ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Insights on NGF-TrkA interaction: structural characterization of two neutralizing antibodies

Thesis submitted for the degree of
" Doctor Philosophiæ "

CANDIDATE
Sonia Covaceuszach

SUPERVISORS
Prof. Antonino Cattaneo
Dott. Doriano Lamba

April 2002

TRIESTE

ISSA - SCUOLA
TERNAZIONALE
SUPERIORE
STUDI AVANZATI
TRIESTE
Via Beirut 2-4

ISAS–INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Insights on NGF-TrkA interaction: structural characterization of two neutralizing antibodies

Thesis submitted for the degree of
" Doctor Philosophiæ "

CANDIDATE
Sonia Covaceuszach

SUPERVISORS
Prof. Antonino Cattaneo
Dott. Doriano Lamba

April 2002

TRIESTE

To Paolo

You take the blue pill and the story ends. You wake in your bed and you believe whatever you want to believe.
You take the red pill and you stay in Wonderland and I show you how deep the rabbit-hole goes. Remember that all I am offering is the truth. Nothing more.

MORPHEUS (The Matrix)

INDEX

- ***CHAPTER 1: INTRODUCTION*** *pag 1*
- Physiological relevance of NGF/TrkA interaction *pag 1*

KO mice analysis *pag 3*

Physiological role of NGF/TrkA interaction *pag 4*

- *In early embryogenesis and target innervation* *pag 4*
- *In hyperalgesia* *pag 4*
- *In neuronal plasticity* *pag 6*

- Implications in diseases states *pag 7*
  - *In neurodegenerative diseases and nervous system injuries* *pag 7*
  - *In certain chronic inflammatory or neuropathic pain states* *pag 10*
  - *In several human malignancies* *pag 10*
- Structure- function studies on NGF/TrkA Interaction *pag 11*
- Structural studies on NGF and TrkA *pag 15*
  - *NGF structure* *pag 15*
  - *NGF-TrkAd5 structure* *pag 16*
  - *NGF-TrkAd5 structure* *pag 20*
- Blocking antibodies: an additional way to study NGF/TrkA interaction *pag 22*
  - *αD11 antibody* *pag 23*
  - *MNAC13 antibody* *pag 24*
- Structural studies on antibody molecules *pag 25*

