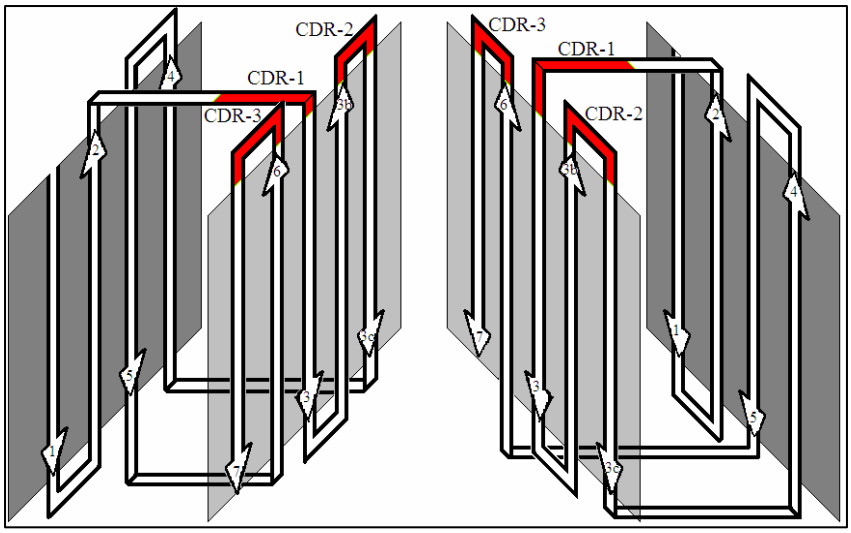


Figure 6: Topology of the all β immunoglobulin fold. CDR: complementarity-determining region. Immunoglobulin folds from the heavy and light chains face each other. The approximation of the CDRs in antibodies results in the antigen binding crevice.

The thermodynamic aspects of antibody-antigen interactions share many common points with protein-protein interactions in general, which may be one of the

reasons why antibody-antigen systems are often used in thermodynamic studies of protein interactions. The forces responsible for antigen binding are believed to be mediated primarily by van der Waals interactions, hydrogen bonding and occasional ion pairing. It is somewhat surprising considering the high affinity of antibodies that they would rely so little on the stronger ion pairing. However, contact surfaces of antibody-antigen complexes are usually very extended (commonly 600-900Å^2)[52], which allows for the accumulation of an accordingly great number of weak interactions which result in the strong affinity. As an increasing number of atomic resolution crystal structures were obtained, it became evident that water molecules are playing an important role in antigen interactions. The anti hen egg lysozyme antibody D1.3 was thoroughly used to characterize this phenomenon precisely [53,54]. This experimental framework helped obtain a clear breakdown of the various energetic contributions to binding. Enthalpy contributions provided by direct binding of the antigen to the antibody or to interfacial water molecules were shown to offset the entropic loss resulting from increased order of the loops and of the solvent molecule network in the interface of the complex[55]. In this particular system, the solvent entropy increase resulting from hydrophobic interactions was not found to provide a significant amount of energy to the system. Theoretical considerations based on numeric computations hypothesized that crystal structures could be underestimating the solvation state of antibody-antigen interfaces[56]. The importance of this phenomenon was also highlighted by the finding that reorganizations of the interfacial network of water molecules were efficient enough to compensate the enthalpy losses inflicted by a mutation[57]. Whether or not these considerations remain

true for other antibodies is unclear, but the results of this project will provide additional data addressing these questions.

1.4. Calcium binding in coagulation factors and immunoglobulins

1.4.1. Overview

Calcium is an extremely important element in bioinorganic chemistry. It is a crucial component of bone tissue, where it is present as calcium phosphate[58]. However, soluble forms of calcium also fulfill an impressive range of functions that are mostly mediated by inducing conformational changes upon binding to a variety of proteins. It is a ubiquitous intracellular messenger with roles including, but not limited to, skeletal or cardiac muscle contraction, vision, or gene regulation[59]. Calcium signaling is very rapid, because in a normal situation its cytoplasmic concentration is very low compared to its concentration in the cytoplasmic stores (mostly the endoplasmic reticulum) or in the extracellular or interstitial compartments. These gradients are taken advantage of to obtain fast concentration increases by opening ion-specific membrane channels. Once released in the cytoplasm, calcium binds to protein sensors of the EF-hand class such as calmodulin or troponin, resulting in a dramatically altered conformation which carries the effector functions. More importantly for this project, calcium is a key component in plasma, where its presence at levels of 2-3mM is necessary for proper function of the blood

coagulation pathway. This requirement stems from a variety of interactions that exist between calcium and the different domains constituting the protein factors of blood coagulation, particularly in the Gla domains, where the average K_d values are nearly millimolar. Finally, as our results unequivocally show, calcium can also bind proteins of the immunoglobulin fold to participate in a specific interaction with a protein ligand. In the present review we will only focus on the coagulation factors and on the immunoglobulin fold because of their relevance to the crystal structure that was obtained.

1.4.2. Calcium binding to blood coagulation factors

Calcium binding to the Gla domain: An important number of reactions involved in hemostasis take place on the surface of the injured endothelium. This localization of the coagulation reactions is mediated by the interaction of a specific domain of the factors, the gamma carboxyglutamic acid domain (Gla domain) with the injured cell membrane. This domain is characterized by the presence of gamma carboxyglutamic acids, which are post-translationally derived from the glutamate residues of the polypeptide precursor in a vitamin K-dependent reaction[60]. It is

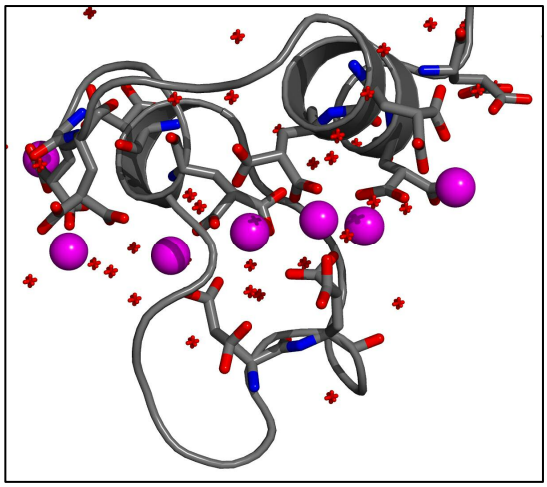


Figure 7: Gla domain, from reference (62), pdb code 2PF2. Calcium atoms are represented as spheres, water molecules as crosses. The gamma carboxyglutamic acid residues are displayed as sticks.

found in coagulation factors II, VII, IX, X, protein C and protein S[61]. The first crystal structure of this domain from prothrombin[62] revealed that the carboxy glutamic acid residues are involved in a network of interactions with 7 calcium ions, which results in the particular all-α domain illustrated on figure 7. Although the binding of calcium was at first thought to alleviate electrostatic repulsion between acidic residues of the Gla domain and the negatively charged membrane phospho-lipids, it was later demonstrated that in the case of the Gla domain of factor X, important structural rearrangements take place upon calcium binding, that result in the exposure of hydrophobic residues[63]. Similar findings were obtained for the Gla domain of protein C[64,65]. Together, these data would readily explain the increased affinity for membranes of the calcium-bound form of the Gla domain. The interaction of calcium bound Gla domains with membranes is not the only function of this domain. In addition, it can participate in mediating protein-protein interactions that are functionally significant. For instance, the Gla domain of protein C enables specific interactions with both the endothelial protein C receptor[65,66] and the cofactor protein S[67].

Calcium binding to the EGF domain: The epidermal growth factor domain (EGF domain), found in proteins of the procoagulant, anticoagulant and fibrinolytic pathways, is a small, all-β protein fold which has in some instances the ability to bind calcium[68]. This occurs in thrombomodulin[26] or protein S[69], the cofactor of aPC. The first EGF domain in protein C also has a calcium binding site, but its role in the function or conformation of the protein remains under study[70]. The situation is very different in factor VII, where the occupancy of the EGF calcium binding site was

found to have an effect in mediating the interactions with tissue factor. Interestingly, this effect was only detected in the presence of the adjacent Gla domain[71]. This type of interaction between a Gla and an EGF domain is also encountered in factor X[72], although it was found that calcium can bind an individual factor X EGF domain[73]. Importantly, the mediation of protein-protein contacts by calcium in EGF domains is not limited to interactions with the Gla domain. This was demonstrated for one of the EGF domains of factor IX[74]. In this particular X-ray crystallography study, the authors obtained a crystal form in which calcium mediates crystal contacts between EGF domains, thus illustrating the ability of calcium to mediate contacts between different EGF domains. Another example is the calcium binding site shared between the EGF-like domains 5 and 6 of thrombomodulin[26]. The calcium-mediated binding of an EGF domain to a different domain is also encountered in proteins that do not belong to the coagulation factor family, such as the complement triggering protease C1s[75] or the extracellular protein fibrillin[74,76].

Calcium binding to the serine protease domain: Lastly, the catalytic domains of some of the blood coagulation factors also share a conserved trypsin-like calcium binding loop[61,77] spanning residues 70-80. In many cases, calcium binding is required to get full activity. This is the case for example in factor VII[78-81], factor IX[82], factor X[83], or protein C[84], although in some instances the presence of a calcium binding site on the substrates complicates the analysis. Importantly, prothrombin does not have a calcium binding site in the 70-80 loop. Instead, a lysine residue is present at position 70, where an acidic residue is usually present, with its epsilon amino group occupying the regular calcium position. The resulting ion pair with the aspartic acid at position

80 further prevents calcium binding, and locks the loop in a conformation essentially similar to that encountered in other, calcium occupied 70-80 loops. Other coagulation factors that contain the serine protease domain however, are all characterized by calcium binding to the 70-80 loop.

Protein C, of which a fragment of the activation region is present in the structure that was obtained during this project, also has a calcium binding site in the serine protease domain[85]. The occupancy of the calcium binding loop has a tremendous effect on both the activation rate[3] (fig. 4) and on the catalytic activity[84]. The effect of calcium on activation was demonstrated by kinetic studies, which revealed that calcium is a potent activator of the activation reaction catalyzed by the T-TM complex, but inhibits the reaction in the absence of TM[3]. The effect of calcium was unequivocally proved to be mediated by a site in the serine protease domain. An E80K mutant of Gla-domainless protein C with an ion pair engineered into the putative Ca^{2+} binding site in the manner of thrombin/prothrombin[42] showed, in the presence of EDTA, the activation behavior of wild type protein C with Ca^{2+} bound. In addition, a proteolytic deletion of the Gla domain would not disrupt the calcium dependence of T-TM activation[85]. This indicates that the ion maintains the serine protease in a conformation that is recognized by the T-TM complex, but not by thrombin alone. This effect is thought to be mediated through a conformational interplay that would occur between the calcium binding loop and the activation loop of protein C. Several experimental results favor this hypothesis. For instance, the introduction by mutagenesis of a tryptophan residue into the activation region dramatically increases the fluorescent properties of the protein in a calcium-

dependent fashion[37]. Alternatively, by replacing either of the two tryptophan residues naturally present in the calcium binding loop of protein C by phenylalanine, it was observed that the fluorescent properties usually seen in the wild type protein were lost[14]. Theoretical modeling studies suggest that in the structure of the zymogen, the activation loop and the calcium binding loop may be in direct contact[17,40]. Finally, the replacement of two aspartic acid residues present three positions away on each side of the scissile bond (positions 10 and 15 in the chymotryspin numbering scheme, or P3 and P3' in the Schechter and Berger nomenclature[86]) by glycines decreases both the affinity of the serine protease domain of protein C for calcium, and the calcium requirement of the activation reaction by T-TM[39]. Overall, the serine protease domain of protein C illustrates the importance that calcium-mediated conformations can have on the biological function of a protein.

1.4.3. Calcium binding to proteins of the immunoglobulin fold

In contrast to the other examples provided so far, the immunoglobulin fold is rarely observed to bind calcium. The Fab molecule is a well populated example of structurally solved immunoglobulin folds, with 652 structures out of the 1512 immunoglobulin structures available in the March 2007 release of the protein databank (PDB)[87]. Among these 652 entries, only 24 (3.68%) have a crystallographic calcium, and only one (0.15%) has a calcium at the antibody-antigen interface[48]. This unique structure is of the Fab of the monoclonal antibody Q-425,

directed at the human CD-4 receptor. Because of its singular relevance to the calcium dependence of HPC-4, this structure will be discussed here.

The authors obtained the structure of the Fab in the presence or absence of calcium. All crystals were obtained in the absence of antigen. As seen on figure 8, the calcium ion is located in the center of the antigen-binding crevice. The complex seen in the 2.9Å crystal structure has a relatively weak dissociation constant of 187µM, as measured by surface plasmon resonance. This relatively mild affinity is due to the fact that the coordination shell of the ion lacks three additional positions, hypothesized to be provided by the antigen. The antibody provides 4 atoms for calcium coordination, located one each in the CDR-H3, CDR-L1, CDR-L2 and CDR-L3. Because of its central position in the antigen binding site, it is hypothesized that the calcium ion acts mostly as an electrostatic bridge between antigen and antibody. Contrary to what will be shown in HPC-4, no structural rearrangement is taking place in Q-425 upon calcium binding, as demonstrated by the absence of major changes between the calcium-bound and calcium-free structures. The backbone traces of these two structures are essentially similar (RMS for backbone atoms of 0.433Å), and the only significant sidechain rearrangement is a 180° rotation of the C_β-C_γ bond in asparagine 100A. The functional character of the

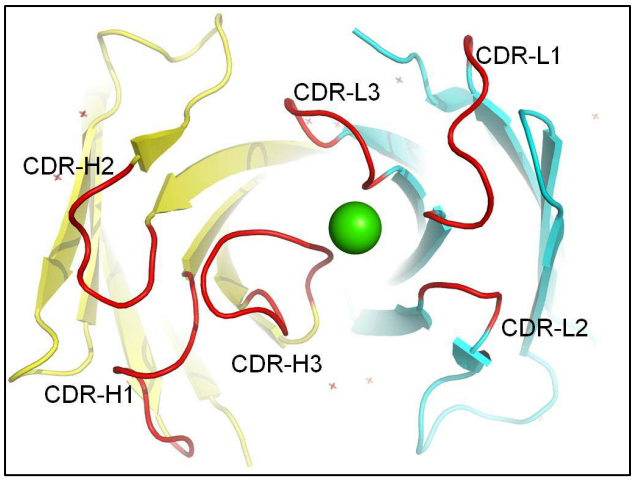


Figure 8: Q-425 antigen binding site. From reference (48), pdb code 2ADJ. Yellow: heavy chain; grey: light chain; red: CDRs. The calcium ion is represented as a green sphere.

calcium binding site seen in the Q-425 structure is emphasized by the several thousand fold increase in antigen affinity resulting from its occupancy. By surface plasmon resonance, the authors demonstrated that the strength of the interaction with calcium was responsible for most of this affinity improvement. They suggested that interfacial calcium was advantageous in this particular case to compensate the limited antigen surface available for binding due to heavy glycosylation.

The sensory domain of the Na^+/Ca^{2+} antiport[88] is a different example of a functionally significant calcium binding site occurring in an immunoglobulin domain. This transmembrane transporter allows the flux of calcium out of the cytoplasm and is gated by the calcium concentration inside the cell. The cytoplasmic calcium levels are detected by the transporter through binding to the intracellular sensory domain. The structure of this domain (fig. 9) revealed that 4 calcium ions were located at the C-terminal portion of the fold, held in place by an extensive network of interactions with acidic residues and water molecules. It is thus very different from Q-425 both in terms of calcium binding mode and function. Whereas calcium is found at the interface between Q-425 and its antigen, the ions found in the Na^+/Ca^{2+} antiport sensory domain are fully coordinated by the immunoglobulin fold. The mode of binding is intimately linked to the function of each protein. In Q-425, calcium binding at the interface helps the formation of the antigen complex by

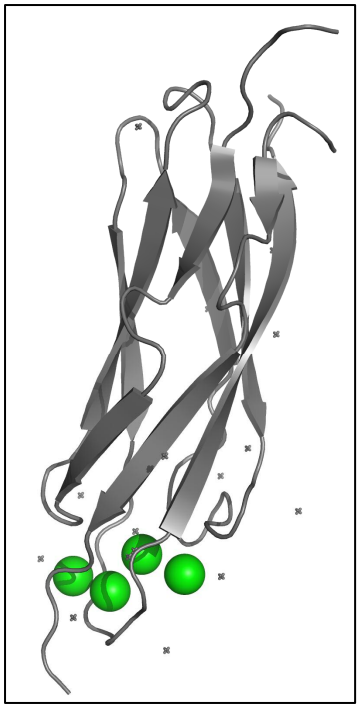


Figure 9: The sensory domain of the Ca^{2+}/Na^+ antiport. From reference (88), pdb code 2DPK. Green spheres: calcium ions.

providing significant energetic contribution. In contrast, calcium binding to the Na^+/Ca^{2+} antiport sensory domain likely induces a conformational change of the antiport which results in its activation[89]. This second example is relevant to HPC-4 in that it illustrates the flexibility of the immunoglobulin fold and its ability to adopt calcium-mediated conformational changes. In principle, such a mechanism could be expected for HPC-4, where binding of calcium to the antibody would mediate a structural rearrangement of the CDR loops that would result in a conformation with high affinity for the antigen. However, the mechanism could not be exactly similar for the following reasons. First, the sensory domain of the antiport is actually an immunoglobulin constant-like domain, and has two strands less than the antigen binding domains of antbodies. If binding of calcium were to occur in exactly the same fashion in HPC-4, the conformational changes would have to be transmitted through the flexible elbow region that separates the variable and constant domains in antibodies. Although not strictly impossible, this mode of action is unlikely. Furthermore, calcium was found at the C-terminal loops of the Ig domain in the sensory domain, but antibodies bind their antigen at the N-terminal side of the β sheet. Still, the sensory domain of the antiport illustrates the ability of the framework of β sheets in immunoglobulins to support calcium-mediated conformational changes in the connecting loops.

The mechanisms of the Na^+/Ca^{2+} antiport and of Q-425 may not be mutually exclusive even though they are quite different. HPC-4 may function according to a combination of the two, where calcium would both induce structural rearrangements in the antibody and increase the affinity of the interaction by providing binding sites

to both the antibody and the antigen. The structure of HPC-4 in complex with calcium and the peptide epitope described in the results section will shed some light on this particular point.

1.5. Generation of antibody fragments by proteolysis

1.5.1. Overview

Crystallization of intact antibodies has been performed only a very few times, and only once to date on the murine subtype of HPC-4, IgG1[90,91]. The most significant obstacle preventing these macromolecules from being crystallized is believed to be the flexibility of the hinge region[92]. This, in addition to the large size of antibody molecules, is the main reason crystallographic studies of antibody molecules are most often carried out on functional fragments from which the constant domain is removed. The first experimental goal of the project was thus to generate functional fragments of the antibody HPC-4 in quantity and purity sufficient for crystallization trials. Several examples of antigen-bound Fab crystal structures have been obtained by similar procedures[81,93-98], further supporting the feasibility of such a method to reach the goals defined by this project.

Established experimental conditions for the generation of Fab, Fab' and Fab'$_2$ fragments of antibodies are presented on figure 5. Briefly, cleavage of the antibody above the hinge region yield Fab fragments in a single step, whereas generation of the Fab' fragments involves cleavage of the antibody below the hinge region in a first step, followed by selective reduction of the hinge disulfides. Fab'$_2$ fragments can be

obtained by modifying the cleavage conditions so that the hydrolysis occurs below the hinge region, yielding a bivalent fragment. Although some evidence exists that the Fc regions of antibodies can have some effect on the antigen affinity of the binding site[99,100], it is generally accepted that the various antibody fragments generated by protease hydrolysis retain essentially the same binding properties as the antibody from which they derive.

A small number of proteases having broad specificity are able to cleave antibodies in relatively homogeneous fragments with antigen-binding capabilities. The choice of the correct protease to use is a delicate process, rendered more complex by the variability of the antibody sequences at the antigen binding site. By using a protease that is very specific for a given sequence or conformation, one ensures that the number of cutting sites on a given antibody will be limited, thus yielding homogeneous preparations of a small number of well defined fragments. However, the high specificity of a protease can also be a hurdle to overcome if the structure or sequence of the particular antibody has any unusual characteristic. Proteases with relatively low specificity such as papain or ficin are often selected for that purpose. On the other hand, the use of a low specificity protease must be accompanied with great caution to avoid the generation of inhomogeneous preparations following cleavage of the antibody at multiple sites. For instance, considering the reported amino acid preferences of the protease ficin[101,102], the antibody HPC-4 has a predicted ficin cutting site about every ten residues. As a result, the conditions for proteolytic cleavage of antibodies have to be adapted for each particular case due to the inherent variability of those molecules. However, some consistency in hydrolysis

conditions can be found among antibodies of a same isotype, due to their sequence similarities. In the case of the murine class IgG1 to which HPC-4 belongs, the use of the broad specificity protease ficin has been shown as optimal[103,104] although only a fairly small number of different IgG1 molecules were investigated. Digestion of IgG1 by pepsin has been shown to be the least efficient among IgG subtypes 2, 3 and 2a[105]. Based on these data and on the existence of a ficin-based kit specifically designed for Fab and Fab'_2 production from murine IgG, the initial condition for the digestion of HPC-4 was to use ficin.

The generation of Fab or Fab'_2 fragments of mouse IgG1 can be performed using a commercial kit from PIERCE ("ImmunoPure"). However, although suitable for most biochemical applications, this kit has several design aspects that are incompatible with protein crystallography. This kit consists of a hydrolysis column onto which ficin is immobilized. The fractions eluted from this column are then submitted to a protein A chromatography that retains the undigested antibodies, but lets the fragments that do not contain a constant domain flow through. By design, this kit would be expected to dilute the sample quite noticeably, as the species of interest are always recovered from the unbound fractions. This mode of operation is inherently devoid of the concentrating effects associated with the self-sharpening elution fronts resulting from specific and saturable interaction of the species of interest and the column[106]. Instead, it dilutes the sample by allowing it to diffuse freely into the buffer fronts that are used to displace the fragments of interest out of the ficin and protein A columns. In addition, the ficin column in the kit is operated by placing the sample onto the column, closing the column, and allowing the hydrolysis

to take place over a set period of time. This mode of operation limits the volume of material that can be cleaved to the void volume of the column, which requires starting with highly concentrated material or making many small batches, and thus increases the risks of variable cutting from batch to batch. In addition, the simple yet substantial operating parameter of enzyme concentration can not be optimized to control the extent of the hydrolysis reaction because it is fixed by the column properties.

1.5.2. Experimental considerations for the use of ficin in HPC-4 hydrolyses

Ficin is a thiol-protease with broad substrate specificity[101,102], and is activated by reducing reagents such as cysteine[101]. The use of reducing agents in the hydrolysis solution is recommended to obtain maximal activity of the protease. However in this particular case, increasing the concentration of reducing agent can only be done to a limited extent. Indeed, the particular substrate that is targeted in our case has a total of 17 disulfide bonds[91,107], some of which are essential for the integrity of the fragments that need to be recovered. Obviously, an intermediate level of reducing potential must be found that allows a reasonable activity of the protease but avoids a denaturing reduction of the substrate and products of the hydrolysis. It must however be considered that a mild unfolding resulting from partial reduction of the antibody may be beneficial in uncovering previously cryptic sites on the antibody for cleavage, as in several documented cases[108-110].

Over the course of the project, different reducing agents were considered to sustain the activity of ficin. For instance, glutathione (GSH) was expected to provide the same equivalent reducing groups as a cysteine solution, but located on a larger molecule. The resulting steric hindrance was expected to limit access of the GSH

reducing groups to the disulfides of the antibody. As a consequence, the resulting reduction was expected to be milder at similar concentration than the reduction that would have been obtained if the thiol groups were present on a smaller molecule. On the other end of the reducing spectrum, the bifunctional agent dithiothreitol (DTT)[111] was expected to provide more reducing efficiency at comparable concentration because of its small size and of the two thiol groups present on the molecule. Other reducing agents encountered in the literature include beta-mercaptoethanol (β-MA)[112], beta-mercaptoethylamine (β-ME)[113], metabisulfite or ascorbic acid[114].

The time and temperature also have a dramatic influence on the extent of the reaction. In particular, the temperature at which the hydrolysis is carried out has an important role by increasing the reaction rates through diminishing the activation energy differences according to the fundamental laws of thermodynamics. As such, temperature is an experimental parameter that can be used to operate on ficin activity.

The generated fragments were evaluated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in order to estimate their size and purity. The SDS-PAGE protocol involves a step where the protein sample is boiled along with the detergent. The presence of even a small amount of a reducing agent carried over from the hydrolysis reaction along with such high temperature induces a partial reduction of the generated fragments, and interferes with the estimate of size. Hence, it is necessary to perform the SDS-PAGE experiments both in native and in reducing conditions. For instance, the cleavage of the heavy chain can be monitored by following the disappearance of the 50kD band present in the reducing lanes. As will be seen, generation of the Fab fragments required some non-reducing lanes, as their

dissociation yields two 25kD fragments that can be indistinguishable from a cleaved heavy chain.

The integrity of the binding site of HPC-4 must be carefully monitored through the process to ascertain that the proteolytic step did not induce cleavage in this region. This consideration is paramount given the importance of structural integrity in the antigen-interacting site for our purposes. To that end, an affinity chromatography resin of immobilized HPC-4 linear epitope was used. In presence of calcium, the only fragments that are able to stay attached to the column during the washing step do so because their binding abilities to the epitope are undamaged. The choice of a hydrophilic material for the beads onto which the peptide is covalently immobilized ensures a low level of unspecific adsorption. Hence, this step represents an efficient way to control that the fragments generated are also functional. In addition, this resin efficiently separates the combining site-containing species from the various other fragments. The calcium-dependence of the antibody is used to elute the fragments with very mild elution conditions that include the calcium-chelating agent ethylene diamine tetra-acetate (EDTA). This is an advantageous feature of HPC-4, especially when considering the near irreversibility of other antibody-antigen interactions. This often requires the use of harsh elution conditions such as extreme pH or high concentrations of denaturing agents or chaotropes that are not always compatible with the easy recovery of properly folded proteins. Consequently, this column is very useful as a first step in the purification of the material. However, it is important to bear in mind that the Fab, Fab', Fab'$_2$ and full-length HPC-4 are equally

well retained. Hence, this step is essentially useless for the separation of these species, if they ever happen to be present in the same solution (see results).

Another important parameter for the reaction is the way it is stopped. Iodoacetic acid (IAA) is an alkylating agent that selectively modifies cystein residues, making it a useful inhibitor for ficin[101]. However, due to its oxidative properties, it has to be avoided because the combining sites of antibodies are rich in tyrosine[46] which are prone to IAA oxidation. Any covalent modification to the antigen binding site could alter the properties of the fragments, and needs to be carefully avoided.

In the case of the generation of the Fab'$_2$ fragments, the reaction is stopped by adding the sulfhydryl-specific alkylating agent N-ethyl-maleimide (NEM). The free cysteines as well as those present in the active site of ficin are quantitatively reacted by NEM, leading to an efficient arrest of the reaction. NEM was used on several occasions for similar puposes, and no spurious oxidation of tyrosine residues was reported[105,115].

2. MATERIAL & METHODS

2.1. Generation of HPC-4 fragments

2.1.1. Optimizations of HPC-4 hydrolysis and purification

The formation of protein crystals for structural analysis is a process that requires extremely homogeneous protein preparations. This requirement stems from the fact that the cohesive forces inside a crystal arise from the interaction of precise surface patches on each molecule within the crystalline lattice. Hence, an inhomogeneous protein preparation can not yield well diffracting crystals with a reasonable likeliness because the surfaces susceptible to form crystalline contacts are different from one molecule to the next, thereby preventing their ordered assembly. Thus, impure protein solutions often are associated with increased tendencies to assemble in a disorganized fashion (precipitation) in conditions of low solubility, rather than to form crystals. In addition, if one of the species actually forms a crystalline lattice, another species may act as inhibitors to crystal growth in the case where they can add onto the lattice but do not provide additional interaction surfaces. As a consequence, protein purity and homogeneity are paramount in protein crystallography. Because the different species of an inhomogeneous preparation may crystallize in very different conditions (a phenomenon used in some industrial purification processes[106]), inhomogeneous preparations also amount to smaller concentrations of each individual species, which would impair their tendency to

crystallize. Taken together, these reasons explain why sample purity and preparation homogeneity are much more important in protein crystallography than in most other biochemistry applications. A significant portion of the project was thus devoted to the development of robust hydrolysis protocols able to generate high amounts of pure and homogeneous HPC-4 fragments that would crystallize more readily.

The methodology that was followed was to test different hydrolysis methods in a systematic fashion. The nature of the fragments that were generated was assessed by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and their function was tested using an affinity chromatography using immobilized antigenic peptide. Although a large number of conditions were assayed over the course of the project, we will only present the final, optimized protocols to obtain the Fab'$_2$ and the non covalently associated Fab ("ncFab"). A condensed account of the steps having yielded these methods is given in the "results" section.

2.1.2. Generation of the Fab'$_2$

Ficin hydrolysis. Ficin (1.97mg/mL) was preactivated by incubation in 1.1mM cysteine for 30 minutes at room temperature as specified by the manufacturer. The activated ficin was then added to a HPC-4 solution that contained cysteine. The final concentrations were 0.386mg/mL (7.72µM) HPC-4, 1mM cysteine, 19.5µg/mL (0.78µM) ficin, 50mM Tris pH7.5 and 2mM EDTA pH 7.5, representing a molar ratio of approximately 10:1. The hydrolysis solution was allowed to incubate on ice at 4°C for 5 days. The reaction was stopped by addition of the alkylating agent N-ethyl maleimide (NEM) to a final concentration of 10mM. The solution was brought to

4mM $CaCl_2$ so that the EDTA was neutralized and that enough calcium would be available for binding to the peptide affinity chromatography column.

Peptide affinity chromatography. The hydrolysis solution was adsorbed to an affinity chromatography medium onto which the antigen peptide was covalently immobilized by primary amine-reactive chemistry. This adsorption took place directly on the packed column for 1hr at room temperature. Following adsorption, unbound material was removed by flowing equilibration buffer (400mM NaCl, 20mM Tris pH 7.5, 4mM CaCl2, 0.02% Na azide) at a very slow rate until the OD at 280nm of the outlet was not detectably higher than the OD of the equilibration buffer. Then, elution buffer was fed to the column (400mM NaCl, 20mM Tris pH 7.5, 4mM EDTA, 0.02% Na azide) while collecting 300μL fractions. The amount recovered was estimated by measuring the absorbance at 280nm and assuming a mass absorption coefficient of 1.72ml/(mg·cm). Fractions containing the highest amounts of protein were pooled for further use. Usual recoveries were in the 60-70% range.

Protein A affinity chromatography. To remove the undigested HPC-4 molecules, the elution fractions of the peptide affinity chromatography were pooled and submitted to protein A affinity chromatography (PIERCE). Protein A is a bacterial protein with high affinity for the constant portion of antibodies, and is thus ideal for this particular application. The chromatography took place as indicated by the manufacturer. Briefly, the column was equilibrated in PIERCE binding buffer (composition not disclosed to customers), then the sample was absorbed. The column was then washed using the binding buffer, and the HPC-4 Fab'_2 fragments were recuperated from these fractions. The column was regenerated by eluting the bound full-length HPC-4 using PIERCE

elution buffer (0.1M citrate pH3.0). Because they do not bind to the column, the Fab'$_2$ fragments were significantly diluted by this purification step. Usual recoveries were in the 60-70% range (36-49% if considering the peptide affinity chromatography purification step).

2.1.3. Generation of the ncFab

Ficin hydrolysis. The method was essentially the same as the one used to generate the Fab'$_2$, with the only differences being the reducing agent (1mM dithiothreitol instead of 1mM cysteine) and the duration of the reaction (4 days instead of 5 for the Fab'$_2$).

Peptide affinity chromatography. The hydrolysis solution was batch-adsorbed to an affinity chromatography medium (20mL bed volume) onto which the antigen peptide was covalently immobilized. This adsorption took place under mild agitation for 3 hours at room temperature. The duration of the adsorption was chosen to be much longer than the typical timescale of antibody binding kinetics to optimize the amount of material bound to the resin. Following adsorption, the loaded resin was packed into a glass chromatography column. Unbound material was removed by flowing equilibration buffer (400mM NaCl, 20mM Tris pH 7.5, 4mM CaCl2, 0.02% Na azide) at a rate of 2mL/min until the OD at 280nm of the outlet was not detectably higher than the OD of the equilibration buffer. Then, elution buffer was fed to the column (400mM NaCl, 20mM Tris pH 7.5, 4mM EDTA, 0.02% Na azide) at a rate of 0.2mL/min while collecting 1.4mL fractions. The amount recovered was estimated by measuring the absorbance at 280nm and assuming a mass absorption coefficient of

1.72ml/(mg·cm). Fractions containing the highest amounts of protein were pooled for further use. Usual recoveries were in the 70-80% range.

2.2. Generation of protein C variant

2.2.1. General considerations

The crystallization of HPC-4 and HPC-4 fragments was also attempted in the form of complexes with protein C. As a vitamin K-dependent blood coagulation factor, protein C has a gamma-carboxyglutamic acid domain in the N-terminal portion of the light chain[116]. The chelation of calcium by the Gla domains of coagulation serine proteases is thought to mediate interactions with injured cell membranes, thus assisting in the spatial localisation of coagulation reactions[117,118]. In our case, the Gla domain was removed from the protein C used in our studies to assist in its crystallization by decreasing the number of separate domains of the protein. Importantly, the removal of the Gla domain was shown not to influence the kinetics of the activation by the thrombin-thrombomodulin complex in solution[85] or the binding to HPC-4[41]. For that purpose, a variant of protein C that is devoid of the gamma carboxyglutamic acid (Gla) domain, GDPC, was used. Initially obtained though proteolysis[85], this variant can also be generated by plasmid expression in a mammalian cell host, as the material used in our study.

2.2.2. GDPC production methods

GDPC was expressed in HEK-293 mammalian cells expressing the sequence from a PL-4 plasmid, a generous gift from Dr. A. R. Rezaie. The cells were grown on the surface of a 10 stack Nunc cell factory (Sigma-Aldrich) with an operating volume of 2L of media supplemented with 5 to 10 % of bovine serum at 37°C under 5% CO_2. Every two days, the protein-rich media was drained from the factories and replaced with fresh media under sterile conditions. After collection, the collected media were supplemented with the reversible protease inhibitor benzamidine and sterilized by addition of sodium azide. Cells that had detached from the factory surface were removed by a "soft" centrifugation for 30 minutes at 3800rpm (4000×g at the bottom of the tubes). The supernatant of this centrifugation step was submitted to a "hard" centrifugation (30 minutes at 10000rpm, or 15300×g at the bottom of the tubes) to remove all cellular debris. The supernatant of the second centrifugation was pooled in 4L batches (1 batch each 2 collections, or 1 batch every 4 days) and stored at 4°C until purification.

GDPC was extracted from the cell culture media using an affinity chromatography column where the antibody HPC-4 was covalently immobilized onto the chromatography beads. After adding calcium to a final concentration of 10mM, the media was batch-adsorbed with equilibrated resin overnight at 4°C under mild stirring (80mL media/mL resin; 50mL resin). The resin was then allowed to sediment for one hour, after which the cleared supernatant was removed. The remaining slurry was packed onto a glass column under steady flow of equilibration buffer (100mM NaCl, 20mM Tris pH 7.5, 3mM $CaCl_2$, 10mM benzamidine, 0.02% sodium azide).

After equilibration in 3 bed volumes, the adsorbed protein was submitted to a high salt wash of 2 bed volumes in high salt buffer (1M NaCl, 20mM Tris pH 7.5, 3mM $CaCl_2$, 10mM benzamidine, 0.02% sodium azide). The salt concentration was then brought back down and the benzamidine removed by extensive wash (3-5 column volumes) into low salt buffer (25mM NaCl, 20mM Tris pH 7.5, 3mM $CaCl_2$, 0.02% sodium azide). The extent of this last step was essential to remove any trace of benzamidine. Indeed, this compound has a very high absorption coefficient at 280nm that interferes readily with protein concentration estimates done spectrophotometrically. The GDPC was then eluted off the column by switching to a very mild elution buffer that differs from the previous buffer only by the presence of EDTA instead of calcium, and that is generally extremely mild compared to conventional affinity chromatography elution buffers. The elution (25mM NaCl, 10mM EDTA pH 7.5, 20mM Tris pH7.5, 0.02 sodium azide) occurred at a flow rate of 8.2mL/hr in 5mL fractions. The protein concentration of the fractions was then evaluated by measuring their absorbance at 280nm with a 1cm light path, using an absorption coefficient of $1.6 mL \cdot mg^{-1} \cdot cm^{-1}$. Fractions with an absorbance greater than 0.04 were kept at 4°C until further purification.

Once an estimated 10mg of GDPC were accumulated, the corresponding fractions were pooled and submitted to a mono-Q ion exchange step using an Äkta FPLC chromatography system (GE healthcare). The protein pool was loaded under a 1mL/min flow rate (25mM NaCl, 20mM Tris pH7.5, 0.02% Na azide) while monitoring the 280nm absorbance of the eluate to detect column overloading. The column was then submitted to a linear gradient of NaCl concentration, sweeping from

25 to 600mM in 30mL under constant flow rate. Both the absorbance and the conductance of the eluate were monitored during the elution. 500µL fractions were collected, and peak fractions were pooled and stored at 4°C until formation of the complex for crystallization.

2.3. Characterization of HPC-4 fragments

2.3.1. Fluorimetry

The fluorescent properties of the naturally occurring amino acid tryptophan (and, to a lesser extent, tyrosine and phenylalanine) are often used as conformation probes in protein studies. Indeed, the fluorescent emission spectrum of these residues depends on the polarity field that surrounds them, which is modified whenever conformational changes occur within a given range. The measure of fluorescence intensity changes at a fixed wavelength or as fluorescence spectrum is often used to monitor an event that modifies the conformation of a protein. This technique has the advantages of simplicity of the setup and of high sensitivity. It is however dependent on the presence of a tryptophan residue in the vicinity of the conformational event, which must be present naturally or added by mutagenesis. In the case of HPC-4, the fluorescence emission between 305 and 400nm was shown to depend on the calcium concentration when a 12-mer epitope peptide is present[41]. To assess the function of the ncFab fragments, fluorescence measurements were performed in increasing concentrations of calcium and in the presence of various concentrations of the 9-mer epitope peptide used in the crystallization studies.

Fluorescence measures were taken on a PC controlled SLM 8000 fluorescence spectrometer with high voltage parameters of 950V and 400V for the signal and reference channels, respectively. The titrations were carried out using a mixture of 1μM ncFab and 0, 0.1, 0.5, 1.0, 2.0 or 4.0μM of epitope peptide. All solutions were made in 10mM tris pH7.5, 0.02% Na azide using chelex treated water to allow a strict control over the calcium concentration. The solutions contained 6kD polyethylene glycol to a final 0.1% to minimize the loss of signal by nonspecific adsorption on the walls of the cuvette. Titrations were performed by adding small volumes of concentrated calcium solutions to the cuvette inside the spectrometer. The excitation wavelength was set at 280nm, and the emission intensity was measured in a 16nm-wide wavelength window centered around 340nm. With this setup, count rates were in the 20000cpm range, allowing a good sensitivity while operating well under the 50000cpm maximum imposed by the saturation limit of the photomultipliers. After each addition of calcium, the sample was mixed by pipette reflux and allowed to equilibrate for at least 3min. The fluorescence intensity was measured during three 1 minute periods with sample mixing in between and sampled at one measure every 3 seconds (60 datapoints minimum). Shutters were closed at all times except while measuring to ensure minimal photobleaching. The signal intensity obtained with closed shutters was averaged over the length of the experiment and removed from the signal values with open shutter (black noise removal). The signal intensity was averaged for each calcium concentration, and corrected for dilution assuming a linear dependence between the ncFab concentration and the fluorescence intensity (Beer-Lambert law adapted to fluorescence, valid for concentrations below 10^{-5}M). The

fluorescence value F obtained at the various calcium concentrations was divided by F_0, the fluorescence obtained in the absence of calcium. The value $F/F_0 - 1$ was fitted to a single site binding model, as recommended[119]:

$$\frac{F}{F_0} - 1 = \frac{\left(F'_{\max} - 1\right) \cdot [Ca^{2+}]}{K_d + [Ca^{2+}]} \tag{1}$$

Specifically, equation (1) was linearized by considering the reciprocal:

$$\frac{F_0}{F - F_0} = \frac{K_d}{F'_{\max} - 1} \cdot \left(\frac{1}{[Ca^{2+}]} \right) + \frac{1}{F'_{\max} - 1} \tag{2}$$

The experimental data (plotted as $F_0/(F - F_0)$ against $1/[Ca^{2+}]$) were submitted to linear regression using equation (2), and the value for K_d was calculated as the ratio of the slope $K_d/F'_{\max} - 1$ and the intercept $1/F'_{\max} - 1$. This procedure was repeated at all epitope concentrations.

2.3.2. Modeling of HPC-4

Further characterization of HPC-4 fragments was carried out *in silico* by building a molecular model. The immunoglobulin fold is characterized by a very small variability in sequence, which transcribes into accordingly small structure variability. Apart from the hypervariable loops and the elbow angle between the variable and constant domains, the structures of antibodies are usually quite similar. Thus, it was decided to map the sequence of HPC-4 on to the structure of a

resembling IgG1 in order to provide working hypotheses about HPC-4 function in the form of an approximate molecular model. The model was never intended to provide any reasonable accuracy at the atomic level, but rather to coarsely map out the positions of certain unusual features of the HPC-4 sequence. In particular, two stretches of electronegative residues are present in the sequence of HPC-4 at the CDR-H2 and CDR-H3. These were examined in more details to evaluate their potential as calcium binding sites. The antigen was not included in the modeling because no satisfying templates are available. An important consequence of this shortcoming in the modeling process was that by construction it would not be informative on the possibility of shared calcium coordination between HPC-4 and the antigen. However, before the discovery of the experimental structure described in the present work, the hypothesis of full coordination of calcium by the antibody was considered plausible, based on terbium fluorescence results[41] and on the apparently extremely low affinity of the epitope peptide alone for calcium. In this hypothesis, the binding of calcium to the antibody was believed to impart conformational changes in the antigen binding region resulting in a form with high affinity for the antigen, thus explaining the strong and mutual requirement for both calcium and the antigen to get ternary complex formation.

The antibody Mab61.1.3[91] (pdb code 1IGY) was used as template because its sequence is relatively close to that of HPC-4 (90.7% and 74.8% for the light and heavy chains, respectively), with most of the variations occurring as expected in the variable domains. First, the sequences were aligned using an EBLOSUM62 matrix and the MatcherP program[120]. Then, using the program "MAIN"[121], the residues in

the Mab61.1.3 structure were successively changed into their counterpart in HPC-4 as predicted by the sequence alignment. When gaps or insertions occurred, backbone rearrangements were carried out, and the resulting segment was energy minimized.

Because of its hypothetical nature, and because the predictions it generated were partially proved wrong by the crystal structure, the model obtained will only briefly be described here, and will not be the object of a specific results section. Two candidate calcium binding sites were identified. The first involved the triplet of aspartic acid residues in the CDR-H2 region, which were modeled in a loop that would allow calcium binding without much backbone rearrangement. The presence of calcium at this location readily explained the tryptophan intensity changes that are observed upon calcium binding[41], because the triplet is located immediately after a doublet of tryptophan residues. The second identified possibility for a calcium binding site was the crevice between the CDR-H1 and the CDR-H3. Although it was less consistent with the fluorescent properties of HPC-4, this hypothesis was closer to the truth, as the crystal structure later revealed (see results section).

2.4. Crystallization

2.4.1. Preparation of the complexes and buffer exchange

Formation of the complexes. HPC-4, HPC-4 Fab'$_2$ and HPC-4 ncFab were mixed with solutions of GDPC or of the antigenic peptide so that molar ratios of 1.1 GDPC/binding site or 1.5 antigenic peptide/binding site were obtained. The valency

of the antibody or antibody fragments was taken into account. The complexes were allowed to form in the presence of 15mM $CaCl_2$.

Buffer exchange. Crystallization trials (screening) require protein solutions that have a salt content barely above the level required for solubility and that are very mildly buffered. Hence it was necessary to buffer-exchange the affinity chromatography elution pool into a "screen" buffer (20mM NaCl, 5mM TRIS pH 7.5, 5mM CaCl2, 0.02% Na azide). A 25mL sephadex-based G-25M column that was extensively equilibrated in screen buffer was used for that purpose. Because the size of the pores on these beads is much smaller than even a small protein, but much greater than the buffering components, it acts as a de-salting step where buffer components are held up inside the pores, whereas protein are eluted much faster because their diameter prevents them from entering the pores. The complex pools were separated in fractions representing 40% of the resin volume. Each aliquot was loaded onto the resin, and eluted slowly while collecting fractions. The fractions having the highest absorbance at 280nm were pooled together and stored at 4ºC. Recoveries were usually in the 90-100% range. Overall protein recoveries were about 42-53%, or 63-80% when computing in the necessary loss of the constant fragment of the antibody.

2.4.2. Concentration

Protein solutions destined to be assayed for crystallization generally need to be in concentrations in excess of 10 g/L. The protein complexes obtained as described previously were submitted to centrifugal filtration on a 10kD MWCO spin filter (Millipore). This filter has a molecular weight cutoff sufficiently small to prevent

protein complexes from passing through. During operation, the quantity of protein atop the filter stays constant, while the volume decreases as the buffer is filtered through the pores by the effect of centrifugal forces, resulting in an increase in concentration of several fold magnitude. Concentration factors in the range of 60 were routinely obtained with this setup, and allowed to reach protein concentrations in the order of 15-20 g/L. The diluted pools typically consisted of 30mL at an OD of 0.45-0.50, and were centrifuged down to about 500µL.

To avoid nonspecific, irreversible adsorption of the protein sample on the filter, it is sometimes recommended to first pass a solution of dilute gelatin to block the nonspecific adsorption sites of the filter. In our hands however, it was found that the small loss of protein on the filter was preferable to the introduction of a protein contaminant associated with the gelatin pretreatment. This was especially important considering the importance of sample purity in protein crystallography. Furthermore, it was noticed that the maximum flow rate attainable was severely decreased with a pretreated filter. Consistently, the flow rate of the filtration decreased while the protein concentration was increasing. This observation could result from two effects[106]. The first effect is the increase in retentate viscosity associated with the increasing protein concentration. The second effect that could explain this observation is concentration polarization. This phenomenon results in a localized gradient of protein concentration where the highest concentration is closest to the filter. In practice, it seemed that a film of highly concentrated protein was present on the filter at the end of the centrifugation step. One of the technical difficulties in using this type of filter is to recover the retentate quantitatively. Most of it can be recovered,

but a thin film always remains close to the filter. Because of the phenomenon of concentration polarization, one would expect that this thin film would be where the protein is the most concentrated. Hence, the concentration gradients still remaining above the surface of the filter were disrupted by mix-flowing the retentate with pipette refluxes, thereby equalizing the protein concentration in the bulk and in the thin film remaining after sampling. It is believed that this simple procedure can substantially improve the recovery of centrifugal filtration steps.

2.4.3. Crystallization

Crystallization trials were performed using the hanging drop diffusion method. This experimental setup consists of a microwell plate (24 wells per plate) in which various crystallization solutions can be placed. The protein, initially present in very mildly buffered water (20mM NaCl, 5mM tris pH 7.5, 5mM $CaCl_2$, 0.02% Na azide), is diluted 1:1 with the crystallization solution to be assayed. A 2uL drop of this diluted solution is deposited on a glass coverslip, and this coverslip is inverted and placed on top of the well and finally sealed using vacuum grease. After a period of time, a physico-chemical equilibrium is reached while water or other volatile molecules contained in the protein drop pass into the well solution through the vapor phase. Because of the great volumetric excess of the well compared to the protein drop (480µL:2µL in our case), the protein ends up in a solution whose composition closely approximates that of the well solution.

Importantly, this technique permits to determine the solubility behavior of the protein in a wide variety of crystallizing conditions without requiring any protein-

consuming step of buffer exchange. An additional benefit of this technique is that it is easy to follow the changes in solubility by observing the drops with a microscope. In our case, the plates were examined immediately following setup, then daily for a week, then about weekly for longer periods of time. A Leica 95 binocular microscope equipped with polarizer was used for the observations. Because most of the drops equilibrate over days, the first week is the most important period to monitor carefully. Observations such as the type of response (heavy/light precipitate, phase separation, clear drop, presence of crystals) or the size of crystals if applicable were recorded. Interpretation of the drop results was considered with respect to biophysical considerations about protein crystallization (see results).

2.4.4. Crystallization conditions database

Crystallization trials typically take hundreds of trials before a successful or even promising crystallization condition is found. In order to be able to keep track of all tested conditions in a quick and easy way, a program was created ("crystal_plates.py"), written in the python programming language (the source code is available upon request). As an input, this program takes a series of text files (one file for each crystallization plate) in which the composition of each well of the plate and other relevant information is written under precise but intuitive syntactic rules. Each time a new plate was set up, a corresponding file was written and saved. The python program actually consists of a graphical interface for easy retrieval of all the information relevant to a particular plate. In addition, the program can compile lists of recorded conditions matching a user-specified request from the data available to it on

all plates. The first benefit of this program was its ability to display the content of all wells of a specified plate in a friendly format. This feature was used to generate tidy and synthetic recording sheets on which microscopy observations were then reported. Analysis of the microscopy observations was then straightforward because the corresponding well buffer compositions were available at a glance.

The second and most useful feature of this program was its ability to search across all recorded plates for conditions that matched a request input by the user, and to output a list of the corresponding wells locations and composition (fig. 10). It is often the case that a small subset of conditions with a particular property (presence of a specific precipitant or precipitant combination, specific pH value or range,...) gives rise to a similar outcome in the protein drops of a single plate. It then becomes of interest to know which of the conditions tried on other plates also belong or resemble to this subset, and to assess whether the protein behaves with some kind of consistency across this subset. However, with hundreds of trials, this task is often impractical to do by hand, and can easily yield inaccuracies. The automated compilation of lists of conditions matching a user-specified request is an attractive

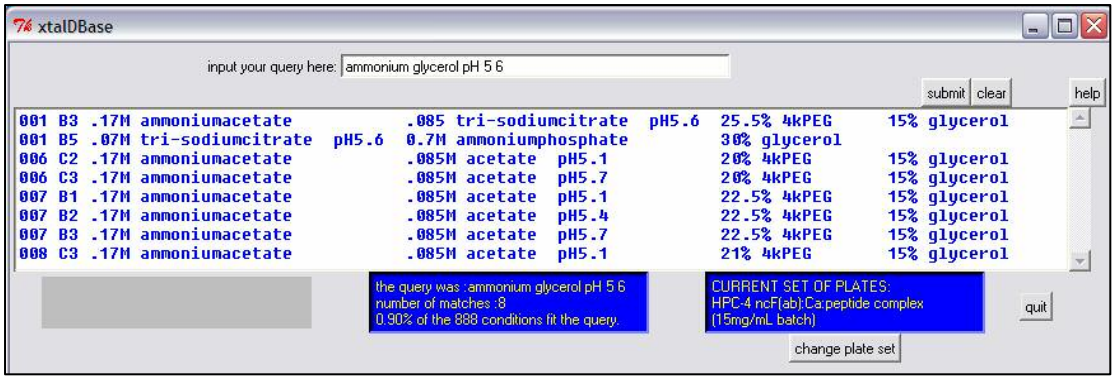


Figure 10: Example output from crystal_plates.py. The program is displaying conditions containing ammonium and glycerol and a pH comprised between 5 and 6.

solution to this issue, and was used in several instances to rapidly map out phase diagrams that synthesized information collected from several crystallization plates.

It is often necessary to quickly compile a list of all conditions that contain a particular precipitant, or that have a certain pH value in order to identify trends in the solubility behavior of the protein. For instance, if a small subset of conditions hints that the protein may always precipitate when placed at pH values smaller than 5, it is very useful to be able to quickly generate a list of all conditions that have such a pH value.

2.4.5. Macroseeding: theory & methods

The crystals obtained, described in "results", were still too small to likely diffract X-rays in a significant manner. After having exhausted most of the usual optimizing methods for the growth conditions, we decided to attempt to apply the macroseeding technique[122,123]. Macroseeding can be understood by considering crystallization as the two sequential events of nucleation and growth. The first event refers to the random formation of an assembly of several protein molecules into organized crystal nuclei of a critical size. The second event, growth, can occur if the solution is in a supersaturated state, corresponding to a protein concentration higher than the equilibrium solubility. If this is the case, then the presence of a nucleus can trigger the return to equilibrium by providing the energy barrier required to overcome supersaturation. In macroseeding, the critical nuclei are introduced into a drop of supersaturated protein ("target drop") in the form of an already formed and individual crystal ("seed"). Crystals were transferred by using a nylon loop of a few hundred μm

in circumference. The addition of a seed to a target drop triggers the return to equilibrium by allowing supersaturated protein molecules to add onto the surfaces of the crystal that is provided, subsequently increasing its size. The procedure can be applied iteratively until the desired crystal size is reached. Prior to transferring the crystal to the target drop, it is recommended to mildly dissolve it ("etching") so that the shell of protein that was added when the crystal stopped growing is removed. This procedure is believed to improve the characteristics of the surface of the seed so that protein molecules of the target drop can adsorb more easily.

The volumes of the growing crystals being macroseeded were measured over time. This data was applied to a mathematical model of macroseeding in order to derive a growth rate and other parameters characteristic of our system. The corresponding derivation is available in appendix X.1..

2.4.6. Cryoprotection methods

The collection of protein crystal diffraction data is done at temperatures of about 100K for two main reasons. The first is to slow down the damaging process imparted on the protein molecules of the crystal by the radiative energy of the X-ray beam[(124)]. The second reason is that in most of the cases, decreasing the temperature actually improves the diffraction characteristics of a given crystal. However, the water present in the crystal drop tends to form an expanded lattice of organized molecules upon freezing, which can be damaging to the protein lattice. To avoid the extraneous diffractions spots associated with the formation of ice, it is necessary to transfer the crystal into a solution able to freeze as amorphous glass rather than as crystalline ice. This process is referred to as "cryoprotection". Several molecules can

be added to a water-based solution to cryoprotect it, such as glycerol, low molecular weight polyethylene glycol (PEG), alcohols or polyols. The method by which the crystal is transferred from the growth solution to the cryoprotecting condition also has some importance. It is possible to transfer it directly to a cryoprotecting solution drop, to have the crystal grow in a cryoprotecting condition, or to dialyze the growth solution against a cryoprotecting solution. In the case of direct transfer, the optimal timescale for transfer can vary from seconds to days, and the precise sequence of solutions through which the crystal is passed before being put in the final cryoprotecting solution has great importance on the outcome[125]. Because all of these effects are mutually acting on the resulting diffraction characteristics of a particular crystal, the determination of the optimal cryoprotecting condition for a given crystal has to be done empirically.

The evaluation of a cryoprotecting protocol is often performed merely by assessing its ability to maintain apparent crystal integrity and by the quality of a diffraction snapshot obtained after cryoprotection. However, this method lacks proper control, since the quality of the obtained snapshot results not only from the cryoprotecting protocol that is assessed, but also from the inherent quality of the crystal. Indeed, even though two protein crystals can have identical aspects under a microscope, their diffraction pattern can vary widely, even in some cases for crystals that were grown in the same drop. As a result, a good cryoprotecting method could be discarded on the grounds of a poor diffraction pattern, when in fact it was the inherent quality of the crystal that caused the poor diffraction. To abrogate this source of variability, we used a two-step methodology to optimize the cryoprotection process.

A crystal was first mounted in its growth solution using a MicroRT mount system (MiTeGen, Ithaca NY) and an X-ray diffraction pattern was recorded at room temperature. The quality of the snapshot was used as an indication of the inherent quality of the crystal, thereby suppressing the crystal variability encountered when comparing cryoprotections performed on different crystals. The crystal was then submitted to the particular cryoprotection procedure being tested, after which an X-ray diffraction pattern was recorded at 100K. The magnitude and nature of the differences observed between the diffraction of a given crystal before and after cryoprotection were used to compare the cryoprotection methods with one another without interference from crystal quality variablity.

Because none of the conventional cryoprotectants offered satisfactory results, we decided to investigate the small organic ion malonate. Although initially mentioned as a salt with uncommon propensity to crystallize proteins[126], malonate later became recognized as a cryoprotectant that is easy to use for protein crystals that are grown in high concentrations of ions[127]. Malonate was expected to provide good results in our particular case, where the protein crystals had been grown in ammonium sulfate concentrations in excess of 1.6M.

2.5. Structure determination

2.5.1. Initial diffraction studies

Prior to the X-ray diffraction dataset that took place at the CHESS synchrotron, all X-ray related experiments were performed on the Rigaku rotating

anode copper X-ray generator of the Crystallography Program at the Oklahoma Medical Research Foundation. This device is equipped with a MAR345 laser readout image plate detector of 345mm in diameter. Although it is less powerful than a synchrotron source, this generator was amply sufficient for initial work. For shots made at room temperature, the loop containing the crystal was mounted on the goniometer, and exposed for 10 minutes.