- About this thesis *pag 28*
- References *pag 30*

PART 1: BIOCHEMICAL CHARACTERIZATION and in vitro FUNCTIONAL STUDIES oN MNAC13 ANTIBODY

CHAPTER 2: Epitope mapping by serial deletions *pag 44*

- Introduction *pag 44*
- Materials and methods *pag 47*

Cloning of TrkA subdomains in pMAL2C *pag 47*

Cloning of TrkA ig1 like subdomain deletion mutants in pMAL *pag 48*

Protein expression and extraction. *pag 50*

Protein extraction from bacterial cells *pag 50*

Protein extraction from bacterial periplasm *pag 50*

Refolding procedure *pag 51*

ELISA assay *pag 51*

- Results *pag 53*
- Discussion *pag 60*
- **References** *pag 62*

CHAPTER 3: Epitope mapping by phage display *pag 63*

- Introduction *pag 63*

- Materials and methods *pag 68*

Phagmidic library *pag 68*

Solution panning *pag 68*

Phage storage and amplification *pag 69*

Phage ELISA assay *pag 70*

Phage immunoscreening *pag 70*

Sequencing *pag 71*

- Results *pag 72*
- Discussion *pag 76*
- References *pag 80*

PART 2: STRUCTURAL CHARACTERIZATION OF TWO NEUTRALIZING ANTIBODIES

CHAPTER 4: scFv expression, purification and characterization *pag 83*

- Introduction *pag 83*
- Materials and methods *pag 85*

Insect cell culture *pag 85*

Cloning of αD11 scFv in pAcGp67B *pag 85*

Changing αD11 scFv linker *pag 86*

Transfection and selection of αD11 scFv recombinant Baculovirus *pag 89*

Amplification of αD11 scFv recombinant baculovirus *pag 89*

Protein expression by baculovirus system *pag 90*

Purification of αD11 scFv by metal ion affinity Chromatography *pag 90*

Western blot *pag 91*

ELISA assay and Competitive ELISA assay *pag 91*

- Results *pag 93*

αD11 scFv expression by baculovirus system and purification *pag 93*

αD11 scFv characterization by size exclusion chromatography *pag 96*

αD11 scFv linker clipping by subtilisin *pag 98*

αD11 scFv linker change *pag 102*

- Discussion *pag 105*
- References *pag 108*

CHAPTER 5: Fab Purification

- Biochemical methods and results *pag 111*

Purification of Mabs *pag 111*

Papain cleavage *pag 113*

Purification of the FabMNAC13 *pag 114*

Purification of the FabαD11 *pag 115*

- Discussion *pag 117*

CHAPTER 6: Fab Crystallization *pag 120*

- Methods and results *pag 125*

  FabMNAC13 crystallization *pag 125*

  Fab αD11 crystallization *pag 127*

- References *pag 131*

CHAPTER 7: Data collection and processing *pag 133*

- Data collection *pag 136*

  Cryoprotection and crystal mounting *pag 136*

  Preliminary analysis and data collection *pag 136*

  - *FabMNAC13 data collection* *pag 137*
  - *Fab αD11 data collection* *pag 138*

- Data processing *pag 141*
  - *FabMNAC13 data processing* *pag 141*
  - *Fab αD11 data processing* *pag 143*

- References *pag 145*

CHAPTER 8: Phase solution *pag 146*

- Molecular replacement *pag 149*
- Phasing by molecular replacement *pag 152*
  - *FabMNAC13 phasing* *pag 154*
  - *Fab αD11 data phasing* *pag 157*

- References *pag 163*

CHAPTER 9: Fitting, Refinement and Validation *pag 165*

- FabMNAC13 *pag 168*

  Fitting and Refinement *pag 168*

  Validation *pag 170*

- FabαD11 *pag 174*

  Fitting and Refinement *pag 174*

  - *Crystal form II preliminary fitting and refinement* *pag 174*
  - *Crystal form III fitting and refinement* *pag 176*

  Crystal form III validation *pag 177*

- References *pag 182*

CHAPTER 10: FabMNAC13 structure *pag 184*

- Overall structure *pag 184*
- Crystal packing *pag 185*
- Antigen binding site *pag 186*
- Discussion *pag 192*
- References *pag 200*

CHAPTER 11: FabαD11 structure *pag 202*

- Overall structure *pag 202*
- Crystal packing *pag 203*
- Antigen binding site *pag 208*
- Discussion *pag 212*
- References *pag 216*

CHAPTER 12: Conclusions and perspectives *pag 217*

- Insight on NGF-TrkA interaction *pag 218*
- Antibody's humanization *pag 221*
- Design of CDRs peptides *pag 223*
- Rational drug design and development of peptide mimics *pag 225*
- References *pag 226*

CHAPTER 1: Introduction

NGF (Nerve growth factor) is a prototypical member of the neurotrophin family of growth factors that includes brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), neurotrophin-4/5 (NT-4/5) and neurotrophin-6 (NT-6) and promotes the differentiation and survival of distinct populations of neurons in both the central and the peripheral nervous systems (Barde, 1994). Two types of cell-surface receptors are engaged by the neurotrophins for signalling.

A first class, composed by to the Trk receptor tyrosine kinase family, binds selectively the neurotrophins with affinity in the picomolar range. Different neurotrophins interact selectively with their cognate Trk receptor and display specific sets of biological activities on peripheral and central nervous system neurons. Both functions are required both for development and maintenance of neuronal populations.

In contrast to the Trk receptors, p75NTR (p75 neurotrophin receptor), the unique member of the second class of receptors, binds to all neurotrophins with similar, but lower affinity in the nanomolar range. p75NTR is a distantly related member of the tumor necrosis factor receptor family. When it binds NGF, in the absence of Trk receptors, it promotes apoptosis during normal development. In the presence of Trk instead, it cooperates to increase the affinity of neurotrophin binding and/or signalling efficiency even at low concentrations of the ligand (Kaplan & Miller, 1997).

Physiological relevance of NGF/TrkA interaction

NGF elicits most of its biological effects by binding with high affinity a specific member of tyrosine kinase receptor family, named TrkA. Docking of TrkA with NGF initiates receptor dimerization, catalytic

autophosphorylation of cytoplasmic tyrosine residues on the receptor, and a signal transduction cascade shown in Figure 1 (Kaplan and Stephens, 1994).