The most direct method to determine whether a crystal seen in a drop is protein or salt is to mount it on a loop and observe its X-ray diffraction pattern. Because their unit cells are small, salt crystals typically yield very few diffraction spots which are separated from one another by great distances. On the opposite, protein crystals that have very large unit cells (commonly in the order of 10nm or more) yield numerous diffraction spots that are close to one another. Thus, X-ray diffraction can actually be used as a relatively quick assay to determine the nature of a crystal. In addition, it is advantageous because in the case of protein crystals, it provides initial estimates of the quality of the packing, the dimensions of the unit cell and the number of protein copies per unit cell. Over the course of the project, several apparently nice crystals were discarded on the ground of a poor diffraction pattern.

It has been reported in some instances that even though a protein crystal has a nice morphology as seen under a microscope, its diffraction characteristics are still poor. This can often be attributed to defects in the crystal packing that are not apparent using simply a microscope, such as crystal twinning on a microscopic level, or to crystal mosaicity. Considering this, testing a crystal on an X-ray diffractometer is so far the only valid test to determine how promising one crystal is in terms of its

potential use for structure determination. The choice is often difficult to make, because at this stage the crystals are not optimal in size, and hence even if they are packed perfectly, they are not expected to give rise to optimal diffraction patterns. Still, in our case the earliest diffraction patterns we obtained were searched for signs of twinning or generally bad packing quality. Importantly, the in-house diffractometer was also used to obtain accurate estimates of the space group of the crystal, the unit cell dimensions, and of the number of proteins present in the asymmetric unit. The minimum resolution attainable was also determined. Indeed, for reasons inherent to the physics of crystal diffraction, the highest resolution spots generally diffract with the least intensity. Since the average intensity of diffracted spots is directly related to the intensity of the incident beam, a crystal will diffract at a higher resolution at a synchrotron than at a rotating anode generator. Hence, the maximum resolution obtained at the OMRF X-ray diffractometer was used as a lower bound to the expected resolution of the synchrotron dataset.

Because it is sufficient for dataset collection, this X-ray source was also amply sufficient for evaluating the damages undergone by a crystal during a cryoprotection procedure. For instance if after cryoprotection a crystal is still intact enough to give an indexable diffraction shot, the unit cell dimensions were measured and compared to the values at room temperature. Any significant deviation from room temperature values was considered as indicative of an important rearrangement of the crystalline lattice. In general however, damages incurred during the cryoprotection were so extensive that the resulting diffraction pattern could not even be indexed.

Finally, when the cryoprotection method was optimized, crystal candidates were screened on the rigaku diffractometer to determine which were the most promising and deserving of being sent to the synchrotron. That way, the time-consuming step of crystal screening was already made when we arrived at the synchrotron, and the limited time available at the beamline was entirely used for data collection.

2.5.2. Handling of crystals at cryogenic temperatures

For diffraction images made at 100K (at the synchrotron or at the rotating anode diffractometer), a propane-freeze procedure was adopted. Briefly, a bath of liquid nitrogen was setup, in which a 15mL eppendorf and a cryovial were put. Care was taken not to drop liquid nitrogen in the vessels. Gaseous propane was then blown into the 15mL eppendorf to condense it at liquid nitrogen temperature. At atmospheric pressure, liquid nitrogen is at its boiling point of 77.36K, or -195.8°C, which is just below the melting point of propane (85.5K, or -187.6°C). The propane blown was cooled by the walls of the 15mL tube, thus liquefying it for a short period of time (approximately 3-4 minutes) before freezing. During this brief period, the liquid propane was transferred to the cryovial, and the crystal mounted on the loop was rapidly immersed in the cryovial. The cryovial was then left in the nitrogen bath, effectively encasing the crystal in a shell of frozen propane. The crystals that are frozen according to this procedure can be stored in liquid nitrogen for extended periods of time. Specially insulated and secured dewars allow shipping the crystals at liquid nitrogen temperature. To mount such cryoprotected crystals, a thin layer of

propane was melted between the cryovial and the propane shell encasing the crystal loop by contact with the gloved hand. The vial containing the loop was then immediately attached onto the magnetic spindle (the loop is mounted on a magnetic base). Because of the thin layer of liquid propane lining the cryovial, the crystal loop (attached to the spindle) is now unfastened from the vial, but still encased in a shell of melting propane. This step is very important, because if the vial is removed while still fastened to the loop by solid propane, there is a risk of detaching the loop from the mount and to lose the crystal. The vial is then removed, exposing the frozen propane shell to the gaseous stream of 100K nitrogen. This procedure ensures that the crystal is maintained under 100K at all times because it is placed under the cryocooling system before the propane is fully melted. The temperature variations undergone by the crystal are thus minimized.

2.5.3. Data collection

The interaction of light with the electrons of a protein crystal provides the information about the structure of the molecules present inside of it. In order to collect information on a length scale relevant to protein structure (10^{-9}-10^{-10}m), it is required to impart a radiation of matching energy, which corresponds to X-rays (10^{-8}-10^{-11}m). Mostly two technologies exist to generate X-rays for macromolecular crystallography, which differ by their flux, brightness and brilliance. These characteristics are important because protein crystals are very small in size, and it is necessary to concentrate the radiation on a space of size comparable to that of the crystal. Furthermore, high resolution spots diffract with the least intensity, and the

most direct method to increase the resolution attainable with a given crystal is to use a greater impinging X-ray beam intensity.

As any electromagnetic radiation, X-rays can be obtained by accelerating charged particles. In a rotating anode type generator, the acceleration happens when a flux of electrons is stopped by a metallic anode within a vacuum tube. However, the resulting emission is anisotropic, which limits the flux of photons arriving on a particular area, and necessitates the use of optical elements to obtain proper collimation of the radiation on to the sample. The alternative to a rotating anode is to use the synchrotron radiation as an X-ray source. A synchrotron consists of a circular vacuum vessel of several hundred meters in circumference in which charged particles (electrons or positrons in this case) travel with a nearly relativistic velocity. To avoid collision of traveling particles with the walls of the synchrotron, it is necessary to accelerate them on a centripetal direction. It is this acceleration which results in the emission of X-rays in a geometrically well defined envelope, resulting in high brightness, flux and brilliance. Additionally, the electrons or positrons can be accelerated along a periodic trajectory when traveling down a straight section using special electromagnets, or "insertions devices" (wigglers and undulators). The energy spectrum of synchrotron radiation depends among other factors on the radius of curvature of the path of the particles, their velocity, and on the strength of the magnetic field used to provide the acceleration. With the setups of current facilities, synchrotron radiations have a very high flux, and are inherently very bright and collimated because of the precise control allowed over the trajectory of the electrons. This is very advantageous compared to rotating anode type generators for X-ray

crystallography applications, because it is necessary to irradiate the small volume of the crystal with the maximum amount of X-ray photons.

After necessary preparation[128], data collection took place at the Cornell High Energy Synchrotron Source (CHESS) at Cornell University, Ithaca, NY. We requested a time allocation at the F-1a hutch because it is well suited for crystals of small size. This particular hutch is located on a 24 pole positron wiggler insertion device. The X-ray beam delivered to F-1a has a rectangular shape of 3.76mm in width and 0.944mm in height, with a radial divergence of 1mrad (horizontal) × 0.4mrad (vertical). The flux arriving at the sample is 3.0×10^{11}photons/sec. The diffracted rays are imaged with an ADSC Q4 charge coupled device. This hutch is equipped with a conventional cryocooling system with auto-refill capability. The spindle and the goniometer are fully servo-controlled, and the crystal centering is assisted by computer. The crystal centering program developed by CHESS displays a live image of the crystal overlayed with the position of the X-ray beam symbolized as a cross hair (fig. 11). By double-clicking on the position of the crystal at several spindle angles, the program sets the goniometer axes so that the crystal is always irradiated by the beam regardless of the spindle angle. Image collection is described in details in the results section.

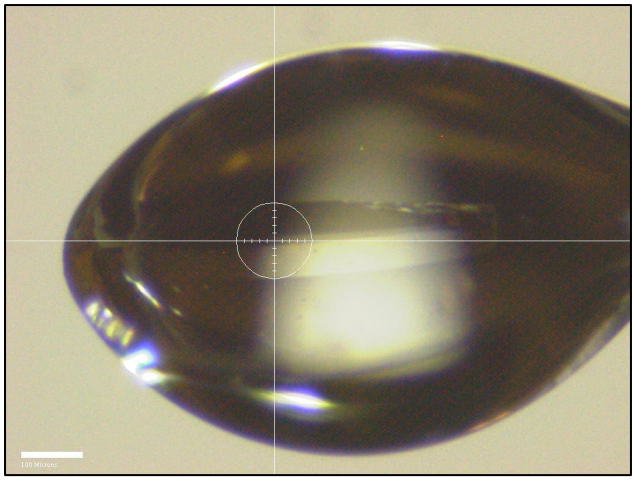


Figure 11: Photograph of the crystal loop being centered. The cross-hair indicates the center position of the X-ray beam. White bar: 100μm.

2.5.4. Structure solving

The process of structure solving from the raw dataset of diffraction pattern images obtained at various orientations of the crystal relative to the X-ray beam axis starts with the numeric process of scaling. This computation is necessary to bring the positions and intensities of the spots present on different images (spindle angles) onto a single frame of reference. The program DENZO[129] was used to that effect.

Unfortunately we were not able to optimize the crystal generation process enough to have heavy atom derivative crystals available for the synchrotron beam time allocated to us. As a result, direct phasing methods could not be applied, and the structure was thus solved by molecular replacement. In this phasing method, the phases of a different, structurally solved protein which shares high structural similarity with the protein that is studied are used as initial approximation. This initial approximation is then iteratively improved using the experimental dataset. This approach, although raising questions of model bias, is generally accepted as accurate, especially when as in our case, the structure is known to be close to that of the model used to generate the initial phase estimate. Because of its high level of sequence homology with HPC-4, the structure of the Fab MNAC13 (pdb code 1SEQ)[130] was used as search model. The translation and rotation of the search model into the measured unit cell for HPC-4 crystals was performed using the program AMoRe[131]. The variable and constant portions of the search model were separated in the input file to improve the Patterson search[132]. This procedure allowed differences in elbow angle between 1SEQ and the final search model. The iterative unit of the refinement process was a reciprocal space refinement of the structure using the CNS suite[133],

followed by a real space model building step using MAIN[121]. The refinements were carried out with simulated annealing only in the first few rounds. At each step, $2F_o$-F_c and F_o-F_c Fourier difference maps were calculated by combining phases from the refined model and the intensities from the experimental dataset. The model building was carried out by fitting the atoms of the current model into maps contoured at $1.0\times\sigma$ ($2F_o$-F_c) and guided by $2.0\times\sigma$ and $-2.5\times\sigma$ contouring of the F_o-F_c map. Periodically, a composite omit map was calculated and used in model building[134]. At each step, the statistical values R_{work} and R_{free}[135] were used to evaluate progress and the quality of fit to the experimental data. When the R values were not significantly improved by further refinement rounds, the structure was submitted to rigorous geometry checks using the programs PROCHECK[136] and WHATCHECK[137].

3. RESULTS

3.1. HPC-4 ficin hydrolysis

3.1.1. Generation of Fab'$_2$

Although a commercial kit is available from PIERCE ("ImmunoPure"), in our hands it did not yield satisfactory results when strictly following the manufacturer's indications (fig. 12). One of the recommended cysteine concentrations actually generated the Fab fragment but it was obtained as a very fuzzy band, indicating the presence of several distinct cleavage products that may have been suitable for biochemical applications, but would unlikely crystallize very easily. In addition, the cysteine required to obtain proper cleavage in the ficin column was found to interfere with the integrity of the resulting fragments, as the 50kD fuzzy band obtained from the ficin column could not be retrieved again in the unbound fractions of the protein A column. This result was attributed to the reduction of the interchain disulfides. It was thus decided to replace the first step of the kit by hydrolysis using soluble ficin, which would allow a better control over all variables in the hydrolysis step. The hydrolysis was stopped by alkylating agents such as iodoacetic acid (IAA) or N-ethyl maleimide (NEM), providing the additional advantage of completely blocking the

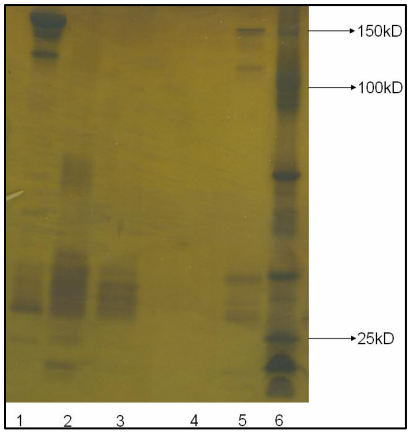


Figure 12: SDS-PAGE for PIERCE ImmunoPure kit. Lane 1: starting material; lane 2: digestion mix; lanes 3,4: unbound fractions from the protein A column; lane 5: bound fractions from the protein A column; lane 6: molecular weights.

reduction of the sample that was observed initially using strictly the protocol of the kit. For the initial optimizations, the hydrolysis was set arbitrarily at 23h and room temperature as a starting point. The effect of reducing agent concentration was evaluated in a first step. The fragmentation patterns obtained around 1mM cysteine (fig. 13) revealed that there was no clear cysteine concentration threshold between the cleavage into either the 100kD Fab'$_2$ or the 50kD Fab fragments. This finding was somewhat surprising, because in the PIERCE kit, the reducing agent concentration is supposed to entirely determine the outcome of the hydrolysis in terms of the nature of the fragment that is obtained. In particular, no reducing agent concentration was found that would both result in the disappearance of the full-length HPC-4 band and limit the amount of 50kD Fab generated. A reasonable middle point of 0.6mM cysteine was identified, which yields the Fab'$_2$ as major product while almost entirely suppressing the full-length HPC-4 band. However at this concentration there was already a detectable amount of Fab present, which would be hard to separate by conventional methods. Indeed, these fragments would be expected to be both equally unretained by the protein A chromatography and equally well retained by the antigen peptide affinity chromatography. Although it would be possible to separate them by gel filtration, this

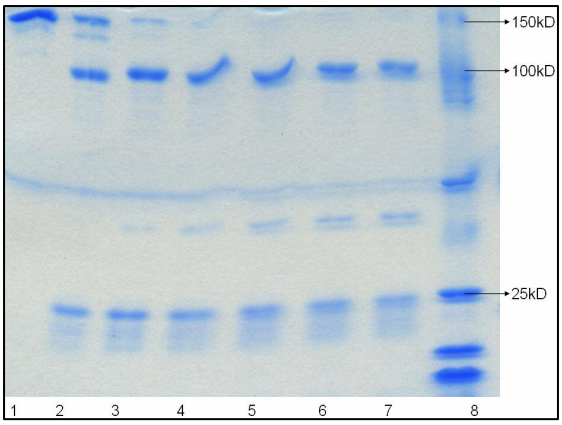


Figure 13: Effect of cysteine concentration on 23hrs ficin hydrolysis at room temperature. Lane 1: starting material; lane 2: 0mM; lane 3: 0.3mM; lane 4: 0.6mM; lane 5: 0.9mM; lane 6: 1.2mM; lane 7: 1.5mM; lane 8: molecular weights.

strategy was not chosen because it is often associated with important dilution of the products.

The effect of temperature was then investigated at the previously identified cysteine concentration (0.6mM) in an attempt to decrease the amount of Fab generated while maintaining a good yield of Fab'$_2$ produced. As seen in figure 14, decreasing the temperature dramatically limited the extent of the hydrolysis. It was found that after 4 days, the 100kD Fab'$_2$ band was noticeably enriched, but no Fab was detected.

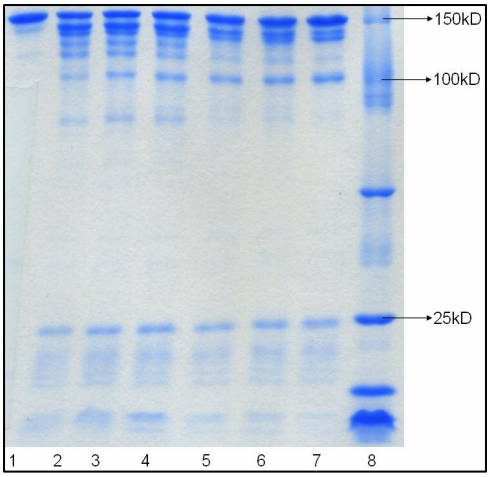


Figure 14: Ficin hydrolysis timecourse at 0°C; 0.6mM cysteine. Lane1: starting material; lanes 2-7: days 1-6; lane 8: molecular weights.

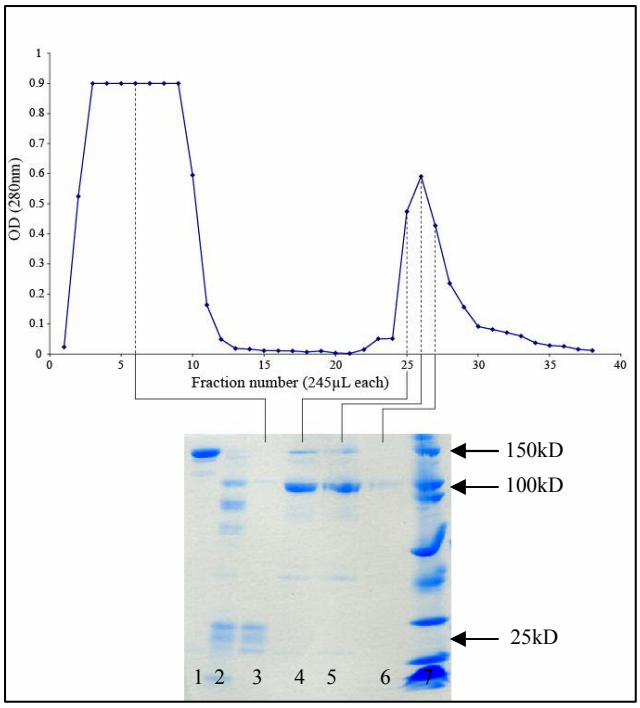


Figure 15: Affinity chromatography of the Fab'$_2$ preparation. Top pane: elution profile of the column. Elution started at fraction 21. Lower pane: SDS-PAGE. Lane 1: starting material; lane 2: hydrolysis mixture; lane 3: unbound fraction, lanes 4-6: elution fractions; lane 7: molecular weights.

Contamination by undigested HPC-4 was considered preferable than by the Fab, because it would be easier to purify using a step of protein A chromatography as recommended in the PIERCE kit. This condition was further optimized in terms of duration and cysteine concentration. A time series at 0°C in increasing concentrations of cysteine revealed that an optimum was

obtained between days 4 and 6 at 1mM cysteine. To ensure that the fragments obtained under these conditions were still functional, they were submitted to purification on the peptide antigen affinity column. As seen on figure 15, the fragments were very well retained by the column and recovered in a rather sharp elution peak, which was considered as indicative of a retained binding ability. The full-length HPC-4 that was obtained in the peptide column elution fractions was successfully removed using the protein A column of the PIERCE kit (fig. 16) and precisely applying the recommendations of the manufacturer.

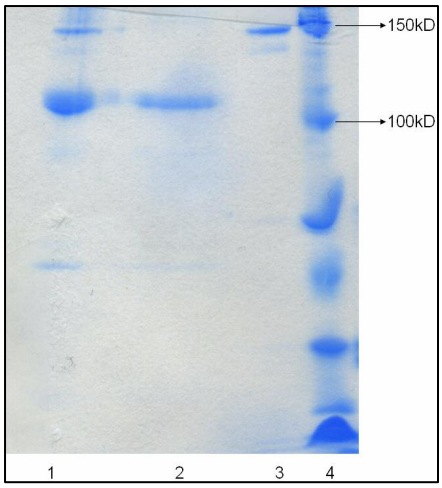


Figure 16: Protein A purification of Fab'$_2$ fragments: SDS-PAGE. Lane 1: elution fraction form the peptide column (starting material); lane 2: unbound fractions; lane 3: bound fractions; lane 4: molecular weights

3.1.2. Generation of the Fab fragment

Attempts at generating the Fab fragments were made either directly or through reductive dissociation of the Fab'$_2$ fragment obtained by the method previously developed.

Dissociation of Fab'$_2$ The Fab'$_2$ fragments were brought to different concentrations of cysteine (1-10mM) and incubated for 1h at room temperature or at 37ºC. This method did not yield any detectable levels of dissociation. Increasing concentrations of cysteine (30-90mM) allowed us to obtain some dissociation, and resulted in a heterogeneous mixture of fragments grouped around 25kD and 50kD. This size heterogeneity as seen on non-reducing lanes was attributed to the partial reduction of

the intrachain disulfides. This was expected to yield fragments that would not be functional, so this method was no longer investigated.

Direct hydrolysis: As a starting point towards finding suitable conditions for the generation of the Fab fragment by hydrolysis, the cysteine concentration was raised from the value used for the Fab'$_2$. The goal was to increase the likeliness of reducing the disulfide linkages of the hinge region. A hydrolysis of 23h at room temperature with 7mM cysteine yielded appreciable amounts of the 50kD Fab and total disappearance of the Fab'$_2$, but the preparation was heavily contaminated by a doublet of bands of approximately 25kD in size. These species were found also in the elution fractions of the epitope affinity column, possibly indicating their ability to bind specifically. Separation of these three species was attempted using a gel filtration chromatography on Sephadex resin G-75 (fractionation range for globular proteins: 3-80kD). To improve peak separation, a bed length twice longer than the manufacturer's recommendation was used.

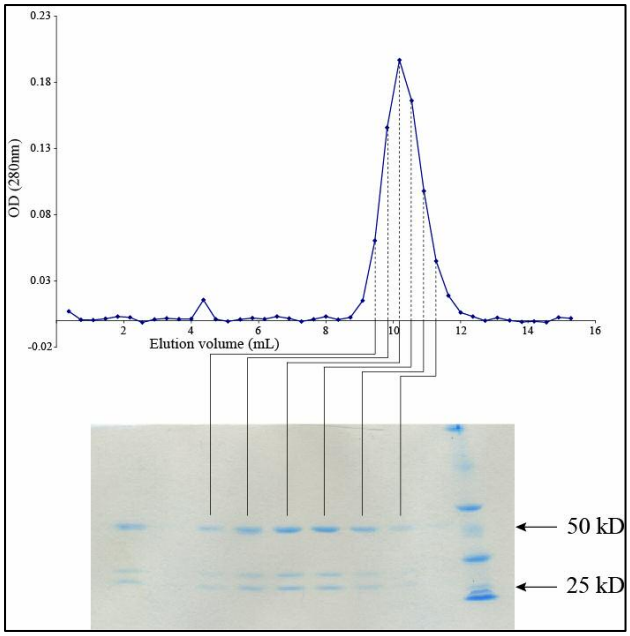


Figure 17: G-75 fractionation attempt. Top pane: elution profile. Bottom pane: corresponding SDS-PAGE. The starting material is shown on lane 1. Last lane: molecular weights.

Unfortunately, a single elution peak was detected, and the two species were equally present

throughout all absorbing fractions (fig.17). It was hypothesized that the 25kD bands were binding nonspecifically to the peptide epitope affinity column, and that they may be associated to it with a lower affinity. Thus, differential elution of the Fab and of the 25kD contaminants was attempted using variations of the wash buffer.

As seen on figure 18 (top pane), buffers containing 2M NaCl or 2M NaCl and 0.5M of the chaotrope guanidine hydrochloride were not able to selectively detach the 25kD fragments. Instead, the 25kD fragments were eluted as a single peak along with the Fab under regular elution conditions. Because the interactions may have been hydrophobic, the same experiment was repeated using a lower salt concentration (10mM) or by adding 30%(v/v) ethanol to the wash buffer. These measures were aimed at decreasing the strength of the hydrophobic interactions with the column by decreasing the polarity of the solvent. Again, this was useless to selectively elute the 25kD fragments from the column, as all the

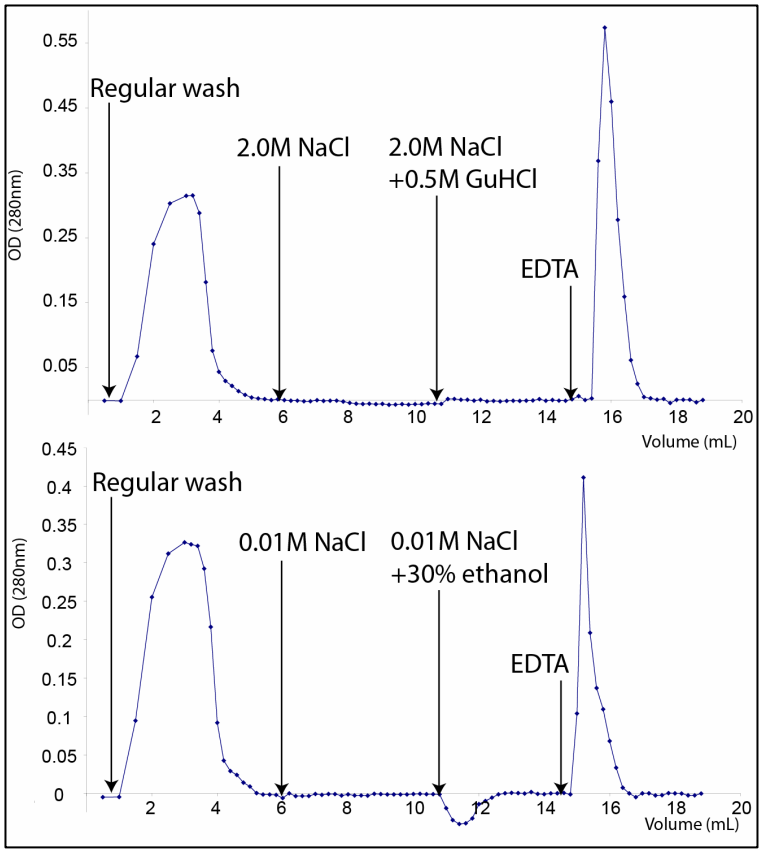


Figure 18: Attempt at separating the 25kD fragments from the 50kD fragment on the peptide epitope affinity chromatography column. Top pane: attempts at high solvent polarity. Lower pane: attempts at low solvent polarity. In either case, the elution peak contains the two fragments (not shown).

proteins came out as a single peak once calcium was replaced by EDTA (fig. 18, lower pane).

At this point, it became increasingly plausible that the interaction of the 25kD contaminants with the column was actually specific. Indeed, any nonspecific interaction would not be expected to be sensitive to mild variations in the chemical composition of the buffer such as the replacement of 4mM calcium by 4mM EDTA that triggered the release of the contaminants. Furthermore, the structure of the Fab fragment is in fact a disulfide linked dimer of two chains of molecular weights close to 25kD. It was thus hypothesized that the observed contaminants may actually be a reduced form of the Fab fragment or a form where the ficin cut occurs a little above the interchain disulfide bond. In the resulting form, the light chain and the cleaved heavy chain would not be disulfide linked, but would stay associated in solution owing to the extended surfaces of interaction that exist in antibodies. This hypothesis of a non covalent Fab, or "ncFab" was consistent with all of the experimental data. Indeed, the striking similarity in behavior regarding the interaction with the peptide epitope column tends to indicate that the two chains could form a functional binding site in a non covalent way that would be able to attach to the column.

The G-75 gel filtration results also contributed to this hypothesis, because the lack of separation may have been caused by the association of the two 25kD bands into a noncovalent assembly of 50kD in apparent molecular weight.

It was thus decided to attempt to generate a pure solution of the ncFab. For that purpose, the more potent reducing agent dithiothreitol (DTT) was used to facilitate the reduction of the hinge disulfides. A hydrolysis of 23h at room temperature was submitted to epitope affinity chromatography, and was insufficient to deplete the 50kD species. Fortunately, it was rapidly found that depletion of the Fab in favor of the ncFab could be realized merely by keeping the DTT concentration at 1mM, increasing the hydrolysis duration to 4 days and decreasing the temperature to 0°C (see fig. 19). As a confirmation of the association of the two 25kD bands into a 50kD species, the solution obtained was submitted to molecular weight correlation using a dynamic light scattering apparatus. The average of 50 measurements of a solution having count rates in the 0.13-0.15s^{-1} range (acquisition time: 10s, signal/noise detection threshold: 2.5, assumed geometry: spherical) was found to be 47.6kD, in excellent agreement with our hypothesis. Intrinsic fluorescence experiments were later performed, and confirmed the proper

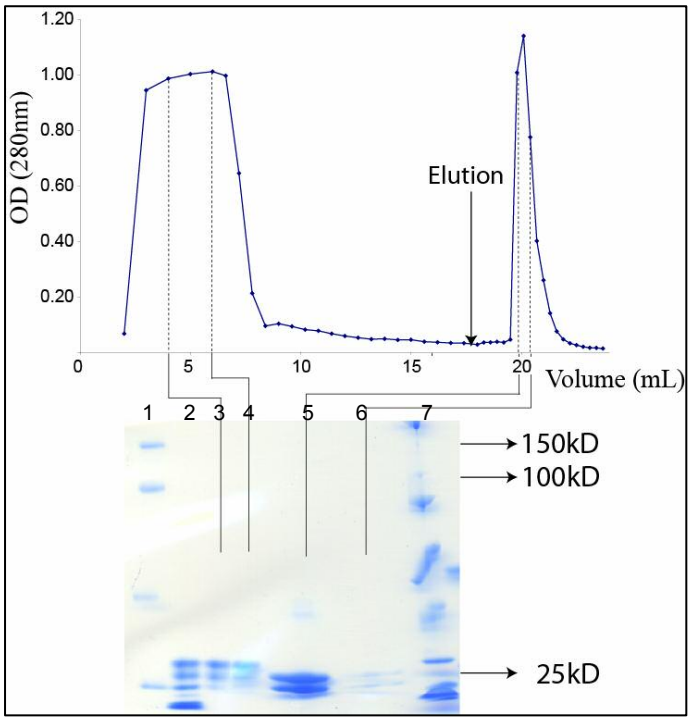


Figure 19: Affinity chromatography of the ncFab preparation. Top pane: elution profile of the column. Elution start indicated by arrow position. Lower pane: SDS-PAGE. Lane 1: starting material; lane 2: hydrolysis mixture; lanes 3-4: unbound fractions, lanes 5-6: elution fractions; lane 7: molecular weights.

function of the ncFab. Most conclusively, the 2.3Å crystal structure obtained also showed unequivocally the dimeric nature of the 25kD bands in the crystalline phase, and their non covalent association into a fully functional Fab form.

3.2. Fluorescence titrations

A typical fluorescence run is shown in figure 20. As seen on the raw data trace, the fluorescence signal increases when calcium is first introduced. This is consistent with the results for the full length HPC-4[41]. The signal appears to decrease upon further additions of calcium, but this is only due to the dilution effect resulting from the extra volume of the calcium solution, as the corrected dataset shows. Figure 21 shows the calcium titrations made at various peptide concentrations. In the absence of peptide, a fluorescence intensity decrease of at the most 2% was observed. The significance of this decrease is unclear, but it could in principle result from the occupancy of a weak calcium binding site on the ncFab in the absence of peptide. In the characterization of the full-length HPC-4[41], the fluorescence intensity was also found to decrease in the absence of peptide, but with a higher magnitude (5%). The valency difference between HPC-4 and its ncFab may be at the origin of this apparent discrepancy.

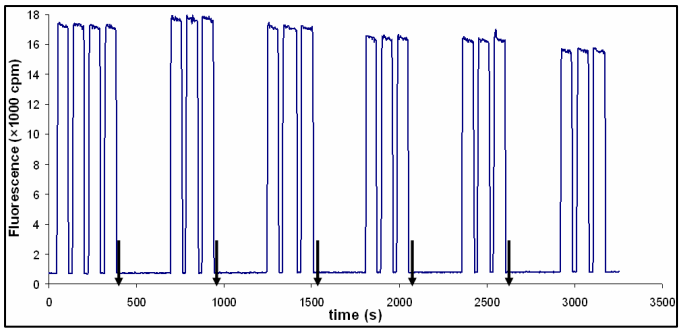


Figure 20: Typical fluorescence experiment. Shutters are closed except during measurements. Arrows: addition of calcium to 0.1, 0.2, 0.3, 0.9, 2.0mM $CaCl_2$. NcFab concentration: 1μM; peptide concentration: 1μM.

The titration of calcium dramatically increased the fluorescence signal of the ncFab in up to 2.0µM of peptide. This suggests the occurrence of a conformational change in the ncFab structure induced by calcium and peptide. This effect can not originate from a conformational change in the peptide, as it does not have any residues with significant fluorescence emission at the wavelengths used. The precise nature of this conformational change can not be determined by this experiment alone,

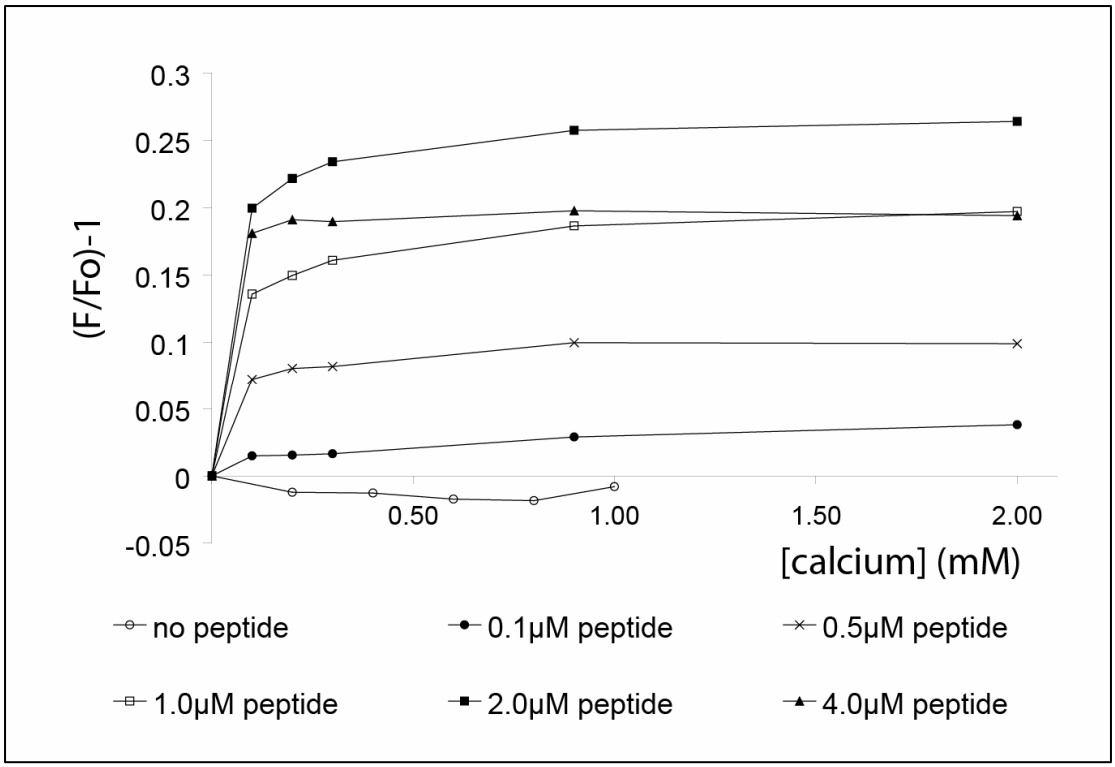


Figure 21: Calcium-induced intrinsic fluorescence intensity changes of the ncFab in presence of various peptide concentrations. The ncFab was present at 1µM. Peptide concentrations are indicated. λ_{ex}*=280nm;* λ_{em}*=340nm.*

but the crystal structure of the complex between ncFab, calcium and peptide obtained in the present project offers attractive hypotheses that will be discussed in the next section. Interestingly, this trend in maximal fluorescence intensity was only observed up to 2.0µM peptide. In the presence of 4.0µM peptide, the fluorescence increase is smaller. This could be the result of a weak binding of calcium to the peptide that

would occur only at high peptide concentrations, scavenging it from productive interaction with the ncFab. This hypothesis is unlikely, as calcium binding to the 12-mer epitope peptide used in the HPC-4 characterization was never detected. However, the 9-mer used in this study is devoid of a positively charged lysine residue present in the 12-mer, which would be expected to impair the ability of the 12-mer to interact with calcium.

The dissociation constant of the ncFab for calcium responded somewhat differently than the maximal fluorescence intensity (fig. 22). It kept decreasing at all peptide concentrations assayed, unlike the maximal fluorescence which reverted the trend at 4μM peptide. Generally, this indicates that binding of calcium to the ncFab is tighter in the presence of peptide. The highest calcium affinity, measured at 4μM peptide, was 8.6μM, in approximate agreement with the 6.5μM reported for HPC-4. These K_d values are not strictly comparable however, because they were obtained at 4 or 2μM peptide for the ncFab or the HPC-4, respectively. For reference, the K_d of ncFab for calcium in the presence of 2μM peptide was only 34μM. This difference may be explained by the highest avidity of HPC-4 over the monovalent

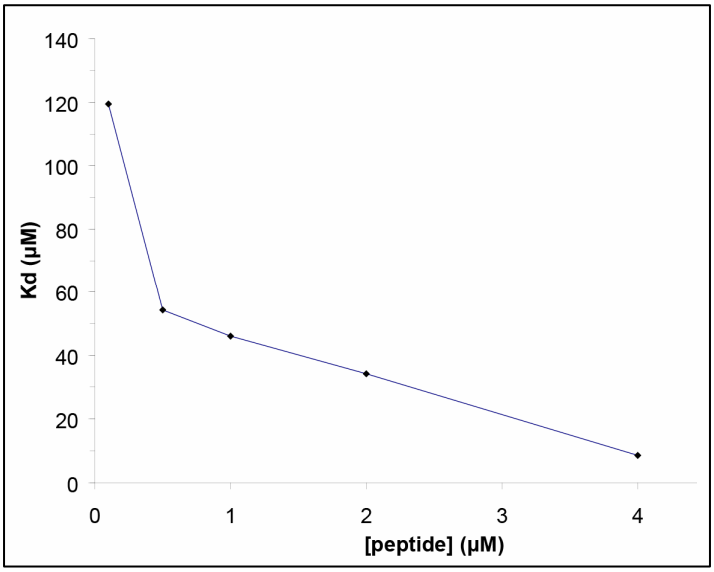


Figure 22: Dependence of the K_d for calcium of the ncFab in the presence of various concentrations of peptide. The K_d values were derived from fig.21 as described in methods.

ncFab. Still, the highest affinities reported for either HPC-4 or the ncFab are in good agreement. Overall, these results indicate that the binding of calcium to the ncFab may induce a significant conformational change in the presence of peptide.

3.3. Crystallization

3.3.1. Crystallization

Crystallization trials were done for the complexes of full-length HPC-4 and GDPC, HPC-4 Fab'$_2$ and GDPC, HPC-4 ncFab and GDPC, and finally HPC-4 ncFab and the antigen 9 residue peptide. All crystallization trials were made in the presence of calcium. Unfortunately, none of the GDPC-containing complexes yielded crystals or promising conditions, despite numerous attempts. The number of trials for each complex is represented in table 1. As a reference, a particular protein complex is generally considered difficult to crystallize after 400 unsuccessful trials[138]. In the absence of crystallization conditions, an iterative "edge search" approach was taken to tentatively identify conditions yielding supersaturation. In a first step, the conditions in which the protein precipitates or yields any type of phase separation were identified. A new series of trials were made with concentrations of the precipitating component evenly spaced from their initial concentration to zero. Observations made on this series effectively constitute a one-

Complex	Number of trials	Crystal hits
HPC-4 (full length) :Ca^{2+}:GDPC	480	0
HPC-4 Fab'$_2$:Ca^{2+}:GDPC	144	0
HPC-4 ncFab :Ca^{2+}:GDPC	984	0
HPC-4 ncFab :Ca^{2+}:peptide	432	4

Table 1: Summary of the crystallization trials

dimensional, sparse-sampled phase diagram. On this diagram, it is then possible to identify two regions, one where the precipitant concentration is high enough that the protein changes phase, bordered by a second region where the protein stays in solution. Two successive concentrations yielding a different solubility response of the protein constitute a solubility edge. Because protein crystals grow out of supersaturated solutions, it is of interest to find the highest precipitant concentration that maintains the solubility of the protein. Such a concentration is high enough to be in the supersaturated region that enables the random event of crystal nucleation, but low enough not to lose most of the protein in a useless precipitated phase. A second set of trials was setup with precipitant concentrations in small increments evenly spaced within the previously identified solubility edge. The approach can be repeated for all the constituents of the initial condition, and is also to some extent applicable to pH. Although the cryoprotectant present in some conditions certainly affects the solubility of proteins, it was not generally used as a variable in edge searches in order to maintain the cryoprotection level of the initial condition. The successive application of this method on increasingly smaller regions of the solubility diagram and in response to all the components of the initial condition allows a precise characterization of the protein behavior. However, because crystal nucleation is kinetically governed when the protein is in the supersaturated region, it remains an inherently random event. The likeliness of obtaining protein crystals can only be increased by identifying a great number of different supersaturated regions. By doing so, one increases the odds of obtaining one condition where the kinetics of nucleation is favorable.

Because crystals of the HPC-4 Fab in complex with calcium and the antigen peptide were obtained, this case will be developed here. The complex was assayed for crystallization using the commercial crystallization solution screens “Cryoscreen”, “Crystal screen II”, “Natrix”, “Membfac” and “Index” from Hampton Research. Crystals were obtained overnight in the index solutions number 4, 5 and 6 (2.0M ammonium sulfate buffered with 0.1M bis-tris pH 6.5, 0.1M hepes pH 7.5 and 0.1M tris pH 8.5, respectively). The solutions at pH 6.5 and 7.5 yielded small, spherical aggregates of needle-like crystals (fig. 24). The solution at pH 8.5 yielded a shower of microcrystals. These solutions were recreated with our own chemicals, and similar crystals were obtained in the same length of time.

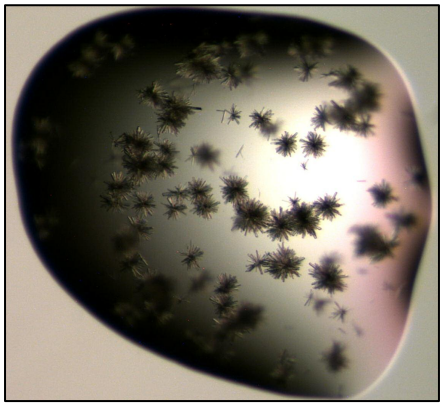


Figure 24: Spherical crystal aggregates obtained overnight in Hampton Index Screen condition 5 (2.0M ammonium sulfate; 0.1M hepes pH 7.5)

Because the presence of sulfate together with the calcium contained in the protein solution would likely enable the formation of calcium sulfate crystals, the nature of the crystals was evaluated. Fine-tuning a crystal growth condition is a very time-consuming and hands-on process, so it is important to determine the nature of a crystal beforehand to avoid spending time and reagents on optimizing the growth of a crystal that eventually turns out to be a trivial salt crystal. An ideal test would be to measure an X-ray diffraction pattern of the unknown crystal. However, this test would have been inconclusive in our case, because the small size of the crystal alone could have explained the absence of a protein-like diffraction pattern as well as would

have their actually being salt crystals. Hence, the following experiments and observations were made on these candidate protein crystals.

First, the crystallization trials were repeated using the protein buffering solution (20mM NaCl, 5mM tris pH 7.5, 5mM CaCl2, 0.02% Na azide) instead of the actual protein solution. Because this solution contains as much calcium as the protein solution but no protein, the presence of crystals with a morphology similar to that observed in the initial solutions would likely have indicated that these were in fact calcium sulfate crystals. No crystals were observed in these drops after 2 days. However, this experiment was not sufficient to prove the protein nature of the crystals. The presence of high protein concentration is expected to lower the chemical activity of the water, and hence it could potentially decrease the solubility of calcium sulfate. In other words, it was still plausible that the presence of high protein concentration is required to trigger the crystallization of calcium sulfate. Replacing the protein solution by plain calcium-containing buffer is therefore not a strict control, because both the presence of protein and the chemical activity of the water solvent are altered. The absence of crystals was a promising indication, but could still have been attributed to the differences in chemical activity between the initial condition and the test condition.

A second test was performed using Hampton's “Izit” dye solution. This dye takes advantage of the fact that salt crystals are packed much more densely than protein crystals, and are devoid of large solvent channels within the crystalline lattice unlike typical protein or biomolecular crystals. The molecular weight of this blue dye is engineered to be able to diffuse into pores as small as the typical diameter of the

solvent channels found in protein crystals. Hence, provided that the dye solution does not interfere with the solubility of the crystal, any color uptake by the crystal indicates diffusion of the dye into the solvent channels, strongly indicating that the observed crystal is a protein crystal. Two dilutions of the dye (undiluted or diluted 10 times in water) were added to a drop that contained the undetermined crystals. The addition of dye immediately dissolved the crystals. The drops were then allowed to equilibrate with the reservoir solution. After 12 hours, crystals grew back from solution with a vivid blue color. As a control, calcium sulfate crystals were grown and the drops colored with the same dye. These crystals had a different morphology (longer and thinner needles), did not dissolve upon addition of the dye, and did not take up the blue color of the dye. Although this result strongly indicated that the crystals were not salt crystals, they did not demonstrate that the crystals contained the protein complex. It is known that small peptide crystals can form that have relatively high solvent content compared to salt crystals, so the remote possibility that the crystals were actually crystals of the antigenic peptide could not be rigorously dismissed. Furthermore, the izit dye had previously been demonstrated to permeate into crystals of peptide as small as 21 residues long, but not in crystals of a 6-mer or smaller peptides[139]. In the case where the peptide would have dissociated from the antibody (for instance if the excess sulfate present had decreased the concentration of free calcium in solution to such an extent that the antibody would lose its affinity), it could in principle have crystallized as a crystal with solvent content high enough for the izit dye to be able to permeate. Still, the observation that the izit dye was able to permeate inside the crystalline clusters was a strong indication that they were in fact

the expected crystals of the HPC-4 ncFab, presumably in complex with calcium and the antigenic peptide.

Because the packing interactions of a protein crystal are very different from those of a salt crystal, the mechanical properties of these two types of crystals are very different. Salt crystals tend to be very resistant and brittle, and break into well-defined shards when stressed above failure. Protein crystals tend to behave mechanically more like a hydrogel, being degraded throughout the stress range applied. A simple mechanical stimulation of an undetermined crystal with a needle under a microscope can often provide invaluable information about its nature. The crystals obtained were submitted to this test. They disintegrated upon contact with the needle when an amount of pressure considerably smaller than that required to break calcium sulfate crystals was applied. The crystal fragments that were obtained did not have the same aspect as the originating crystal, and resembled a typical protein precipitate instead.

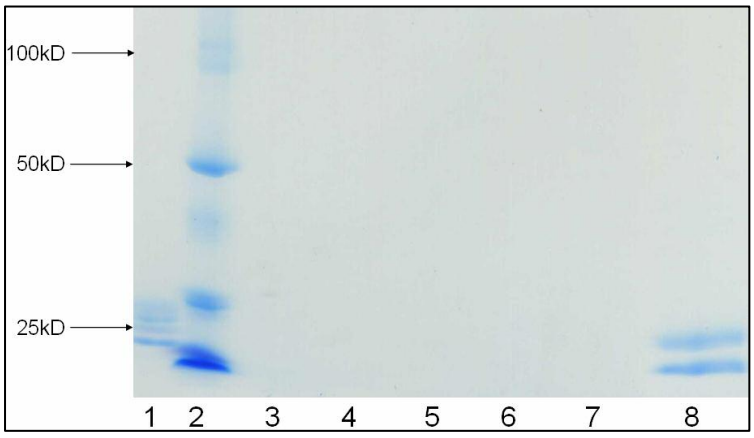


Figure 25: SDS-PAGE of dissolved crystals. Lane 1: protein solution used to grow the crystals; lane 2: molecular weights; lane 3: crystallization buffer; lanes 4-7: solutions used to rinse the crystals; lane 8: dissolved crystals.

Finally, the content of the crystals were assessed by gel electrophoresis. 30 crystalline clusters were harvested from the originating drop, and were rinsed multiple times in reservoir solution

to avoid contamination by the initial protein solution that contained them. The crystals were then dissolved in SDS-PAGE migrating dye, boiled, and submitted to gradient gel electrophoresis. As controls, the wash solutions and the protein solution used to obtain the crystals were also put on the gel. As can be seen in figure 25, only the lane containing the dissolved crystals had protein bands, and the protein bands matched exactly those obtained with the protein complex solution. This last test unequivocally established the protein nature of the crystals.

Although promising, these crystals were still too small to diffract X-rays sufficiently to evaluate their potential usefulness in structure determination. Indeed, in some instances, the packing of a protein crystal can be too poor to provide reliable diffraction data even if the crystal has an apparently good morphology. The interacting effects of ammonium sulfate concentration, buffering molecule and pH were then evaluated. Crystals with better morphology were obtained using bis-tris at pH 6.5 as buffering agent, and by decreasing the ammonium sulfate concentration to the 1.70-1.85M range. Because these optimized crystals were still too small, it was decided to further improve them by macroseeding.

3.3.2. Optimization of macroseeding procedure

The solubility pattern of the HPC-4-peptide complex was established in conditions close to the nucleating condition. In the cases where clear drops were obtained, it was determined if they were undersaturated or supersaturated by transferring a small crystal of defined size. The dissolution or decreased size of the crystal after an incubation period was taken as an indication that the drop was

undersaturated. If the crystal grew in size, or yielded the formation of additional crystals in the drop, the condition was considered supersaturated.

An important and recurrent problem was the appearance of new small crystals in the target drop after seeding. The crystals would appear either independently from the seed ("secondary nucleation"), or would directly grow off of the surface of the seed, but in incorrect orientations ("sprouts"). These small crystals were considered deleterious because they diverted some of the protein molecules away from the seed, thus decreasing its rate of growth. In addition in the case of sprouts, the quality of the seed itself was decreased so much that their diffraction pattern would worsen.

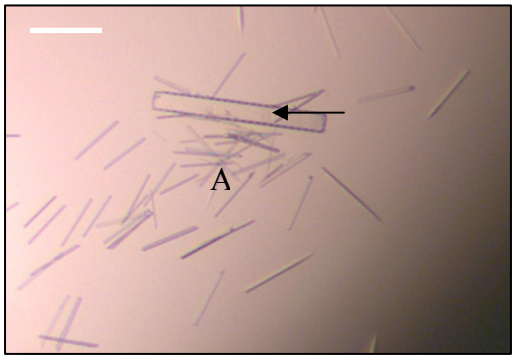


Figure 26: Example of secondary nucleation. Small crystals (A) divert protein away from the main seed (arrow). White bar:100μm

Secondary nucleation (see fig. 26) was attributed to the accidental transfer of microscopic crystal nuclei along with the principal seed during macroseeding. It could be almost entirely prevented by implementing a washing step into an ammonium sulfate concentration of 1.5-1.7M. Because this range of concentration does not allow the nucleation of crystals or even support preexisting crystal growth, it was hypothesized that the crystals put in presence of this solution would dissolve. Because of its size, the seed was only dissolved to a small extent (relative to its total size), but small crystals and nuclei were entirely dissolved, thereby decreasing the amount of secondary nucleation induced by their introduction into the supersaturated target drop. It was important not

to leave the crystals for extended periods of time in the wash solution, because they tended to shatter in large fragments after a few minutes of wash time.

Appearance of sprouts (see fig. 27) was attributed to the presence of some sort of defects on the surface of the crystals, which would act as nucleation centers inducing the addition of protein molecules onto a lattice of a different orientation relative to the seed. First, the implementation of the washing step, although efficient to suppress secondary nucleation, seemed to actually increase the number of sprouts. The addition of 10mM $CaCl_2$ to the washing solution resulted in a more efficient washing that yielded dramatically less heterogeneous surface nucleation. This observation was also considered as indicative that the HPC-4 present in the crystal was still saturated in calcium despite the high ammonium sulfate concentration. Interestingly, the presence of calcium allowed much longer wash times than previously, when they were limited by crystal shattering from its core. It would indeed be expected that the transfer of a calcium-containing crystal to a drop lacking the ion would induce the dissociation of calcium from the protein of the crystal by mass action. The resulting heterogeneity of protein in the crystal could likely cause the shattering of crystals occurring in calcium-free wash solution that was observed. The partial dissociation of calcium from the protein copies located on the outermost shell of the crystals would also likely yield patches of calcium-free copies of the protein from which the heterogeneous crystals could grow. In any case, the

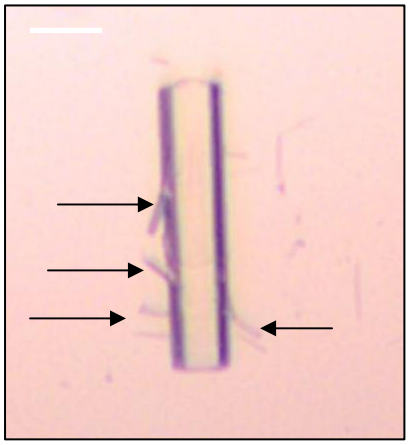


Figure 27: Sprouting. Arrows: sprouts growing off of the main seed. White bar: 100µm

differential responses obtained in the presence or absence of calcium were an indication that the crystals were actually saturated with calcium, and hence probably also with the peptide antigen.

A second source of sprouting was the mechanical manipulations of crystals. The crystals were relatively fragile, and shattered very easily upon application of a very mild mechanical stress. Although the nylon loops that were used for macroseeding were flexible, any contact of the loop with the crystal might still be expected to create surface defects. This source of sprouting was unavoidable even when using loops much bigger than the size of the crystal because it was often necessary to gently pull the crystal with the loop to have it escape the surface tension of the originating drop. Sprouts were also sometimes mechanically dissected from the seed, with some success if the dissection was followed by an extended wash. However, dissection of sprouts generated additional surface defects in an incontrollable manner that could not be removed even after extended wash. Once this was noticed, the presence of sprouts was used as a signal to consider stopping the iterative macroseeding of a particular crystal.

Despite these problems, several crystals were amenable to dramatic increases in size, as shown in figure 28. The rate of

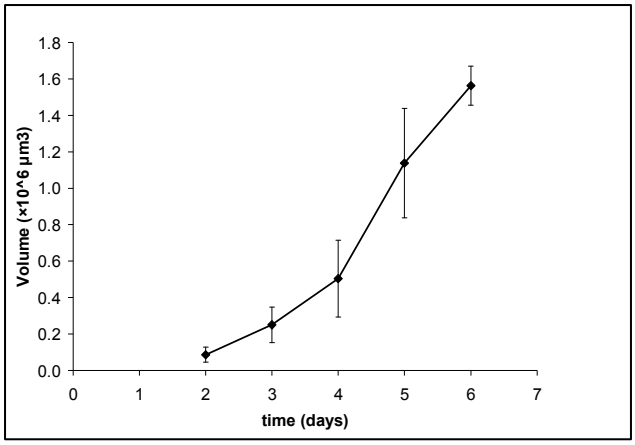


Figure 28: Time course of crystal volumes during macroseeding-assisted growth. Crystals were too small at day 1 to be accurately measured. The points represent an average calculated from 6 crystals. Bars: standard error on the average value.

macroseeding growth, as defined and derived in appendix X.1., was of 6.1μm·day^{-1}, or 6.8 μm·day^{-1} when considering a decay of 70 days (also defined in appendix X.1.). Final sizes of 40-50×40-50×700-800μm^3 were usually obtained over 5 macroseeding iterations. Increasing the number of macroseeding iterations invariably increased sprouting problems, thus limiting the final size for the crystals. The decision to stop macroseeding cycles was quite arbitrary because the appearance of sprouts was somewhat unpredictable. Crystals with cross sections as small as 40×40μm^2 routinely yielded diffraction patterns nearly good enough for data collection. Thus, this particular threshold was considered as a minimum. Crystals yielding sprouts before reaching this threshold were mechanically dissected for one or two macroseeding rounds and discarded if sprouts persisted. Macroseeding of crystals bigger than this threshold and devoid of sprouts was considered very carefully, and performed with all the precautions possible.

Many crystals had a thickest part at the center of their long axis, and had a cross sectional area that decreased mildly and smoothly towards the extremities. These crystals usually had excellent diffraction patterns and were less prone to sprouting. If left equilibrated for days after the last macroseeding step, they would slowly fill until a fully prismatic geometry was adopted.

3.3.3. Cryoprotection

Cryoprotection trials were made by taking a reference room temperature diffraction pattern of a crystal, transferring it from their growth solution to a tested cryoprotecting solution, taking a new diffraction shot at 100K and comparing the

diffraction properties before and after the cryoprotecting treatment. This procedure eliminated the crystal to crystal variability from the list of experimental unknowns.

Initial attempts were made using the conventional cryoprotectant glycerol, brought at a concentration of 30% in the macroseeding buffer. Although good amorphous glass formation was obtained merely by quickly dipping the crystal in the cryoprotectant, its diffraction pattern was of dramatically worse quality than before treatment. This was attributed to the partial disruption of the crystalline lattice by the important osmolarity gradients bound to occur during the brief equilibration period. Hence, a gradual replacement of the macroseeding solution by the cryoprotecting solution, based on a published method[125], was adapted. A drop of the macroseeding solution was deposited onto a coverslip, and the crystal was transferred to this drop. Then, while monitoring the crystal under the microscope, a small volume of the solution was removed and replaced by the same volume of the cryoprotecting solution, effectively increasing the cryoprotectant concentration surrounding the crystal. This operation was repeated several times until the initial macroseeding growth solution was significantly diluted into the cryoprotecting solution. This technique has the advantage of limiting the mechanical manipulation of the crystal, and also to allow a precise control of the cryoprotectant concentration profile throughout the procedure. A simple series analysis of the drop composition with respect to the cryoprotecting agent can be used to demonstrate that the concentration time profile can be precisely operated with the following variables: the wait period in between solution replacements; v, the volume of solution that is replaced at each step; V the total volume of the drop; C_f the concentration of cryoprotectant that is used in

the exchanges and n, the number of exchanges performed. The simple derivation in appendix X.2. shows that the concentration at the n-th step is obtained by the following expression:

(X.2.6.) $$C_n = \frac{v \cdot C_f}{V} \cdot \sum_{i=0}^{n-1} \left(\frac{V - v}{V} \right)^i$$

In practice, it was found that evaporation of the drop was negligible only when the drop volume V was sufficiently large. A volume of 10μL was estimated sufficient for a total exchange time of 10 minutes or less. The concentration C_f was determined empirically for each cryoprotectant by immersing an empty loop into various concentrations, and fixing the loop onto the X-ray diffractometer with the cryostated nitrogen gas stream system on. The quality of the cryoprotection was estimated visually by the clear or hazy aspect of the drop or by the formation of ice crystals. As equation (X.2.6.) predicts, at fixed V and C_f, the highest concentration step increase, occurring in the first step, is dictated by the volume v. This volume was set at 2μL, so that the steepest gradient would only be to go from 0 to 20% of the final cryoprotectant concentration. Plots of (X.2.6.) were also used to determine that under these conditions, 22 steps are required to obtain an effective concentration in the drop of 99% of C_f. Unfortunately, adopting this step-increase protocol did not improve the cryoprotection, maybe because the 20%×C_f (amounting to 6% in glycerol concentration) initial concentration increase was still too high for the crystalline lattice to withstand. A smaller value for C_f of 20% was also tried using the step increase method, because this C_f was still cryoprotecting and because the highest concentration increase would only amount to 4% in glycerol concentration, which may have been less damaging to the lattice. This did not prove to be helpful either.

Observations made during the optimization of the crystal growth conditions indicated that the presence of even small concentrations of glycerol impaired the nucleation, growth rate and the morphology of the crystals. This observation was an indication that glycerol may improve the solubility of the protein, or in other words be disruptive for a crystalline lattice. If this hypothesis were correct, it would mean that regardless of its final concentration, glycerol would never be a successful cryoprotectant for this crystal. Thus, other shapes of gradients that could have been obtained by varying v and C_f at each step were not investigated because crystals were in short supply.

Instead, the effects of different cryoprotectants were assessed because the differential sensitivity of crystals to cryoprotecting agents is widely accepted[124]. The organic solvent paratone was assayed. This petroleum derivative is transparent to X-rays, and does freeze as a glass. Because paratone is not miscible with water, it is in principle possible to maintain the crystal in its growth solution within the drop of cryoprotectant. However because of its high viscosity, it proved difficult to transfer crystals without breaking. Another difficulty was to thoroughly remove the thin film of water that surrounds the crystal within the oily phase. This was required to prevent ice formation outside the crystal. Finally, paratone did not prove effective in cryoprotecting the crystals, presumably because the water remaining in the growth solution present in the solvent channels of the crystal had expanded while cooling to an extent that was damaging to the crystalline unit.

For reasons discussed in the methods section, the small organic acid malonate was used. It was determined that concentrations in excess of 3.2M were required to

obtain the cryoprotecting effect. As a safety precaution, a concentration of 3.4M was adopted. A quick dip in 3.4M malonate did not suitably cryoprotect the crystal. However at the macroseeding pH of 7.5, the majority of the malonate molecules have both their carboxylic acid functions dissociated. The resulting negative charges make it a potential calcium chelator. With such a high concentration, even if the acid functions are neutralized by sodium and even if the calcium has a low affinity for malonate, the mass action may be sufficient to dissociate the calcium present in the crystal. This reasoning was invoked to explain the poor diffraction of the crystal cryoprotected by a quick dip in malonate. To test this hypothesis, a similar procedure was attempted but using a malonate solution supplemented in a calcium concentration matching that of the macroseeding condition.

As seen on figure 29, this simple measure dramatically improved the diffraction properties of a frozen crystal, and quick dip in 3.4M malonate and 10mM calcium was selected as the definitive

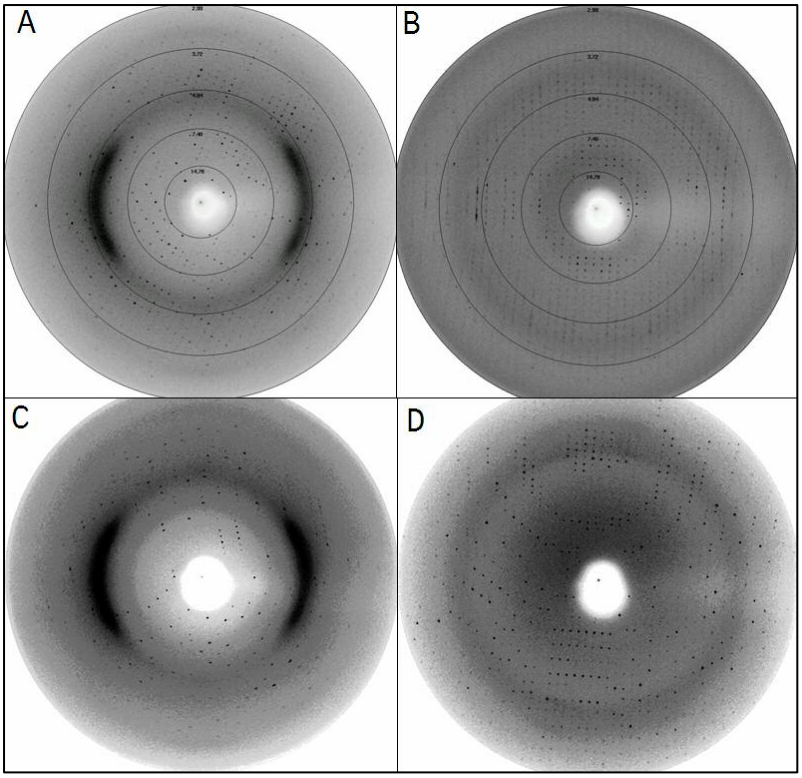


Figure 29: Cryoprotection. A,B: diffraction of a crystal before and after rapid dip in 3.4M malonate; C,D: diffraction of a crystal before and after rapid dip in 3.4M malonate + 10mM $CaCl_2$. Diffractions A and C were obtained at room temperature; diffraction B and D were obtained at 100K.

cryoprotection procedure for the crystals. The differential behavior of the crystals in response to the presence or absence of calcium was considered a promising indication that they were in fact constituted by the calcium-bound form of the antibody. Despite the fact that the crystals were grown in the presence of calcium, the presence of the ion inside the crystalline lattice was not a certainty because the high concentration of sulfate present in the crystallization condition could still have driven the calcium away from the complex forming the crystal.

3.4. X-ray crystal structure

3.4.1. Diffraction data collection

A total of three diffraction datasets were collected on a single crystal at a wavelength of 0.9128Å. The "STRATEGY" program was used to determine the most advantageous range of spindle angles to expose in order to obtain the best completeness. The position of the crystal inside the loop, and the resulting orientation of the crystal relative to the beam unfortunately yielded consequent blind regions of diffraction. Nevertheless, a total predicted 99 % completeness was expected. The first 173 frames were collected on the central part of the crystal, with 40s exposures over spindle oscillation angles of 0.5°. Heavy radiation damage was visible at the end of this collection before the full intended angle sweep was reached, as a bright area on the live image of the crystal and a mild decrease in number of high resolution diffraction spots. The spindle was then moved to the right extremity of the crystal (in the orientation shown in fig.11), where a second dataset of 421 images was taken. For reasons attributable to the quality of the crystal in this particular area, the resolution of this dataset was not very high. It was not included in the set of intensities used in

the structure solving process. The spindle oscillation angle was maintained, but the exposure time was decreased to 20 s to slow down radiation damage to the crystal. Finally, a third dataset was collected on the left side of the crystal with 387×20s exposures of 0.5°. A total of 2,029,045 spots were collected, scaled, and merged to 18,413 unique reflections, equivalent to 94.6% of the theoretical total number of spots obtainable (94.1 % in the highest resolution shell).

3.4.2. Structure solving

Images were indexed, scaled and the dataset was merged using DENZO and SCALEPACK. The molecular replacement program AMoRe was able to find a proper orientation for the search model. We obtained a crystal with one copy in the asymmetric unit in a $P2_12_12_1$ unit cell, of dimensions $68\times76\times87 Å^3$.

Refinement of the oriented search model into the final structure was performed using the CNS[133] crystallographic suite. The first step of the structure solution was to replace the amino acids of the search model (pdb code 1SEQ[130]) by their counterpart in HPC-4. The program alignP[120] was used to make a sequence alignment which determined the equivalence between the search model and the working HPC-4 model residues. In this task, the areas with the most problem were the insertions and deletions. Indeed, sometimes it was evident that the sequence alignment had placed these in disagreement with the electron density. In this case, the density was considered more indicative than the alignment because it contained more experimental information even at this early stage. When the density was not clear enough to provide a direct answer, some errors in the sequence offset remained for

several refinement rounds, until the electron density map became clear enough to point out the problem. This occurred for instance in the CDR-H2, where the sequence had to be shifted one residue downstream after it became evident that this would facilitate the modeling of a proline residue 6 positions away. This correction could only be made after 6 rounds of refinements had improved the quality of the map enough to perform this modeling task with confidence. Changing the residues of the working model for their counterpart in HPC-4 was very useful for the quality of the model, and expectedly improved the R factors more than any subsequent rounds of building and refinement.

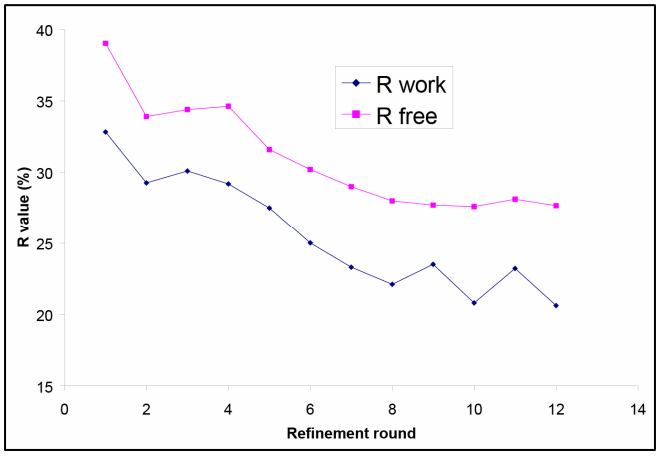


Figure 30: Evolution of the crystallographic R values during refinement. Refinement rounds 8 and 10 were performed according to geometric constraints.

An overview of the progress of the R factors over various refinement stages is represented on figure 30. An almost steady improvement of the R factors is visible in the first 8 rounds. The short-lived R values increases seen in rounds 2, 3 and 4 (R_{free}) or 2 and 3 (R_{work}) was due erroneous input files given to the CNS program. These turned out to be inconsequential. After 8 rounds of refinements, most of the changes that could be done to the model were suggested by the geometry check programs WHATCHECK and PROCHECK. It was noticed that in some regions, the model had to be distorted away from the electron density to satisfy the canonical geometry. As a result, although beneficial in terms of geometric correctness of the

model, these modifications changed it to a state that did not explain the observed intensities as well as did the model without geometric corrections, and R values increased. Over the last refinement rounds, the improvements in R values obtained by refinement were canceled out by the geometric constraint corrections, and it was decided to arrest the refinement process. Final R values of R_{work}=20.62 and R_{free}=27.66 were obtained after 10 rounds of reciprocal space refinement. A summary of refinement statistics is shown in table 2.

Unit cell geometry:	$P2_12_12_1$; a=67.9Å,b=76.3Å,c=87.4Å; α=β=γ=90°
One copy per asymetric unit	
Resolution range:	50-2.3 Å
Average signal to noise ratio:	I/σ=14.7
Reflections:	16,634 used for refinement 842 used as test set
Completeness:	94.6% overall; 94.1% in the highest resolution shell
Refinement statistics:	R_{work} =0.21 R_{free} =0.28 R_{merge} =0.11
Bond geometry:	$RMSD_{length}$=0.008Å $RMSD_{angle}$ =1.49°

Table 2: Summary of refinement statistics.

The density for the peptide and for the calcium became visible very early in the refinement process, but these were not modeled until round 5. Although placing the peptide could in principle have been possible, the proper orientation would have been difficult to find in the early maps. Furthermore, this decision has prevented any model bias from affecting the density for the peptide. When it became evident that additional atoms needed to be positioned in the density to maintain the steady

improvement of the R factors, water molecules were placed in the peptide density. This way, the sharpness of the peptide density still improved and no bias was taken as to the orientation of the peptide. The positioning of the peptide was in large part realized by placing the sidechain of arginine 12 in a relatively large protrusion on the peptide density. With this constraint, the two possible polarities of the peptide were attempted, and only the one with the carboxy terminus nearing the CDR-H3 provided a good match to the rest of the density. The position of the peptide did not explain a large portion of density that was visible both on the $2F_o$-F_c map and on the positively contoured F_o-F_c map, surrounded by negative and partial negative charges. Because the relative positions of the electronegative atoms approximately satisfied the classic pentagonal bipyramidal geometry, a calcium ion was modeled in this region. By modeling in the peptide and calcium, the quality of the map in this region and elsewhere improved and R values decreased significantly.

In the final model, only 6 residues of the total 536 have their backbone angles in generously allowed regions of the Ramachandran plot, and no residues are found in the disallowed regions except glycines. Geometric checks carried out on the structure reveal that the proline puckering amplitudes and phases are within normal ranges.

3.4.3. The structure of HPC-4: general observations

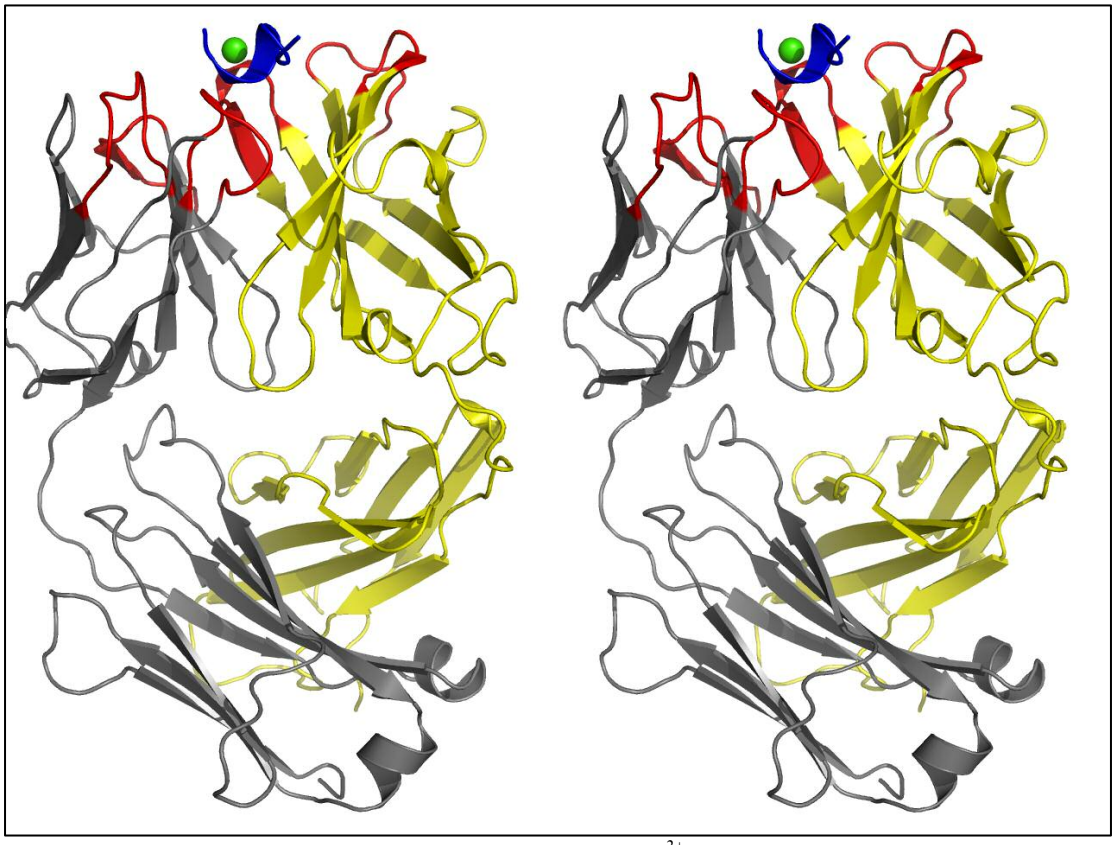


Figure31: Overall stereo view of the HPC-4 ncFab-Ca^{2+}-epitope peptide complex. β-sheeted regions are represented as arrows. Grey: light chain; yellow: heavy chain; red: CDRs; blue: epitope peptide; green sphere: calcium ion.

The structure of HPC-4 (fig. 31) is a typical example of the immunoglobulin fold. It is constituted of 4 Ig domains, 2 each in the heavy and light chain. The antigen binding site is occupied by calcium and by the peptide, as will be described in more details in the next paragraph. The ficin cleavage has occurred precisely one residue above the heavy-light disulfide bond, and the new carboxy terminal residues of the heavy and light chain remain in close proximity. The electron density of the light chain ends unambiguously and cleanly after the two C-terminal oxygens of E212, suggesting strongly that this is the ficin cleavage site. In the heavy chain, the electron density is continuous for the main chain atoms through R213 although weak density

beyond this position leaves ambiguity about the identity of the residue at the C-terminus. The variable domains of the heavy and light chains each have 9 β strands, while the constant domains only have 7. Strands 3c in both the heavy and the light chains are very short. The connectivity of the strands satisfies the canonical topology in both the variable and constant domains. Each domain possesses a disulfide bond linking strands 2 and 6.

One of the first features made visible by the 2.3Å resolution level of the structure is an extensive network of hydrophobic interactions that maintains the heavy and light chains in close proximity despite the absence of a covalent bond. In the variable domains, residues P_{44}, L_{46}, F_{98}, and W_{91} of the light chain and residues I_{37}, L_{45}, W_{47}, W_{52}, V_{93}, M_{95}, M_{100} and W_{103} of the heavy chain constitute a hydrophobic core that maintains the two domains in close interaction. In the constant portion of the ncFab, the hydrophobic core contains residues F_{117}, V_{132}, F_{134}, and L_{159} in the light chain and L_{124}, L_{141}, F_{166}, P_{167}, and V_{169} in the heavy chain. The variable and constant domains are separated by rather long loops devoid of secondary structure sometimes referred to as “elbows”. In the light chain, the elbow runs from K_{106} to T_{113}, while in the heavy chain it runs from V_{111} to S_{120}. A number of helical portions are found in the structure, most noticeably on the C-terminal side of the constant domain of the light chain, where the loops connecting strands 1 and 2 and 5 and 6 have more than one complete helical turn. Other regions appear helical but their φ and ψ angles do not strictly satisfy the definition for helicity.

The G_{127} to Q_{131} loop in the C-terminal part of the constant domain of the heavy chain has the highest B factor value, and was difficult to model. Atoms of this loop were weighed at zero during the refinement, but a likely conformation was built nonetheless in the model. Noticeably, this stretch was also absent in the phase search model, possibly because it is highly mobile. Another interesting feature found in HPC-4 is the presence in the first strands of the variable domains of a noticeable kink at the level of proline 9 (heavy chain) or proline 8 (light chain). This feature is visible also in the phase search model or in Q-425.

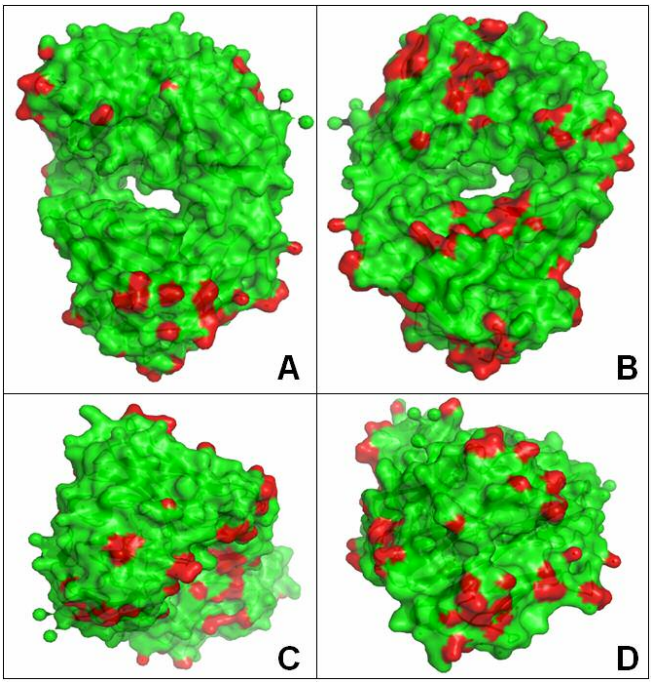


Figure 32: Crystal contacts. The structure is represented as a van der Waals surface. Red patches indicate atoms less than 4Å away from atoms of a neighboring molecule. A: orientation with long axis in the plane of the sheet, antigen binding site on top and heavy chain on the right; B: 180° rotation of A along the vertical axis; C: 90° rotation of A along the horizontal axis; D: -90° rotation of A along the vertical axis. In C, the orientation is such that one looks down into the antigen binding site.

The packing interactions of the protein molecules in the crystal are relatively minor, which suggests that the conformation seen here is relatively close to the energy minimum in solution. As seen on figure 32, most of the crystal contacts occur on the sides of the molecule. Most interestingly, the antigen and antigen binding regions appear almost entirely devoid of any contacts. In the variable portion of the

molecule, only two places are found where backbone atoms are located less than 4Å away from a neighboring molecule. The first position is a small patch formed by G_{44} in the heavy chain and G_{100} in the light chain. It is in contact with the carboxy terminal residues of the light chain of a neighbor. The second place is the $S_{65}G_{66}$ dipeptide on the back sheet of the light chain, which contacts the loops connecting the strands 6 to 7 of the constant portion of a neighboring light chain. Importantly, both these regions do not diverge noticeably from their equivalent position in the search model, even though the packing interactions are very different. This suggests that they retained their initial conformation despite being in close proximity to a crystal neighbor.

An unusual feature of our structure is the presence of no less than 6 *cis*-prolines among a total of 26 prolines present in the structure. Although uncommon, the presence of *cis*-prolines is not an anomaly, as this pseudo-imino acid is the only residue to which this conformation is not almost energetically entirely forbidden[(140)]. The 23% representation of *cis*-prolines seen in our structure is still higher than the average 5%[(141)], but it does not seem unusual considering the 6 and 7 *cis*-proline residues present in the search model or the calcium-dependent antibody Q-425, respectively. All *cis*-proline residues found in HPC-4 have an equivalent *cis*-proline in the search model.

3.4.4. The antigen and the antigen binding region

One of the earliest recognizable features of the map was the presence of a large segment of electron density between the N-terminal portions of the variable domains of the heavy and light chains. This density was acknowledged as peptide, as

it was in first approximation large enough to contain at least 7 residues. The peptide seen in the final structure represents the least model-biased part of the molecule, as it was modeled in the latest stages of refinement. Furthermore, no choices were made about its orientation until the electron density map acquired a level of detail that allowed unambiguous placement of 8 of the 9 residues of the peptide.

The detailed description of the antigen peptide that follows mentions two numbering schemes. One of them is based on the starting residue of the protein C heavy chain. Thus, the first residue of the peptide is actually residue number 8, as it is the eighth residue of the heavy chain of protein C. For reference, the scissile bond, cleaved by the T-TM complex in the protein C activation reaction, is the peptide bond between residues 12 and 13. Peptide residues will be mentioned as the one letter code followed by their position on the protein C heavy chain sequence in subscript. A different nomenclature relevant to the peptide would be the Schechter and Berger numbering scheme[86], applicable when protein C is considered as T-TM substrate. This different convention, designated for protease substrates, takes the scissile bond as a reference. Residues preceding the scissile bond are referred to as Pn,···, P3, P2 and P1 ("non primed reigon"), while residues following the scissile bond are referred to as P1', P2', ···, and Pn' ("primed region"). In this nomenclature, the scissile bond is located between residues P1 and P1'. The two numbering schemes are shown in figure 2.

The antigen binding groove of the antibody (fig. 33) is mostly formed by the CDR-H2, CDR-H3, CDR-L1 and CDR-L3. Although they are not in direct contact with the peptide antigen, CDR-H1 and CDR-L2 probably play an important role in

maintaining the peptide-contacting CDRs in a shape suitable for high affinity binding of the antigen. With the exception of CDR-L3, all CDRs are relatively short, and do not protrude significantly towards the groove. As a result, the antigen binding site of HPC-4 is relatively deep, and is rimmed by the CDRs. Most of the bottom base of the binding site is formed by the elements of secondary structure of the heavy chain. The rather deep shape of the groove sets HPC-4 apart from other protein-binding antibodies, which are generally characterized by shallow antigen binding sites[142].

The N-terminal aspartic acid of the peptide is not seen in the structure. This is most likely because even though the residue is actually present, it is too mobile to diffract significantly. The electron density was however sufficiently clear to model the following glutamine with confidence. Overall, the peptide adopts a very distinct conformation where the residues up to P_{11} do not have a clearly defined secondary structure but where residues R_{12} to I_{14} adopt a helical conformation on 5/6 of a turn, followed by the carboxy-terminal aspartic acid, which does not belong to the helix.

The structure shows an extensive array of interactions taking place between the antigen, calcium and the antibody. As mentioned above, the first residue of the peptide (D_7) is invisible in the current structure, probably because of its high mobility. The fact that Q_8 is visible is somewhat surprising because no obvious contacts with the antibody are found at first. However, by displaying the crystal neighbors it becomes evident that the sidechain of this residue is almost within bonding distance of the S_{56} residue of a neighboring light chain molecule (3.92Å). This partial interaction likely stabilizes Q_8 into place. The great distance between the interacting partners results in a rather weak stabilizing effect, as reflected by the high B-value of

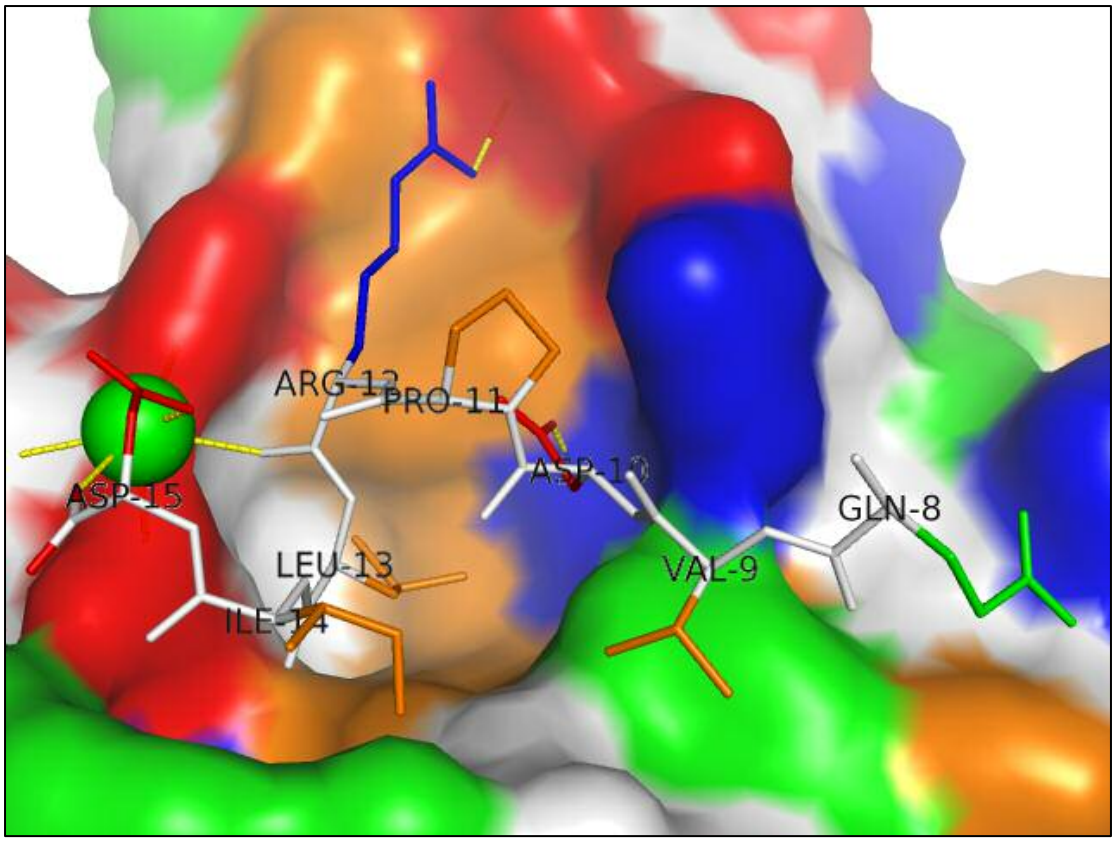


Figure 33: Detailed view of the peptide and peptide-binding site. Antibody: solvent-accessible surface representation; peptide: sticks representation; calcium: green sphere. Peptide and antibody are color-coded according to chemical properties (blue: basic, red: acid, white: backbone, orange: hydrophobic, green: uncharged polar). The heavy chain is visible in the upper portion of the image, and the light chain is visible in the lower portion.

the glutamine sidechain (in the 60-80 range). Still, it is sufficient to maintain it stable enough for X-ray diffraction to occur. The V_9 residue is involved in a hydrophobic interaction with Y_{94} in the CDR-L3. The backbone atoms of the rest of the CDR-L3 constitute a small and rather shallow pocket providing an additional surface feature into which V_9 is loosely held. The next residue, D_{10}, is maintained tightly in place by an ion pair with H_{50} and a hydrogen bond with W_{52}, both in the heavy chain strand immediately preceding the CDR-H2. These strong interactions pull the acidic sidechain towards the bottom of the antigen binding groove. The next residue is the P_{11} proline. During the structure determination, a *cis*-proline isomer was assessed, but did not fit the density as well as the regular *trans* isomer, which was consequently

adopted. As a result of backbone geometry constraints, the sidechain of this residue is pointing towards the solvent cavity. This is a little surprising considering the nonpolar character of the proline heterocycle. However, this may result from the geometric constraints imposed to the backbone by the surrounding residues. Indeed, R_{12} is tightly maintained in its position by a strong ion pair with D_{54} of the CDR-H2. In addition, this residue is also maintained by a cation-π interaction between the nitrogen atoms of the guanidinium group and the aromatic system of W_{53} in the CDR-H2. The combined strengths of these interactions may compensate for the unfavorable solvent exposure of P_{11}. R_{12} is the first of three residues present in a helical conformation. Instead of participating in hydrogen bonding with the nitrogen atom of residue n+4 as would classically be seen in an α helix, the carbonyl oxygen of R_{12} is in chelating contact with the calcium ion. This is possible because the corresponding hydrogen bond acceptor (which would be D_{15}) does not belong to the helix, thus letting the R_{12} carbonyl free and available for interacting with the calcium ion. The upwards direction of the carbonyl group is maintained by the next residue, L_{13}, whose sidechain is present in a very deep hydrophobic cavity of the antigen binding site. The bottom of this cavity is formed by three residues (M_{95}, A_{100A} and M_{100B}) which surround the CDR-H3. Two tryptophan residues, W_{91} in the light chain and W_{52} in the heavy chain, constitute the sides of the hydrophobic pocket which docks L_{13} into place. The downwards position of this leucine is important for calcium binding, as it forces the carbonyl oxygen of the previous residue (R_{12}) in an upwards position. The next residue, I_{14}, is surprisingly exposed to solvent for a hydrophobic residue. Although the hydrophobic residues V_9 and P_{11} in the peptide are in close proximity,

the surface offered for binding to the light chain is mostly polar. Indeed, T_{31} and N_{92} are the closest sidechains available. Although the light chain also has the hydrophobic residues Y_{32} and W_{91} close by, their sidechains are pointing towards the core of the protein, and their backbone atoms participate in hydrogen bonding with one shared water molecule. Thus, this residue appears to be mostly stabilized by van der Waals interactions with the antibody. The last residue of the peptide, D_{14}, is involved in the shared binding of calcium together with the CDR-H3 in the antibody. Both its carboxylic acid groups (one in the sidechain and one at the carboxy terminus) participate in the coordination shell of the ion. Noticeably, D_{14} does not have a binding site on the antibody, as the high density of negative charges present in the CDR-H3 would predict. The situation is however very different when the ion is present. In this case (seen in the structure), the binding of calcium not only shields the peptide from the negative charges of the CDR-H3, but drives the binding event through the energy gain offered when completing the full coordination shell of the calcium ion. The next section will describe the calcium binding site in more details.

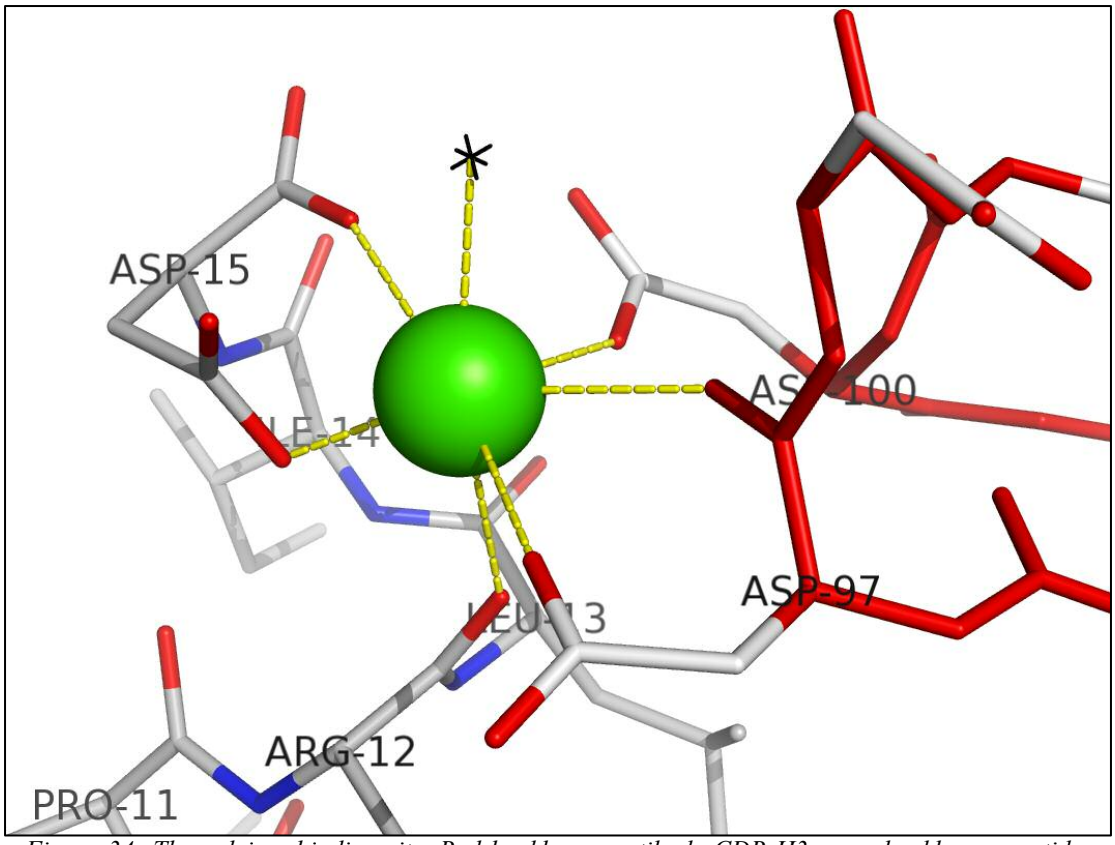


Figure 34: The calcium binding site. Red backbone: antibody CDR-H3; grey backbone: peptide; green sphere: calcium ion; black cross: water molecule. The peptide and the antibody both provide 3 coordination positions. The top vertex of the coordination shell is occupied by a water molecule.

The calcium-binding site has a pentagonal bipyramid geometry (7 coordination sites), with both the peptide and the antibody heavy chain participating in binding (fig. 34). Noticeably, the antibody only binds calcium through atoms of the CDR-H3, which provides two carboxylates (D_{97}, D_{100}) and one backbone carbonyl (D_{97}). The peptide also provides a carbonyl (R_{12}) and two carboxylates (both the sidechain and the carboxy-terminus of D_{15}). The electron density for the ion does not appear as an unconnected sphere, but rather as a protrusion connecting the carbonyls of R_{12} in the peptide and of D_{97} in the CDR-H3. The resulting distances of the oxygen atoms with the ion are of 2.63 and 2.70Å, respectively. The pentagon forming the base of the bipyramid comprises all three calcium binding CDR-H3 atoms and is completed by the two carboxylic acid groups of the peptide. It is not strictly on a

geometric plane, as the carboxy terminal oxygen of the peptide is located about 2Å above the other atoms. The bipyramid is completed by the carbonyl of R_{12} in the peptide (bottom vertex), and a water molecule 2.43Å above the calcium ion forming the top vertex. The calcium is located almost exactly at the intersection between the planes formed by the carboxylic acid functions of D_{97} and D_{100} in the antibody and of the peptide carboxy terminus. The calcium is noticeably out of the planes of the carbonyl functions or of the sidechain carboxylate of D_{15}. To conclude, atoms from the antibody, the peptide epitope and a water molecule participate in all possible coordination positions of the calcium ion, resulting in slightly distorted pentagonal bipyramid geometry. The composition of our calcium binding site is in relative agreement with a statistical analysis that we carried out on the 10605 calcium binding sites present in the 2006 release of the PDB using an in-house made python program. In this analysis, it was found that the "average" calcium binding site contains 2.9 oxygen atoms from carboxylic acid functions (glutamic acid or aspartic acid), 0.77 oxygen atom from water molecules, 2.89 backbone carbonyl oxygen atoms, 0.34 asparagine side chain oxygen atom and 0.06 rare binders like threonine or glutamine side chain oxygen atoms. The calcium binding site seen in the present structure has a slight overrepresentation of carboxylate oxygens (4 instead of about 3 in average), and a slight underrepresentation in backbone carbonyl oxygens (2 instead of about 3 in average). The presence of a single water molecule is in good agreement with the average 0.77. When considering that the carboxylate oxygen is probably replaced by a carbonyl oxygen in GDPC (see discussion), then the calcium binding site has a canonical composition.

4. DISCUSSION

4.1. Packing environment

Except in a few cases (notably insulin, oxidases in the peroxisome, or hemoglobin in the hemoglobin C disease, as reviewed elsewhere[143]), proteins rarely form crystal packing interactions in their biological context. As a consequence, it is important to assess whether the conformation observed in a crystal structure is affected by crystal contacts or not. Although in principle the physiologic conformation may well pack readily into a crystalline lattice, it is of importance in any crystal structure endeavor to assess the effect of crystal packing on the natural conformation of the protein and hence on the interpretation of its function. Apart from their contributions to crystal formation, packing interactions can be detrimental to structural studies because they are capable of distorting flexible loops into non-physiologic conformations. Although in most instances, the structures obtained by X-ray crystallography are compatible with results from techniques operating in solution (essentially NMR or other spectroscopic methods), the validity of a crystal structure always needs to be assessed.

To address packing environment, the interactions with the solvent that occur in the crystalline phase can be considered. As a first approximation, the Matthews coefficient[144] V_M can be used to that effect. It is a measure of the "protein density" inside the crystal, and indirectly provides information on its solvent content. It is analogous to a partial specific volume, and is expressed in $Å^3$/Da. Usual values of V_M are between 1.5$Å^3$/Da for densely packed crystals and 4.0$Å^3$/Da or more for loosely

packed crystals. In our case, the Matthews coefficient was of 2.4Å^3/Da, a value well within normal ranges of packing density. A recent statistical analysis of the distribution of Matthews coefficients in the PDB elegantly demonstrated a correlation between densely packed crystals and diffraction to higher resolutions[145]. Evidently both high resolution and limited packing interactions are desirable, but this statistical trend indicates that these goals are most often incompatible. In our case, the resolution of 2.3Å seems to be in a good agreement with this correlation, as V_M in the order of 2.2-2.8Å^3/Da were found to be the most frequent in structures with resolutions comprised between 2.01 and 2.40Å (average 2.6Å^3/Da). Still, although it is a good indicator of the tightness of the crystal packing, the solvent content of a protein is not a direct indicator of the nature of the crystal contacts. By examining the crystal structure it is possible to characterize which surfaces of the protein participate in packing interactions.

As seen in figure 32, most of the crystal contacts in our structure appear on the sides of the molecule, with little packing interaction occurring at the antigen. The distance threshold used on this figure was set at 4Å, a value confidently greater than typical noncovalent bond distances to ensure that the areas that appear free of contact really are at appreciable distance from the next crystal neighbor. The antigen is surrounded by a consequent solvent channel, with only the second residue of the peptide participating in a weak ionic pair with a neighboring molecule. The fact that the first residue in the peptide is invisible in the structure, together with the high B-factors of the second residue indicate that these residues are very mobile, and that the observed crystalline contact may actually be rather transient. In our case, the mere

presence of calcium and peptide in the antigen binding site is a good indication that the potential distortions introduced by the crystallization are not significant enough to disrupt the function of the protein. Hence it is reasonable to think that our structure is similar to the one that would be observed in solution.

4.2. On the structure of the protein C zymogen

Crystallization of GDPC in complex with HPC-4 and calcium was attempted during this project, but unfortunately without success despite the large number of trials. Consequently, the structure we obtained provides only a limited amount of information on the conformation of the activation region of GDPC in the zymogen form. In this section, the conformation of the epitope peptide and its relevance to the structure of zymogen protein C will be discussed.

Although the sequence of a protein is the most important determinant of its structure, the inherent flexibility of polypeptide chains tends to make them adopt a range of conformations around one or more energy minima. This plasticity is often functionally important for instance in enzyme mechanisms, where conformational changes can sometimes be of significance in the catalytic cycle[146]. Other determinants of protein structure include pH or other solvent characteristics, temperature, and most importantly binding of effector molecules or ions, or structural environment. Concerning our structure, these considerations have important implications in terms of the validity of the zymogen protein C peptide conformation that is seen. The interaction with HPC-4 may stabilize the peptide into a conformation that would not be seen otherwise in solution. By the same reasoning, there is a priori no indication that the conformation adopted when binding to HPC-4 is similar to the

conformation of the catalytic complex within the active site of thrombin in the T-TM activation reaction.

Still, the calcium dependence of the binding event and the ability of the protein C peptide to directly bind calcium are somewhat intriguing in this context. It is tempting to propose the hypothesis that the conformation seen in complex with HPC-4 is one and the same with the conformation adopted in the active site of the T-TM complex. One may then propose a model in which a calcium binding site is located in the zymogen protein C at the interface between the activation region and the calcium binding loop. In this model, the coordination positions provided by the activation region would be those seen in the HPC-4 structure, where the P1 arginine and the P3' aspartic acid provide oxygen atoms for chelation. The most compelling argument in favor of this otherwise fragile hypothesis comes from a mutagenesis study of protein C. In this particular study, the acidic residues in P3 and P3' were mutated to glycines, and about a third of the calcium affinity was lost[39]. The loss of affinity would be in agreement with the hypothesis where calcium is bound directly by P3' as seen in our structure. Furthermore, shared calcium binding would also provide a likely mechanism for the conformational interplay occurring in between the two regions during calcium titration experiments, for which there is extensive evidence. For instance, the intrinsic fluorescence change observed by titrating protein C with calcium is lost upon activation, and the introduction of fluorescent reporters in the activation region have a dramatic impact on the fluorescence response to calcium titration[14]. Furthermore, modeling studies did not rule out the possibility of a direct interaction taking place between the calcium binding loop and the activation region in

the zymogen form of protein C[40]. The basic residues in the calcium binding loop of protein C have been proposed to interact with thrombomodulin[26] leaving the acids in P3 and P3' available for ionic interaction. However, several observations raise questions about this candidate mechanism.

First and foremost, this hypothesis is hardly compatible with biochemical evidence. HPC-4 was initially studied for its ability to bind protein C[41]. If the hypothesis previously described were true, then there would be a steric clash between the GDPC calcium binding loop and the CDR-H3 of HPC-4 during the binding event, which would result in a dramatic conformational change of PC. Although not formally impossible, this is quite unlikely. Furthermore, the helical conformation seen in the structure would not be expected to be a good substrate to the T-TM complex. The helical conformation in itself is not unusual for a thrombin substrate, as evidenced by the thrombin structure in complex with fibrinogen (pdb code 1BBR[147]), in which a short helical segment is found 6 residues N-term from the scissile bond. It is rather the position of the helical segment, C-term from the scissile bond, which would make this conformation an unlikely fit in the active site of thrombin. Furthermore, thrombin substrates usually adopt a rather extended conformation. A second problem with this model is that extensive evidence across many coagulation factors argues that the calcium binding loop does not require additional coordinating atoms provided by a different loop in order to bind calcium with high affinity. In fact the shared binding of calcium by the activation region would rather be suboptimal compared to binding to the calcium binding loop alone, which has all coordination positions necessary. Finally, if the calcium binding mode seen in the HPC-4-peptide

structure has any relevance to the conformation of the zymogen, then there should be a resemblance between the CDR-H3 of HPC-4 and the calcium binding loop of protein C. However, these two loops are very different. The calcium binding loop of protein C is 10 residues long, and its first and last residues participate in the binding. In contrast, the CDR-H3 of HPC-4 only contains three residues, and calcium binding residues are only separated by two residues. Consequently, the protein C residues which bind calcium in the HPC-4 structure are unlikely to do so in the zymogen protein C in solution or in complex with T-TM.

This raises a question about the hypothetical structure of PC when in complex with HPC-4. Although we need to consider the possibility that the peptide conformation that we observe may not be the same as the one followed in the protein C-HPC-4 complex, several considerations argue otherwise. First, the activation region is known to be fairly flexible, a condition necessary for its recognition by the rather deep groove of the active site of thrombin. This flexibility can also be invoked to argue that even in the structural background of PC, there are still enough degrees of freedom in the epitope to allow it to adopt the conformation seen in our structure. Furthermore, the alternate case appears quite unlikely. Indeed, it would amount to HPC-4 having developed an antigen binding site able to recognize two different conformations of the same antigen, both in a calcium dependent fashion. This is highly unlikely considering the scarcity of antibodies having a calcium binding site. Regarding the calcium binding, one could argue that the carboxylic acid present at the terminus of the peptide is absent from PC, and hence that the conformation of the peptide is necessary different. However, the carboxy terminus would be replaced in

PC by a backbone carbonyl. The corresponding oxygen, although less electronegative than the one of a carboxy terminus, is still partially charged because of the resonant property of the peptide bond. The enhanced structural rigidity of a polypeptide backbone over a carboxy terminus could compensate for this less than optimal chemical function. Noticeably, the replacement of the carboxy terminus by a carbonyl would constitute a more usual composition for the calcium binding site, with 3 acidic oxygens, 3 carbonyl oxygens and one water molecule. Incidentally, the present position of the carboxy terminus would orient the next two residues of PC in a position which does not overlap with any antibody atom, consistent with the absence of binding of the D_{15}-Q_{27} peptide[41]. The same is true for the N-terminus of the peptide, which points away from the antibody in a direction where the PC protease domain would not clash with the antibody.

4.3. Correlation with fluorescence results

The fluorescent properties of tryptophan are to some extent changed by polarity variations in its immediate chemical environment. Fluorescence changes in the spectral region of maximal emission for tryptophan were observed in the ncFab in the presence of calcium and epitope peptide. In the absence of structural information, it is often difficult to infer the nature of the changes that occur in the vicinity of the reporter residues, but the structure we obtained allows some hypotheses. Two mechanisms seem most likely in our case. First, the fluorescence change could directly report the binding of calcium in the immediate vicinity of one or more tryptophan residues. The presence of a positive electrostatic charge would likely

modify the electronic system of the aromatic rings of tryptophan, causing the fluorescence change. The other possibility would be that binding of calcium and peptide transmit conformational changes to one or more tryptophan residues, causing the polarity of their immediate environment to change, and inducing their increased fluorescence response. These two mechanisms are not mutually exclusive, and are not the only ways to explain the observed fluorescence change. Our structure represents the endpoint of the structural transition initiated by the binding of calcium and peptide. Hence, only assumptions can be made as to the nature of the structural changes responsible for the fluorescence change.

Several tryptophan residues appear in direct contact with the peptide. At the bottom of the combining site, W_{91} (light chain) and W_{52} (heavy chain) most likely see their environment change upon calcium and peptide binding, either through hydrophobic interaction for W_{91} or through ionic interaction for W_{52}. The electronic system of W_{53} is also most likely affected by the cation-π interaction that occurs with the ε nitrogen of R_{12} in the bound peptide. The fluorescence response observed for HPC-4 ncFab upon calcium and epitope peptide binding can thus be rationalized without any kind of conformational rearrangement to the combining site. As will be discussed in the next section however, such conformational changes are likely to occur, and could in principle participate to the fluorescence response described above. In particular, the ion pair between R_{12} in the peptide and D_{54} in the CDR-H2 likely pulls the backbone in a motion that could transmit to the two preceding tryptophan residues W_{53} and W_{52}. This effect does not directly require the presence of calcium, and could explain the mild response obtained when titrating the peptide in the

absence of the ion[41]. The weak fluorescence response we observed in the absence of peptide, which was reported also for HPC-4, can be explained with the present structure. Indeed, the antibody and the water alone account for most of the coordination positions used to bind calcium. The possible conformational changes occurring when calcium binds to the CDR-H3 would transmit to tryptophan W_{52}, transmitted through M_{95}. The movement of CDR-H3 would also likely affect the tryptophans W_{52} and W_{53} of the CDR-H2 by conjugated movements transmitted through the intercalated CDR-H1 loop. Such a concerted movement of CDR loops would not be without precedent[49].

The reversal of the fluorescence intensity effect that was observed at the highest peptide concentration (fig. 21) can also be tentatively addressed with our structure. Although only the peptide-HPC-4 complex has a full calcium binding site, the peptide still participates in 3 coordination positions, which amounts to as many positions as the antibody provides. The calcium could thus in principle bind to the peptide alone, although with a lower affinity. By increasing the peptide concentration, one could then compensate the low affinity of the half calcium binding site on the peptide, and to start having significant binding nonetheless. This would scavenge the ion away from the antibody, which is the only molecule providing a measurable fluorescent signal. This would then explain why the fluorescence gain decreases at high peptide concentration. Also, it is seen that the full calcium binding site is completed by a D-X_1-X_2-D motif in the antibody (where X_1 is D_{98} and X_2 is Y_{99}). Coincidentally, the first four residues of the peptide used in our fluorescence studies also satisfy this motif, with two aspartic acid present two residues apart (D_7-Q_8-V_9-

D_{10}). It is thus possible to invoke a model in which calcium would mediate the dimerization of the peptide. In this model, one peptide would adopt the conformation seen in the antigen, and the second would form a complete calcium binding site by having its amino terminal segment mimic the CDR-H3 conformation of the antibody. The presence of additional binding sites in the peptide-HPC-4 interaction would explain the higher affinity for calcium of this complex compared to the hypothetical peptide dimer. The absence of fluorescent reporter residues in the peptide makes this hypothesis difficult to test.

4.4. Hypotheses about the mechanism of the switch

The mechanism of the conformational switch occurring during HPC-4 binding can only be determined by comparing our structure with the structure of HPC-4 in the absence of ligands. At the time of composition, crystals of the ncFab had been grown in 5mM EDTA, which diffracted below 2.25Å without any kind of optimization, but were too twinned to provide a dataset. This result is very promising, and may result in a second structure that will be used to test the following hypotheses.

Examination of the distances between charged or partially charged oxygen atoms in the calcium binding site reveal that they are significantly shorter than anywhere else in the protein. By generating a sorted list of the interatomic distances between charged oxygens in the whole structure, it appears that 10 out of the 15 pairs found below 4.5Å are clustered in the calcium binding site (the average distance being 36Å). The four shortest distances are also found there, with the exception of the distance between E_{79} and E_{81} in the light chain, which is maintained by the proximity

of the positive charge of R_{61}. This finding, in addition to the established fact that ionic interactions occur between the positive charges of calcium and the negative charges of the carboxylates, indicates that the ion is responsible for these abnormally short distances. The removal of calcium and peptide would thus result in many negative charges in close proximity to one another. A reasonable expectation is that ionic repulsion is going to take place and will result in the adoption of greater distances between charged oxygens, to conform to more typical distances as seen elsewhere in the structure. Although some of the movement could be accommodated by mere sidechain rotations, backbone rearrangements could be involved as well. Especially, the CDR-H3 conformation seen in the present structure has a pronounced backbone kink in the Y_{99}-D_{100} dipeptide, which is resolved into the next strand by a short helical segment. As a consequence, the CDR-H3 is almost contained in a plane orthogonal to the one formed by strands 6 and 7. This backbone kink is energetically unfavorable, but is probably maintained nonetheless because of the calcium binding enthalpy gain. However in the absence of the ion, this conformation would be expected to be shifted, driven by electrostatic repulsion, in the looser conformation usually seen for CDRs, where the loop is approximately in the same plane as the sheet from which it protrudes. With such a short segment between strands, one might even expect to see a canonical β-β hairpin at this location.

The presence of the negative charge cluster seen in the complex has consequences on the possible mechanism of the binding event. Indeed, it indicates that binding of the peptide to the antigen in the absence of calcium is confronted with a high energy barrier caused by electrostatic repulsion. If at all even possible (as

slight fluorescence changes might indicate), this event would be expected to be unlikely. On the other hand, binding of calcium to the antibody alone does not face any foreseeable energy barriers other than the displacement of whichever positive charge currently neutralizes the acidic functions of the CDR-H3 (in all likeliness a sodium ion from the buffer). Unlike binding of the peptide alone, binding of calcium alone would thus be expected to be more favorable. The Ca^{2+}-HPC-4 complex would then readily bind the peptide because the negative charges of the antibody would be shielded, and because of the energy contribution incurred by completing the coordination shell of calcium. This mechanism is consistent with the characteristic sharpness of the elution peaks obtained on the epitope affinity chromatography (see figs. 18 and 19). In this column, elution is performed by replacing calcium by EDTA in the running buffer. The removal of calcium from the ternary complex then transforms the high affinity binding occurring throughout the peptide into a strong ionic repulsive interaction occurring at the carboxy terminal helix, poorly compensated by the remaining binding sites of the median part of the peptide.

4.5. On the affinity of the peptide-Ca^{2+}-HPC-4 interaction

One of the most striking features of the antigen binding site of HPC-4 is the great number of ionic interactions. Besides the ionic character of calcium binding, two additional ion pairs are found. This observation contrasts with the classical view of antigen binding sites, in which interactions are dominated by van der Waals contacts and hydrogen bonds[52]. Ionic interactions figure among the most stable noncovalent interactions, so their presence may be of functional significance. The

activation region of blood coagulation serine proteases is known to be very mobile, based on the high B-factor values observed in this region of factor VII[148] and of the prethrombin-2 structure[149]. This flexibility in protein C is accommodated by thrombin through ion pairing with R_{15} and β-sheet hydrogen bonding between the active site and the substrate. In HPC-4 however, the fact that all but one charged group of the epitope are faced with an opposite charge in HPC-4 may be a mechanism by which the antibody accommodates the flexibility of the antigen. Noticeably, the R_{15} sidechain central to the peptide is found to be in an ion pair both in the interaction with HPC-4 and with thrombin, where the residue D_{189} pulls and maintains it deep within the active site.

These considerations must however be put into perspective especially regarding the stability of the complex to high salt concentrations. During the course of the project, an experiment was done in which ncFab fragments in complex with the epitope peptide were exposed to 2M salt (see section #3.1.2., fig. 18). This experiment was performed in an attempt to selectively detach ncFab or Fab from the peptide affinity column in order to separate them. It was not a good method for the intended goal (both fragments were equally retained), but it still demonstrated that the interaction can withstand very high salt concentrations. If ionic interactions were important for affinity, one would expect high salt concentrations to be disruptive by virtue of competition of the interacting charges with the sodium and chloride ions present in solution. The resistance of the interaction to high salt can be rationalized by the structure of the complex and by considering the underlying electrostatic attraction. This force is governed by Coulomb's law, which states that it is inversely

proportional to the permittivity of the medium in which the charges are located. The degree of solvent exposure would be expected to affect the local permittivity, and consequently the force of interaction of a given ion pair. In the structure, it can be seen that most ion pairs are located relatively deep in the binding site, where the nonpolar environment would strengthen the interaction. This location towards the central core of the structure would also prevent the ions in the solution from penetrating merely because of their size, which would decrease their effective diffusivity in the vicinity of the ion pair and thus their ability to disrupt it. The enhancement of hydrophobic interactions that is obtained when increasing the salt concentration must also be considered. Thus, the absence of dissociation at high salt indicates that the increase in hydrophobic interaction strength is able to compensate for the relative weakening of ionic interactions.

In the same setting, the fragments were submitted to 2M salt in the presence of 0.5M guanidine, a potent chaotrope agent. The fragments did not detach either, in agreement with the absence of hydrogen bonds seen in the structure.

The behavior of the complex in high salt is also interesting to analyze in terms of calcium binding. The high concentration of sodium ions present in the wash solution might be expected to disrupt the calcium binding by neutralizing the negative charges of the carboxylates. The failure of sodium to compete calcium out of the ternary complex could be attributed to the suboptimal coordination geometry offered by the calcium binding site to sodium. It could also indicate that a fully occupied calcium binding site is to some extent more favorable energetically than the sodium salt form of the separated species. This emphasizes the importance of calcium binding

not only as an electrostatic shield to the repulsion between peptide and antibody, but also as an important energetic contribution through the completion of a full coordination shell.

The retention of the fragments that was obtained in low salt and even in low salt plus 30% ethanol must also be analyzed. Indeed, the previous result obtained in high salt indicated that hydrophobic interactions have an important contribution to the binding strength. By decreasing the polarity of the solvent, hydrophobic interactions are supposed to weaken, yet 0.01M NaCl or even 0.01M NaCl with 30% ethanol were not sufficient to achieve dissociation in the above series of experiments. Thus, these particular ion pairings can compensate the loss or decrease of hydrophobic interaction caused by the combined effects of a 200× reduction in salt concentration and 30% ethanol.

Taken together, these data indicate that the binding strength of the peptide to HPC-4 is contributed both by hydrophobic interactions and by ionic interactions. In addition, calcium provides significant contributions both through ionic shielding and through completion of a full coordination shell. The interacting partners identified at the molecular level in the structure are in agreement with the macroscopic effects seen in affinity chromatography results.

4.6. Comparison with Q-425

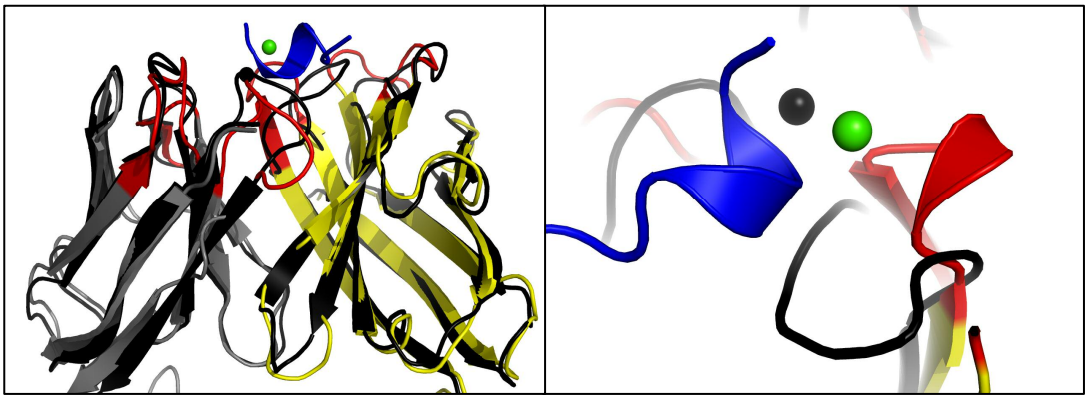


Figure 35: Overlay of HPC-4 with Q-425 (pdb code 2ADJ). Left pane: overview of the variable domains of both antibodies. Q-425 is represented in black; HPC-4 is represented with the same color coding as in fig. 31. Right pane: close-up view of the CDR-H3. In Q-425, CDR-H3 protrudes towards the antigen binding site, whereas in HPC-4, it is kinked backwards and away from the antigen. The alignment rms is 0.938Å. The calcium in Q-425 (black sphere) is located 4.77Å from the calcium in HPC-4.

HPC-4 was the first antibody with a characterized calcium binding site, and the present work provides a basis for comparison with the only other example of a structurally solved calcium dependent antibody[48]. The extent of the comparison is, however, limited by the composition of the complexes. In the case of Q-425, the structure is available in the presence or absence of calcium, but not for the ternary complex whereas in our case, the structure of the ternary complex is the only one currently available. This will no longer be true in the near future, as we were able to grow diffracting crystals of the ncFab in the absence of calcium or peptide which will soon yield the structure of the unliganded form of the HPC-4 ncFab.

The antigen binding sites of HPC-4 and Q-425 are appreciably different in the path adopted by the backbone of the CDRs and especially in CDR-H3. In our structure, this CDR forms the rim of the binding site, and is kinked away from the groove. In Q-425, the CDR has one more residue, and the resulting loop forms the

bottom base of the combining site by protruding forward into the antigen binding groove. It extends so far that it comes in direct contact with the CDR-H2 and by doing so prevents the access of the CDR-H1 to the binding site. In HPC-4, the CDR-H1 contributes to the rim of the antigen binding site but does not directly interact with the antigen. It is however in close contact with W_{53} in the CDR-H2 and D_{97} in the CDR-H3, which are direct antigen ligands.

More importantly, the two antibodies seem to differ by the mechanisms underlying their calcium dependence. Surface plasmon resonance and structural studies demonstrated that the increase in antigen affinity occurring upon calcium binding to Q-425 arises only from minor structural rearrangements. In this case, the effect of calcium on affinity is thus merely limited to the completion of the coordination shell. This absence of conformational rearrangement can be attributed to the presence of only 2 charged oxygen atoms in the surroundings of the calcium ion. In the ion-bound structure, these atoms are located 2.49Å apart, and they shift to 3.10Å in the calcium-free form without involving important backbone rearrangements. In the case of HPC-4 discussed earlier, in addition to the bridging mechanism seen in the structure, it is believed that binding of calcium to the antibody is inducing conformational rearrangements. Thus in Q-425, the ion strictly acts as a bridge between antigen and antibody, whereas in HPC-4 the ion has the additional effect of shaping the antigen binding site into a high affinity conformation.

In the report of the Q-425 structure, the authors attributed the scarcity of calcium-dependent antibodies seen in the PDB to some selective disadvantage of this particular mode of antigen binding. The increased rigidity of the antigen binding site

conferred by calcium binding and the multicoordinated nature of the binding were proposed as detrimental. This reasoning is based on the assumption that the structures available in the PDB represent an accurate depiction of the variety of binding modes that exist naturally, which may or may not be the case considering the many millions of antibodies present in a typical immune system[150], or the estimated $\sim 1.7 \times 10^8$ different combinations potentially accessible by the combinatorial effects of V(D)J recombination, somatic hypermutation, and class switching[59]. The requirement for calcium was also proposed to be favorable only through the increased binding energy for antigens with limited surface available for binding such as the heavily glycosylated Q-425 antigen CD-4. Why the glycosylated surfaces would be unavailable for typical antigen interaction remains unclear, since there are examples of carbohydrate specific antibodies for instance in allergies[151] or in heparin-induced thrombocytopenia[152] to name a few. Still, this would be in relative agreement for HPC-4, where the binding strength of calcium is taken advantage of to accommodate the unusual flexibility of the antigen.

5. CONCLUSION

We have solved the 2.3Å structure of the calcium-dependent antibody HPC-4 ncFab in complex with calcium and the epitope peptide. The structure reveals a singular and unprecedented binding mode for calcium. In contrast with the only other known structure of a calcium-binding antibody, the calcium ion is believed to induce profound conformational changes to a region of the backbone of HPC-4, in addition to its more conventional role as a bridge between antibody and antigen. Thus, our findings represent an important step forward in the understanding of the structural biology of metalloproteins in general and immunoglobulins in particular.

Together with the initial biochemical characterizations of HPC-4, the present work provides a solid experimental framework which could form the base of many additional studies. As announced earlier, a diffracting crystal of the ncFab grown in the absence of peptide or calcium was obtained, and diffracted to promising resolutions. This will likely yield a second structure in the near future. In the meantime, the structural rearrangements occurring upon calcium and peptide binding that we postulated based on our structure represent testable hypotheses that could in principle be addressed using nuclear magnetic resonance spectroscopy techniques. In addition, the different residues involved in the particularities of HPC-4 identified in our structure could be modified through selective mutagenesis in order to better dissect their relative contributions to the overall function of HPC-4.

The calcium dependence of this antibody has a strong potential also in terms of bioseparations and biosensing. In these fields, the high affinity of antibodies

generally prevents them from being used because the conditions required to dissociate the antigen of interest (for instance a therapeutic agent in bioseparations or a measured analyte in biosensing) are often incompatible with operation requirements. Thus, the ability to rationally design antibodies with reversible binding properties represents an important future milestone. Our structure provides a solid foundation which can readily be used to that effect.

BIBLIOGRAPHY

(1) P. H. Reitsma; *Protein C deficiency: from gene defects to disease*; Thromb Haemost; 78(1); 344-50; 1997
(2) C. T. Esmon; *The protein C pathway*; Chest; 124(3 Suppl); 26S-32S; 2003
(3) C. T. Esmon, N. L. Esmon; *Protein C activation*; Semin Thromb Hemost; 10(2); 122-30; 1984
(4) C. T. Esmon; *Regulation of blood coagulation*; Biochim Biophys Acta; 1477(1-2); 349-60; 2000
(5) N. L. Esmon, W. G. Owen, C. T. Esmon; *Isolation of a membrane-bound cofactor for thrombin-catalyzed activation of protein C*; J Biol Chem; 257(2); 859-64; 1982
(6) K. Fukudome, C. T. Esmon; *Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor*; J Biol Chem; 269(42); 26486-91; 1994
(7) F. B. Taylor, Jr., G. T. Peer, M. S. Lockhart, G. Ferrell, C. T. Esmon; *Endothelial cell protein C receptor plays an important role in protein C activation in vivo*; Blood; 97(6); 1685-8; 2001
(8) J. M. Stassen, J. Arnout, H. Deckmyn; *The hemostatic system*; Curr Med Chem; 11(17); 2245-60; 2004
(9) K. Gomez, J. H. McVey, E. Tuddenham; *Inhibition of coagulation by macromolecular complexes*; Haematologica; 90(11); 1570-6; 2005
(10) J. K. Leach, E. Patterson, E. A. O'Rear; *Improving thrombolysis with encapsulated plasminogen activators and clinical relevance to myocardial infarction and stroke*; Clin Hemorheol Microcirc; 30(3-4); 225-8; 2004
(11) S. Tanabe, T. Sugo, M. Matsuda; *Synthesis of protein C in human umbilical vein endothelial cells*; J Biochem (Tokyo); 109(6); 924-8; 1991
(12) D. C. Foster, C. A. Sprecher, R. D. Holly, J. E. Gambee, K. M. Walker, A. A. Kumar; *Endoproteolytic processing of the dibasic cleavage site in the human protein C precursor in transfected mammalian cells: effects of sequence alterations on efficiency of cleavage*; Biochemistry; 29(2); 347-54; 1990
(13) B. W. Grinnell, J. D. Walls, B. Gerlitz; *Glycosylation of human protein C affects its secretion, processing, functional activities, and activation by thrombin*; J Biol Chem; 266(15); 9778-85; 1991
(14) A. R. Rezaie, C. T. Esmon; *Tryptophans 231 and 234 in protein C report the Ca(2+)-dependent conformational change required for activation by the thrombin-thrombomodulin complex*; Biochemistry; 34(38); 12221-6; 1995
(15) P. Fernlund, J. Stenflo; *Amino acid sequence of the light chain of bovine protein C*; J Biol Chem; 257(20); 12170-9; 1982
(16) J. Stenflo, P. Fernlund; *Amino acid sequence of the heavy chain of bovine protein C*; J Biol Chem; 257(20); 12180-90; 1982
(17) T. Mather, V. Oganessyan, P. Hof, R. Huber, S. Foundling, C. Esmon, W. Bode; *The 2.8 A crystal structure of Gla-domainless activated protein C*; Embo J; 15(24); 6822-31; 1996
(18) W. Bode, D. Turk, A. Karshikov; *The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis,*

overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships; Protein Sci; 1(4); 426-71; 1992
(19) S. Kurosawa, D. J. Stearns, K. W. Jackson, C. T. Esmon; *A 10-kDa cyanogen bromide fragment from the epidermal growth factor homology domain of rabbit thrombomodulin contains the primary thrombin binding site*; J Biol Chem; 263(13); 5993-6; 1988
(20) M. Zushi, K. Gomi, S. Yamamoto, I. Maruyama, T. Hayashi, K. Suzuki; *The last three consecutive epidermal growth factor-like structures of human thrombomodulin comprise the minimum functional domain for protein C-activating cofactor activity and anticoagulant activity*; J Biol Chem; 264(18); 10351-3; 1989
(21) J. Ye, L. W. Liu, C. T. Esmon, A. E. Johnson; *The fifth and sixth growth factor-like domains of thrombomodulin bind to the anion-binding exosite of thrombin and alter its specificity*; J Biol Chem; 267(16); 11023-8; 1992
(22) M. Zushi, K. Gomi, G. Honda, S. Kondo, S. Yamamoto, T. Hayashi, K. Suzuki; *Aspartic acid 349 in the fourth epidermal growth factor-like structure of human thrombomodulin plays a role in its Ca(2+)-mediated binding to protein C*; J Biol Chem; 266(30); 19886-9; 1991
(23) T. Hayashi, M. Zushi, S. Yamamoto, K. Suzuki; *Further localization of binding sites for thrombin and protein C in human thrombomodulin*; J Biol Chem; 265(33); 20156-9; 1990
(24) L. Yang, A. R. Rezaie; *The fourth epidermal growth factor-like domain of thrombomodulin interacts with the basic exosite of protein C*; J Biol Chem; 278(12); 10484-90; 2003
(25) S. Krishnaswamy; *Exosite-driven substrate specificity and function in coagulation*; J Thromb Haemost; 3(1); 54-67; 2005
(26) P. Fuentes-Prior, Y. Iwanaga, R. Huber, R. Pagila, G. Rumennik, M. Seto, J. Morser, D. R. Light, W. Bode; *Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex*; Nature; 404(6777); 518-25; 2000
(27) A. O. Pineda, Z. W. Chen, S. Caccia, A. M. Cantwell, S. N. Savvides, G. Waksman, F. S. Mathews, E. Di Cera; *The anticoagulant thrombin mutant W215A/E217A has a collapsed primary specificity pocket*; J Biol Chem; 279(38); 39824-8; 2004
(28) G. Lu, S. Chhum, S. Krishnaswamy; *The affinity of protein C for the thrombin-thrombomodulin complex is determined in a primary way by active site-dependent interactions*; J Biol Chem; 2005
(29) C. H. Croy, J. R. Koeppe, S. Bergqvist, E. A. Komives; *Allosteric changes in solvent accessibility observed in thrombin upon active site occupation*; Biochemistry; 43(18); 5246-55; 2004
(30) M. A. Parry, S. R. Stone, J. Hofsteenge, M. P. Jackman; *Evidence for common structural changes in thrombin induced by active-site or exosite binding*; Biochem J; 290 (Pt 3)(665-70; 1993
(31) A. R. Rezaie, S. T. Cooper, F. C. Church, C. T. Esmon; *Protein C inhibitor is a potent inhibitor of the thrombin-thrombomodulin complex*; J Biol Chem; 270(43); 25336-9; 1995
(32) J. Ye, N. L. Esmon, C. T. Esmon, A. E. Johnson; *The active site of thrombin is altered upon binding to thrombomodulin. Two distinct structural changes are*

detected by fluorescence, but only one correlates with protein C activation; J Biol Chem; 266(34); 23016-21; 1991
(33) G. Musci, L. J. Berliner, C. T. Esmon; *Evidence for multiple conformational changes in the active center of thrombin induced by complex formation with thrombomodulin: an analysis employing nitroxide spin-labels*; Biochemistry; 27(2); 769-73; 1988
(34) J. G. Mandell, A. Baerga-Ortiz, S. Akashi, K. Takio, E. A. Komives; *Solvent accessibility of the thrombin-thrombomodulin interface*; J Mol Biol; 306(3); 575-89; 2001
(35) B. F. Le Bonniec, C. T. Esmon; *Glu-192----Gln substitution in thrombin mimics the catalytic switch induced by thrombomodulin*; Proc Natl Acad Sci U S A; 88(16); 7371-5; 1991
(36) B. F. Le Bonniec, R. T. MacGillivray, C. T. Esmon; *Thrombin Glu-39 restricts the P'3 specificity to nonacidic residues*; J Biol Chem; 266(21); 13796-803; 1991
(37) L. Yang, S. Prasad, E. Di Cera, A. R. Rezaie; *The Conformation of the Activation Peptide of Protein C Is Influenced by Ca2+ and Na+ Binding*; J Biol Chem; 279(37); 38519-38524; 2004
(38) N. L. Esmon, L. E. DeBault, C. T. Esmon; *Proteolytic formation and properties of gamma-carboxyglutamic acid-domainless protein C*; J Biol Chem; 258(9); 5548-53; 1983
(39) A. R. Rezaie, C. T. Esmon; *The function of calcium in protein C activation by thrombin and the thrombin-thrombomodulin complex can be distinguished by mutational analysis of protein C derivatives*; J Biol Chem; 267(36); 26104-9; 1992
(40) L. Perera, C. Foley, T. A. Darden, D. Stafford, T. Mather, C. T. Esmon, L. G. Pedersen; *Modeling zymogen protein C*; Biophys J; 79(6); 2925-43; 2000
(41) D. J. Stearns, S. Kurosawa, P. J. Sims, N. L. Esmon, C. T. Esmon; *The interaction of a Ca2+-dependent monoclonal antibody with the protein C activation peptide region. Evidence for obligatory Ca2+ binding to both antigen and antibody*; J Biol Chem; 263(2); 826-32; 1988
(42) A. R. Rezaie, T. Mather, F. Sussman, C. T. Esmon; *Mutation of Glu-80-->Lys results in a protein C mutant that no longer requires Ca2+ for rapid activation by the thrombin-thrombomodulin complex*; J Biol Chem; 269(5); 3151-4; 1994
(43) A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia; *SCOP: a structural classification of proteins database for the investigation of sequences and structures*; J Mol Biol; 247(4); 536-40; 1995
(44) E. A. Kabat, T. T. Wu, H. Bilofsky; *Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites*; J Biol Chem; 252(19); 6609-16; 1977
(45) E. A. Padlan; *On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands*; Proteins; 7(2); 112-24; 1990
(46) A. V. Collis, A. P. Brouwer, A. C. Martin; *Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen*; J Mol Biol; 325(2); 337-54; 2003

(47) S. Lea, D. Stuart; *Analysis of antigenic surfaces of proteins*; Faseb J; 9(1); 87-93; 1995
(48) T. Zhou, D. H. Hamer, W. A. Hendrickson, Q. J. Sattentau, P. D. Kwong; *Interfacial metal and antibody recognition*; Proc Natl Acad Sci U S A; 102(41); 14575-80; 2005
(49) Y. Li, H. Li, S. J. Smith-Gill, R. A. Mariuzza; *Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63* Biochemistry; 39(21); 6296-309; 2000
(50) R. L. Stanfield, M. Takimoto-Kamimura, J. M. Rini, A. T. Profy, I. A. Wilson; *Major antigen-induced domain rearrangements in an antibody*; Structure; 1(2); 83-93; 1993
(51) B. C. Braden, R. J. Poljak; *Structural features of the reactions between antibodies and protein antigens*; Faseb J; 9(1); 9-16; 1995
(52) D. R. Davies, G. H. Cohen; *Interactions of protein antigens with antibodies*; Proc Natl Acad Sci U S A; 93(1); 7-12; 1996
(53) B. C. Braden, B. A. Fields, R. J. Poljak; *Conservation of water molecules in an antibody-antigen interaction*; J Mol Recognit; 8(5); 317-25; 1995
(54) D. G. Covell, A. Wallqvist; *Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance of water molecules in the binding epitope*; J Mol Biol; 269(2); 281-97; 1997
(55) T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, R. J. Poljak; *Bound water molecules and conformational stabilization help mediate an antigen-antibody association*; Proc Natl Acad Sci U S A; 91(3); 1089-93; 1994
(56) P. Monecke, T. Borosch, J. Brickmann, S. M. Kast; *Determination of the interfacial water content in protein-protein complexes from free energy simulations*; Biophys J; 90(3); 841-50; 2006
(57) B. A. Fields, F. A. Goldbaum, W. Dall'Acqua, E. L. Malchiodi, A. Cauerhff, F. P. Schwarz, X. Ysern, R. J. Poljak, R. A. Mariuzza; *Hydrogen bonding and solvent structure in an antigen-antibody interface. Crystal structures and thermodynamic characterization of three Fv mutants complexed with lysozyme*; Biochemistry; 35(48); 15494-503; 1996
(58) J. Green; *The physicochemical structure of bone: cellular and noncellular elements*; Miner Electrolyte Metab; 20(1-2); 7-15; 1994
(59) L. Stryer; *Biochemistry*; Freeman & co.; 4th edition(1995
(60) D. W. Stafford; *The vitamin K cycle*; J Thromb Haemost; 3(8); 1873-8; 2005
(61) M. Sunnerhagen, T. Drakenberg, S. Forsen, J. Stenflo; *Effect of Ca2+ on the structure of vitamin K-dependent coagulation factors*; Haemostasis; 26 Suppl 1(45-53; 1996
(62) M. Soriano-Garcia, K. Padmanabhan, A. M. de Vos, A. Tulinsky; *The Ca2+ ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1*; Biochemistry; 31(9); 2554-66; 1992
(63) M. Sunnerhagen, S. Forsen, A. M. Hoffren, T. Drakenberg, O. Teleman, J. Stenflo; *Structure of the Ca(2+)-free Gla domain sheds light on membrane binding of blood coagulation proteins*; Nat Struct Biol; 2(6); 504-9; 1995

(64) L. Zhang, F. J. Castellino; *The binding energy of human coagulation protein C to acidic phospholipid vesicles contains a major contribution from leucine 5 in the gamma-carboxyglutamic acid domain*; J Biol Chem; 269(5); 3590-5; 1994
(65) V. Oganesyan, N. Oganesyan, S. Terzyan, D. Qu, Z. Dauter, N. L. Esmon, C. T. Esmon; *The crystal structure of the endothelial protein C receptor and a bound phospholipid*; J Biol Chem; 277(28); 24851-4; 2002
(66) L. M. Regan, J. S. Mollica, A. R. Rezaie, C. T. Esmon; *The interaction between the endothelial cell protein C receptor and protein C is dictated by the gamma-carboxyglutamic acid domain of protein C*; J Biol Chem; 272(42); 26279-84; 1997
(67) R. J. Preston, E. Ajzner, C. Razzari, S. Karageorgi, S. Dua, B. Dahlback, D. A. Lane; *Multifunctional specificity of the protein C/activated protein C Gla domain*; J Biol Chem; 281(39); 28850-7; 2006
(68) J. Stenflo, Y. Stenberg, A. Muranyi; *Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions*; Biochim Biophys Acta; 1477(1-2); 51-63; 2000
(69) K. E. Persson, J. Stenflo, S. Linse, Y. Stenberg, R. J. Preston, D. A. Lane, S. M. Rezende; *Binding of calcium to anticoagulant protein S: role of the fourth EGF module*; Biochemistry; 45(35); 10682-9; 2006
(70) C. H. Cheng, J. P. Geng, F. J. Castellino; *The functions of the first epidermal growth factor homology region of human protein C as revealed by a charge-to-alanine scanning mutagenesis investigation*; Biol Chem; 378(12); 1491-500; 1997
(71) B. J. Leonard, B. J. Clarke, S. Sridhara, R. Kelley, F. A. Ofosu, M. A. Blajchman; *Activation and active site occupation alter conformation in the region of the first epidermal growth factor-like domain of human factor VII*; J Biol Chem; 275(45); 34894-900; 2000
(72) M. Sunnerhagen, G. A. Olah, J. Stenflo, S. Forsen, T. Drakenberg, J. Trewhella; *The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca2+ binding to the first EGF domain. A combined NMR-small angle X-ray scattering study*; Biochemistry; 35(36); 11547-59; 1996
(73) M. Selander-Sunnerhagen, M. Ullner, E. Persson, O. Teleman, J. Stenflo, T. Drakenberg; *How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH2-terminal EGF-like domain in coagulation factor X*; J Biol Chem; 267(27); 19642-9; 1992
(74) Z. Rao, P. Handford, M. Mayhew, V. Knott, G. G. Brownlee, D. Stuart; *The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions*; Cell; 82(1); 131-41; 1995
(75) L. A. Gregory, N. M. Thielens, G. J. Arlaud, J. C. Fontecilla-Camps, C. Gaboriaud; *X-ray structure of the Ca2+-binding interaction domain of C1s. Insights into the assembly of the C1 complex of complement*; J Biol Chem; 278(34); 32157-64; 2003
(76) S. A. Jensen, A. R. Corbett, V. Knott, C. Redfield, P. A. Handford; *Ca2+-dependent interface formation in fibrillin-1*; J Biol Chem; 280(14); 14076-84; 2005
(77) W. Bode, P. Schwager; *The refined crystal structure of bovine beta-trypsin at 1.8 A resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0*; J Mol Biol; 98(4); 693-717; 1975

(78) J. Schiodt, N. Harrit, U. Christensen, L. C. Petersen; *Two different Ca2+ ion binding sites in factor VIIa and in des(1-38) factor VIIa*; FEBS Lett; 306(2-3); 265-8; 1992
(79) P. Wildgoose, D. Foster, J. Schiodt, F. C. Wiberg, J. J. Birktoft, L. C. Petersen; *Identification of a calcium binding site in the protease domain of human blood coagulation factor VII: evidence for its role in factor VII-tissue factor interaction*; Biochemistry; 32(1); 114-9; 1993
(80) A. K. Sabharwal, J. J. Birktoft, J. Gorka, P. Wildgoose, L. C. Petersen, S. P. Bajaj; *High affinity Ca(2+)-binding site in the serine protease domain of human factor VIIa and its role in tissue factor binding and development of catalytic activity*; J Biol Chem; 270(26); 15523-30; 1995
(81) D. W. Banner, A. D'Arcy, C. Chene, F. K. Winkler, A. Guha, W. H. Konigsberg, Y. Nemerson, D. Kirchhofer; *The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor*; Nature; 380(6569); 41-6; 1996
(82) A. Mathur, D. Zhong, A. K. Sabharwal, K. J. Smith, S. P. Bajaj; *Interaction of factor IXa with factor VIIIa. Effects of protease domain Ca2+ binding site, proteolysis in the autolysis loop, phospholipid, and factor X*; J Biol Chem; 272(37); 23418-26; 1997
(83) L. Chen, C. Manithody, L. Yang, A. R. Rezaie; *Zymogenic and enzymatic properties of the 70-80 loop mutants of factor X/Xa*; Protein Sci; 13(2); 431-42; 2004
(84) A. E. Schmidt, K. Padmanabhan, M. C. Underwood, W. Bode, T. Mather, S. P. Bajaj; *Thermodynamic linkage between the S1 site, the Na+ site, and the Ca2+ site in the protease domain of human activated protein C (APC). Sodium ion in the APC crystal structure is coordinated to four carbonyl groups from two separate loops*; J Biol Chem; 277(32); 28987-95; 2002
(85) A. E. Johnson, N. L. Esmon, T. M. Laue, C. T. Esmon; *Structural changes required for activation of protein C are induced by Ca2+ binding to a high affinity site that does not contain gamma-carboxyglutamic acid*; J Biol Chem; 258(9); 5554-60; 1983
(86) I. Schechter, A. Berger; *On the size of the active site in proteases. I. Papain*; Biochem Biophys Res Commun; 27(2); 157-62; 1967
(87) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne; *The Protein Data Bank*; Nucleic Acids Res; 28(1); 235-42; 2000
(88) D. A. Nicoll, M. R. Sawaya, S. Kwon, D. Cascio, K. D. Philipson, J. Abramson; *The crystal structure of the primary Ca2+ sensor of the na+/ca2+ exchanger reveals a novel Ca2+ binding motif*; J Biol Chem; 281(31); 21577-81; 2006
(89) M. Ottolia, K. D. Philipson, S. John; *Conformational changes of the Ca(2+) regulatory site of the Na(+)-Ca(2+) exchanger detected by FRET*; Biophys J; 87(2); 899-906; 2004
(90) L. J. Harris, E. Skaletsky, A. McPherson; *Crystallization of intact monoclonal antibodies*; Proteins; 23(2); 285-9; 1995
(91) L. J. Harris, E. Skaletsky, A. McPherson; *Crystallographic structure of an intact IgG1 monoclonal antibody*; J Mol Biol; 275(5); 861-72; 1998

(92) V. Chitarra-Guillon, H. Souchon, G. Boulot, M. M. Riottot, R. Mariuzza, D. Tello, R. J. Poljak; *Crystallization and preliminary X-ray diffraction studies of antigen--antibody complexes*; Biochimie; 70(8); 1045-52; 1988
(93) S. Covaceuszach, A. Cassetta, A. Cattaneo, D. Lamba; *Purification, crystallization, X-ray diffraction analysis and phasing of a Fab fragment of monoclonal neuroantibody alphaD11 against nerve growth factor*; Acta Crystallogr D Biol Crystallogr; 60(Pt 7); 1323-7; 2004
(94) M. Huang, B. C. Furie, B. Furie; *Crystal structure of the calcium-stabilized human factor IX Gla domain bound to a conformation-specific anti-factor IX antibody*; J Biol Chem; 279(14); 14338-46; 2004
(95) S. Li, K. R. Schmitz, P. D. Jeffrey, J. J. Wiltzius, P. Kussie, K. M. Ferguson; *Structural basis for inhibition of the epidermal growth factor receptor by cetuximab*; Cancer Cell; 7(4); 301-11; 2005
(96) P. C. Bourne, S. S. Terzyan, G. Cloud, N. F. Landolfi, M. Vasquez, A. B. Edmundson; *Three-dimensional structures of a humanized anti-IFN-gamma Fab (HuZAF) in two crystal forms*; Acta Crystallogr D Biol Crystallogr; 60(Pt 10); 1761-9; 2004
(97) V. Granata, N. G. Housden, S. Harrison, C. Jolivet-Reynaud, M. G. Gore, E. A. Stura; *Comparison of the crystallization and crystal packing of two Fab single-site mutant protein L complexes*; Acta Crystallogr D Biol Crystallogr; 61(Pt 6); 750-4; 2005
(98) X. Shi, Y. Li, C. Bian, G. Zhao, M. Huang; *Crystallization of an anti-factor IX antibody and its complex*; Acta Crystallogr D Biol Crystallogr; 61(Pt 6); 701-3; 2005
(99) M. Torres, N. Fernandez-Fuentes, A. Fiser, A. Casadevall; *The immunoglobulin heavy chain constant region affects kinetic and thermodynamic parameters of antibody variable region interactions with antigen*; J Biol Chem; 2007
(100) R. C. Blake, 2nd, X. Li, H. Yu, D. A. Blake; *Covalent and noncovalent modifications induce allosteric binding behavior in a monoclonal antibody*; Biochemistry; 46(6); 1573-86; 2007
(101) I. E. Liener, B. Friedenson; *Ficin*; Methods in Enzymology; 19(261-273; 1970
(102) D. N. Gosalia, C. M. Salisbury, J. A. Ellman, S. L. Diamond; *High Throughput Substrate Specificity Profiling of Serine and Cysteine Proteases Using Solution-phase Fluorogenic Peptide Microarrays*; Mol Cell Proteomics; 4(5); 626-636; 2005
(103) M. Mariani, M. Camagna, L. Tarditi, E. Seccamani; *A new enzymatic method to obtain high-yield F(ab)2 suitable for clinical use from mouse IgGl*; Mol Immunol; 28(1-2); 69-77; 1991
(104) Y. Zou, M. Bian, Z. Yiang, L. Lian, W. Liu, X. Xu; *Comparison of four methods to generate immunoreactive fragments of a murine monoclonal antibody OC859 against human ovarian epithelial cancer antigen*; Chin Med Sci J; 10(2); 78-81; 1995
(105) T. J. Smith; *Purification of Mouse Antibodies and Fab Fragments*; Methods in Cell Biology; 37(4); 1993
(106) P. Todd, S. R. Rudge, D. P. Petrides, R. G. Harrison; *Bioseparations Science and Engineering (Oxford University Press)*; 2003

(107) P. Parham, M. J. Androlewicz, F. M. Brodsky, N. J. Holmes, J. P. Ways; *Monoclonal antibodies: purification, fragmentation and application to structural and functional studies of class I MHC antigens*; J Immunol Methods; 53(2); 133-73; 1982
(108) J. Rousseaux, R. Rousseaux-Prevost, H. Bazin; *Optimal conditions for the preparation of Fab and F(ab')2 fragments from monoclonal IgG of different rat IgG subclasses*; J Immunol Methods; 64(1-2); 141-6; 1983
(109) A. Coulter, R. Harris; *Simplified preparation of rabbit Fab fragments*; J Immunol Methods; 59(2); 199-203; 1983
(110) J. W. Goding; *Use of staphylococcal protein A as an immunological reagent*; J Immunol Methods; 20(241-53; 1978
(111) K. S. Kan, V. A. Anderson, W. S. Leong, A. M. Smith, A. T. Worth, G. T. Stevenson; *Thioether-bonded constructs of Fab'gamma and Fc gamma modules utilizing differential reduction of interchain disulfide bonds*; J Immunol; 166(2); 1320-6; 2001
(112) E. Ishikawa, S. Yoshitake; *A general and mild procedure for the purification of rabbit Fab' antibodies*; J Immunol Methods; 38(1-2); 117-23; 1980
(113) D. P. Humphreys, O. M. Vetterlein, A. P. Chapman, D. J. King, P. Antoniw, A. J. Suitters, D. G. Reeks, T. A. Parton, L. M. King, B. J. Smith, V. Lang, P. E. Stephens; *F(ab')2 molecules made from Escherichia coli produced Fab' with hinge sequences conferring increased serum survival in an animal model*; J Immunol Methods; 217(1-2); 1-10; 1998
(114) H. Ditzel, K. Erb, G. Leslie, J. C. Jensenius; *Preparation of antigen-binding monomeric and half-monomeric fragments from human monoclonal IgM antibodies against colorectal cancer-associated antigens*; Hum Antibodies Hybridomas; 4(2); 86-93; 1993
(115) W. Zhang, L. A. Marzilli, J. C. Rouse, M. J. Czupryn; *Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody*; Anal Biochem; 311(1); 1-9; 2002
(116) C. T. Esmon, L. DeBault, R. C. Carroll, P. C. Comp, N. L. Esmon; *Biochemical and physiological aspects of protein c*; Ric Clin Lab; 14(3); 455-68; 1984
(117) C. M. Jackson, Y. Nemerson; *Blood coagulation*; Annu Rev Biochem; 49(765-811; 1980
(118) K. G. Mann; *Membrane-bound enzyme complexes in blood coagulation*; Prog Hemost Thromb; 7(1-23; 1984
(119) M. R. Eftink; *Fluorescence methods for studying equilibrium macromolecule-ligand interactions*; Methods Enzymol; 278(221-57; 1997
(120) P. Rice, I. Longden, A. Bleasby; *EMBOSS: the European Molecular Biology Open Software Suite*; Trends Genet; 16(6); 276-7; 2000
(121) D. Turk; *Weiterentwicklung eines Programms fur Molekulgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklarungen*; Ph.D. Thesis, Technische Universitat, Munchen; 1992
(122) T. Bergfors; *Seeds to crystals*; J Struct Biol; 142(1); 66-76; 2003
(123) C. Thaller, L. H. Weaver, G. Eichele, E. Wilson, R. Karlsson, J. N. Jansonius; *Repeated seeding technique for growing large single crystals of proteins*; J Mol Biol; 147(3); 465-9; 1981

(124) E. F. Garman, R. L. Owen; *Cryocooling and radiation damage in macromolecular crystallography*; Acta Crystallogr D Biol Crystallogr; 62(Pt 1); 32-47; 2006
(125) E. Garman; *Cool data: quantity AND quality*; Acta Crystallogr D Biol Crystallogr; 55(Pt 10); 1641-53; 1999
(126) A. McPherson; *A comparison of salts for the crystallization of macromolecules*; Protein Sci; 10(2); 418-22; 2001
(127) T. Holyoak, T. D. Fenn, M. A. Wilson, A. G. Moulin, D. Ringe, G. A. Petsko; *Malonate: a versatile cryoprotectant and stabilizing solution for salt-grown macromolecular crystals*; Acta Crystallogr D Biol Crystallogr; 59(Pt 12); 2356-8; 2003
(128) E. Mitchell, P. Kuhn, E. Garman; *Demystifying the synchrotron trip: a first time user's guide*; Structure; 7(5); R111-21; 1999
(129)
(130) S. Covaceuszach, A. Cattaneo, D. Lamba; *Neutralization of NGF-TrkA receptor interaction by the novel antagonistic anti-TrkA monoclonal antibody MNAC13: a structural insight*; Proteins; 58(3); 717-27; 2005
(131) J. Navaza; *Implementation of molecular replacement in AMoRe*; Acta Crystallogr D Biol Crystallogr; 57(Pt 10); 1367-72; 2001
(132) S. Terzyan, P. A. Ramsland, E. W. Voss, Jr., J. N. Herron, A. B. Edmundson; *Three-dimensional structures of idiotypically related Fabs with intermediate and high affinity for fluorescein*; J Mol Biol; 339(5); 1141-51; 2004
(133) A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren; *Crystallography & NMR system: A new software suite for macromolecular structure determination*; Acta Crystallogr D Biol Crystallogr; 54(Pt 5); 905-21; 1998
(134) T. Bhat; *Calculation of an OMIT Map*; Journal of applied crystallography; 21(279-281; 1988
(135) A. T. Brunger; *Free R-value - a novel statistical quantity for assessing the accuracy of crystal-structures*; Nature; 355(6359); 472-475; 1992
(136) A. L. Morris, M. W. MacArthur, E. G. Hutchinson, J. M. Thornton; *Stereochemical quality of protein structure coordinates*; Proteins; 12(4); 345-64; 1992
(137) R. W. Hooft, G. Vriend, C. Sander, E. E. Abola; *Errors in protein structures*; Nature; 381(6580); 272; 1996
(138) B. Rupp; *Personal communication*; EMBO Practical Course on Protein Expression, Purification and Crystallization (PEPC5; EMBL Hamburg outstation); 2006
(139) R. Cudney; *Personal communication*; Hampton Research Technical Services; 2006
(140) G. N. Ramachandran, A. K. Mitra; *An explanation for the rare occurrence of cis peptide units in proteins and polypeptides*; J Mol Biol; 107(1); 85-92; 1976
(141) S. Lorenzen, B. Peters, A. Goede, R. Preissner, C. Frommel; *Conservation of cis prolyl bonds in proteins during evolution*; Proteins; 58(3); 589-95; 2005

(142) C. Branden, J. Tooze; *Introduction to Protein Structure (Second Edition)*; Garland Publishing; 1998
(143) J. P. K. Doye, W. C. K. Poon; *Protein Crystallization in vivo*; Current Opinion in Colloid and Interface Science; 11(40-46; 2006
(144) B. W. Matthews; *Solvent content of protein crystals*; J Mol Biol; 33(2); 491-7; 1968
(145) K. A. Kantardjieff, B. Rupp; *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals*; Protein Sci; 12(9); 1865-71; 2003
(146) P. F. Cook, W. W. Cleland; *Enzyme Kinetics and Mechanism*; Routledge; 2007
(147) P. D. Martin, W. Robertson, D. Turk, R. Huber, W. Bode, B. F. Edwards; *The structure of residues 7-16 of the A alpha-chain of human fibrinogen bound to bovine thrombin at 2.3-A resolution*; J Biol Chem; 267(11); 7911-20; 1992
(148) C. Eigenbrot, D. Kirchhofer, M. S. Dennis, L. Santell, R. A. Lazarus, J. Stamos, M. H. Ultsch; *The factor VII zymogen structure reveals reregistration of beta strands during activation*; Structure (Camb); 9(7); 627-36; 2001
(149) J. Vijayalakshmi, K. P. Padmanabhan, K. G. Mann, A. Tulinsky; *The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin*; Protein Sci; 3(12); 2254-71; 1994
(150) S. Tonegawa; *Somatic generation of immune diversity*; Biosci Rep; 8(1); 3-26; 1988
(151) F. Altmann; *The role of protein glycosylation in allergy*; Int Arch Allergy Immunol; 142(2); 99-115; 2007
(152) D. B. Cines, L. Rauova, G. Arepally, M. P. Reilly, S. E. McKenzie, B. S. Sachais, M. Poncz; *Heparin-induced thrombocytopenia: an autoimmune disorder regulated through dynamic autoantigen assembly/disassembly*; J Clin Apher; 22(1); 31-6; 2007

APPENDICES

X.1. Characterization of macroseeding-assisted crystal growth

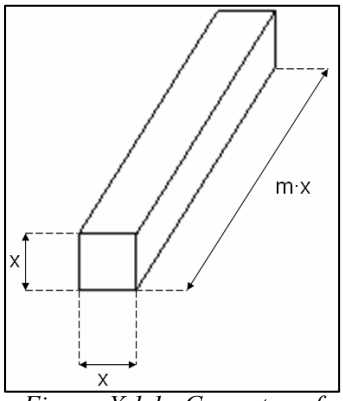


Figure X.1.1: Geometry of the ncFab-Ca^{2+}-peptide crystals.

The aim of this appendix is to find constants able to characterize the growth of the crystals. Experimental crystal volumes measured over time are used to validate a mathematical growth model. The experimental dataset that is used was obtained by averaging the volume of 6 crystals measured daily during macroseeding (see fig. 28).

The crystals were parallelepipeds with a cuboid geometry (see fig. X.1.1.). Their dimensions were measured using the microscope, and their volume was logged. It soon became apparent that the ratio *m* between their longest dimension and their shortest dimension did not vary over time, with an average value of 13.77. The length of the small square sides, x, can thus be used as a characteristic dimension of the crystals. Also at any time *t*, the volume *V(t)* of a crystal is obtained by:

[X.1.1.] $$V(t) = m \cdot [x(t)]^3$$

Similarly, the area *A(t)* available for growth is easily obtained from the geometry as

[X.1.2.] $$A(t) = 2 \cdot x(t)^2 + 4 \cdot m \cdot x(t)^2 = (2 + 4m) \cdot x(t)^2$$

Crystals grow by addition of protein copies on to the surface of a nucleus. It is why many crystal growth models assume that the growth itself is directly proportional to the area of the crystal[106]. In this line of thinking, we will model the evolution of the crystal volume as follows:

[X.1.3.] $$\frac{dV}{dt} = K \cdot A(t)$$

where K (units: μm·day^{-1}) is a "macroseeding growth constant" we seek to determine. This model is a simplification of more complex equations which take into account the protein supersaturation in the growth medium. In our case, the protein concentration in the drops will decrease upon addition of the molecules onto the crystal. However it is regularly brought back to its initial value by the macroseeding procedure. Indeed, macroseeding consists of repeatedly placing the crystals in a fresh drop of known composition. It will thus be assumed that the protein concentration (and thus the supersaturation) is constant over several macroseeding cycles. In our expression, the constant K encompasses supersaturation, diffusion of the protein molecules to the surfaces of the nuclei and kinetics of adsorption as well as possibly other undetermined factors.

To derive an expression for *V(t)*, one needs to replace both terms of the previous equation so that there is only one value varying over time. The characteristic length *x(t)* seems most appropriate:

[X.1.4.] $$\frac{d}{dt}\left[m \cdot x(t)^3\right] = K \cdot \left[(2+4m) \cdot x(t)^2\right]$$

Rearranging this expression yields the following differential equation:

[X.1.5.] $$\frac{dx^3}{dt} = \frac{K \cdot (2 + 4m)}{m} x(t)^2$$

Solving this equation provides us with an expression for the characteristic length as a function of time:

[X.1.6.] $$x(t) = \frac{K \cdot (2 + 4m)}{3 \cdot m} t + cst$$

The integration constant *cst* can be equaled to zero by considering the initial condition of a size of zero at the beginning of the experiment. The previous equation then becomes:

[X.1.7.] $$x(t) = \frac{K \cdot (2 + 4m)}{3 \cdot m} t$$

The experimental data was an average volume, so we can use [X.1.1.] to derive our growth model:

[X.1.8.] $$V(t) = m \cdot [x(t)]^3 = m \cdot \left[\frac{K \cdot (2 + 4m)}{3 \cdot m} t \right]^3$$

After simplification, the final expression for the time dependence of macroseeded crystal volume is obtained:

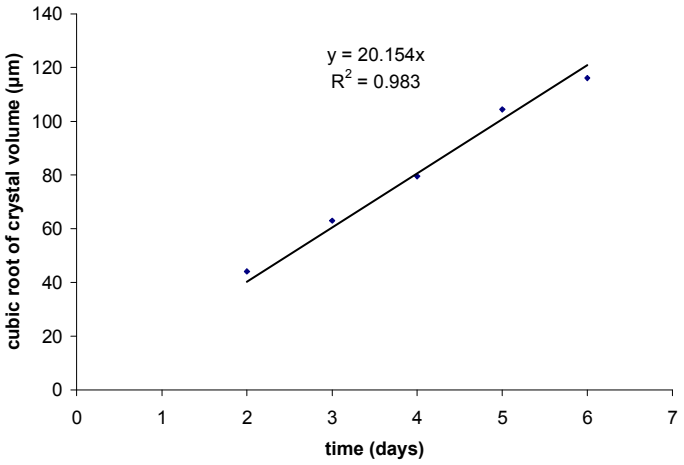
[X.1.9.] $$V(t) = K^3 \cdot \left[\frac{(2 + 4m)^3}{27 \cdot m^2} \right] \cdot t^3$$

Interestingly, the volume shows a cubic dependency on time. Because the numerical value for *m* was determined experimentally (*m*=13.77, bracketed term=36.33), there is only one parameter (K^3) to this model. A simple way to determine *K* is to notice

that the model can be linearized by applying a cubic root operation on the expression for growth:

[X.1.10.] $$\sqrt[3]{V(t)} = K \cdot \left[\frac{(2+4m)}{3 \cdot m^{2/3}}\right] \cdot t$$

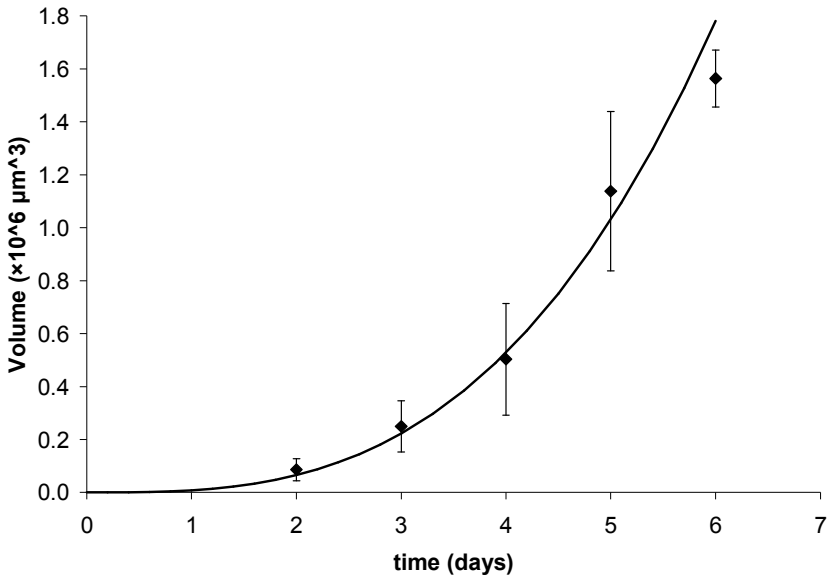
The resulting relationship shows that the cubic root of the volume is in a direct linear relationship with time, where the slope is the product of the bracketed term and of the constant of interest, *K*. The value of the slope is found using a linear regression:



As seen on the graph, this model provides a remarkable correlation to the data (R^2=0.983) with a slope of 20.154. The numerical value for K is found to be:

$$K = slope \Big/ \left[\frac{(2+4m)}{3 \cdot m^{2/3}}\right] = \frac{20.154}{\frac{2+4\times 13.77}{3\times 13.77^{2/3}}} = 6.1 \mu m \cdot day^{-1}$$

The macroseeding growth rate is thus found to be ***K*=6.1µm·day^{-1}**. The following graph shows the quality of the fit in terms of volume.

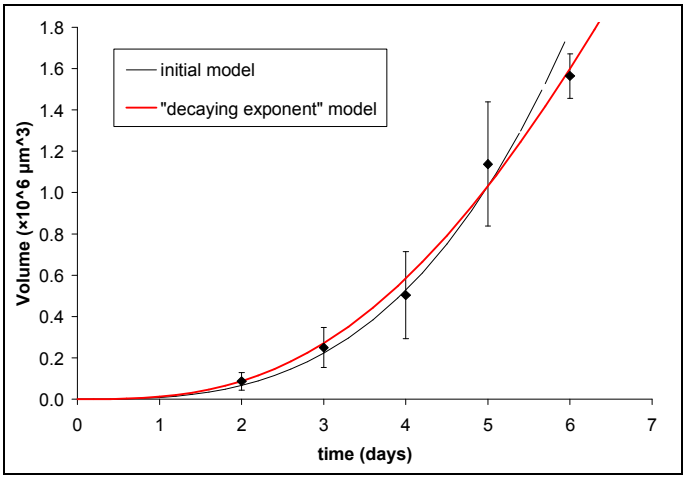


As seen on this graph, the latest volume is smaller than the one predicted by the model. In practice, it was often seen that sprouting would slow down the growth of the seeds. One could then modify the model by introducing a growth constant that decays over time. However, most of the monotonously decaying functions by which one could multiply the constant K would ultimately offset the cubic growth of the crystal, resulting in a size decrease. This is not what was seen in reality, where crystals would slowly stop growing but never decreased after the maximal size was reached. A different term that can be modified in the model to account for the decrease in growth rate would be the cubic exponent itself. Multiplying the exponent by a function that converges towards 0 would result in a smooth transition from cubic growth ($V \div cst \cdot t^3$) to an arrested growth phase ($V \div cst \cdot t^0 = cst$). In physical terms, the decay of the exponent could tentatively be explained as the gradual disappearing of the number of dimensions available for growth due to sprouting. Exponential functions are classically seen in many physical phenomena where it is often used to

model decays. In our case, we propose the following, "decaying exponent", modified model:

[X.1.11] $$V(t) = K^3 \cdot \left[\frac{(2+4m)^3}{27 \cdot m^2} \right] \cdot t^{3 \cdot \exp(-t/\tau)}$$

The additional constant τ would be a "decay time" (units: days) to determine. In the early time points, the exponent term would approximate $3 \cdot e^0 = 3$, and the growth would be cubic with respect to time, just as in the initial model. Then at later time points, the exponential terms would start converging towards *0*, resulting in growth arrest. Using the same dataset and a nonlinear fitting procedure, K and τ values of 6.8μm·day^{-1} and 70 days are obtained, respectively. The average residuals of each model compared to the dataset can be used to compare their goodness of fit. The regular model has an average residual value of $79 \times 10^3 \mu m^3$, whereas the decaying exponent model has an average residual value of $49 \times 10^3 \mu m^3$. The introduction of the decay time, τ into the derived model allows to substantially improve the fit, mostly by better accounting for the slowing down of growth that occurs at our last time point. The two models are plotted on the following graph.



X.2. Cryoprotection

At first, the cryoprotectant concentration C in the drop of volume V is zero. A small volume v is then removed, and replaced by the same volume of cryoprotecting solution of concentration C_f, so that the volume is maintained at V. Thus, at the end of the first step (step number n=1), the concentration in the drop is

[X.2.1.] $$C_{n=1} = \frac{v \cdot C_f}{V}$$

After the second step, a volume v of the drop (presently at concentration C_1) will be replaced by a volume v at the concentration C_f. Thus, the concentration at the end of the second step, C_2, will be:

[X.2.2.] $$C_2 = \frac{(V - v) \cdot C_1 + v \cdot C_f}{V}$$

To find the iterative formula, one can replace C_1 in the precedent equation by its literal expression, yielding:

[X.2.3.] $$C_2 = \frac{(V - v) \cdot \frac{v \cdot C_f}{V} + v \cdot C_f}{V}$$

Factoring (X.1.3.) with the expression for C_1 yields:

[X.2.4.] $$C_2 = \frac{v \cdot C_f}{V}\left[1 + \left(\frac{V - v}{V}\right)\right]$$

Pursuing this for greater values of n yields:

[X.2.5.] $$C_n = \frac{v \cdot C_f}{V}\left[1 + \left(\frac{V - v}{V}\right) + \left(\frac{V - v}{V}\right)^2 + \left(\frac{V - v}{V}\right)^3 + \cdots + \left(\frac{V - v}{V}\right)^{n-1}\right]$$

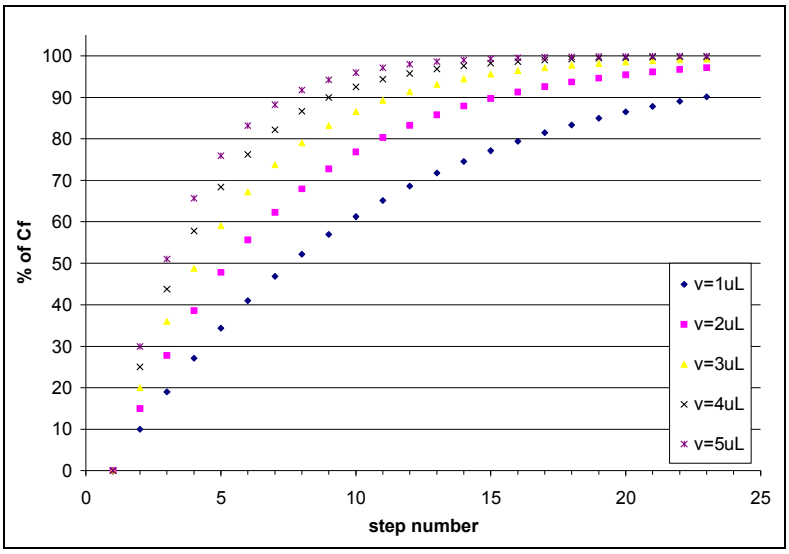
[X.2.5.] can then be rewritten in a more compact form:

[X.2.6.] $$C_n = \frac{v \cdot C_f}{V} \cdot \sum_{i=0}^{n-1} \left(\frac{V - v}{V} \right)^i$$

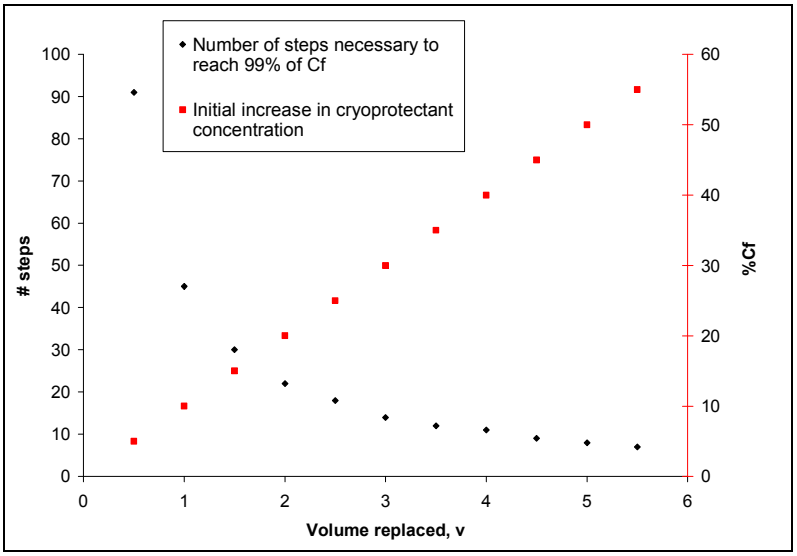
This expression is generally identified as a Taylor series in $(V-v)/V$, and since by construction $-1<(V-v)/V<1$, it also satisfies the definition for a MacLaurin series. This formalism is interesting to examine the behavior of C_n at high values of n, where the following MacLaurin identity is valid:

[X.2.7.] $$\sum_{n=0}^{\infty} \left(\frac{V - v}{V} \right)^n = \frac{1}{1 - \left(\frac{V - v}{V} \right)} = \frac{V}{v}$$

So for a great enough number of solution replacements, equation [X.2.6.] will collapse to the equality $C_n=C_f$, clearly demonstrating the convergence of the [X.2.6.] series. In other words, by repeating the exchange step for enough number of times, it is possible to asymptotically reach the final concentration C_f. This could be expected with no derivation but the precise number of steps required to arrive at a value of C_n that is significantly close to C_f can only be determined using [X.2.6.]. Even though it is not easily amenable to analytical solution (because of the sum form), relevant values can be found by plotting [X.2.6.] for various n and v values that are technically feasible (C_f and V being otherwise dictated by the required cryoprotecting concentration and the evaporation rate, respectively). The following plot was realized for 5 different values of v:



Using the previous plot, the following values are obtained:



To conclude, we chose a replacement volume of 2µL because it allowed to reach 99% of the cryoprotecting concentration in 22 steps (yielding a comfortable 27s in between volume exchanges considering the 10min evaporation limit) while setting the highest osmotic change undergone by the crystal to only 20% of the cryoprotecting concentration.