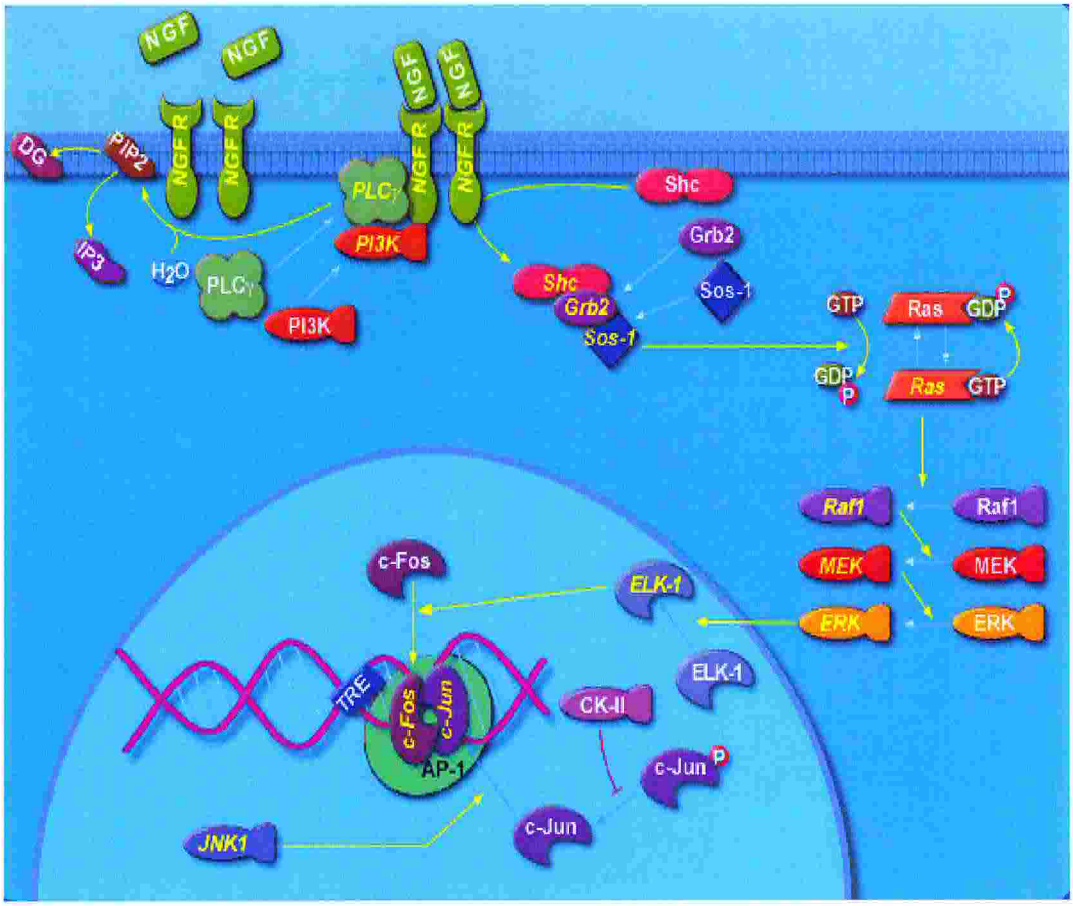


***Figure 1**: Trans-membrane signaling activated by NGF/TrkA interaction*

These signals lead to prevention of apoptotic cell death (Maliartchouk and Saragovi, 1997), promotion of cellular differentiation and axon elongation, and up-regulation of choline acetyl transferase (ChAT) (Hefti *et al.*, 1985) on specific neuronal population of peripheral and central nervous system that express TrkA receptor.

In particular basal forebrain cholinergic neurons are the principal target of NGF action in the central nervous system. These neurons provide most of the cortical and hippocampal innervation (Dutar *et al.*, 1995) and exhibit selective uptake and retrograde transport of NGF (Sieler and Schwab, 1984; Domenici *et al.*, 1994). Concerning the peripheral nervous system the responsive neuronal populations are mainly sympathetic neurons, small size nociceptive spinal sensory neurons and the trigeminal ganglion.

KO mice analysis

The generation of mice carrying germ line mutations in the gene encoding NGF (Crowley *et al.,* 1994) has paved the way to the unique opportunity to explore its role in the development and maintenance of the mammalian nervous system. Similar studies (Smeyne *et al.,* 1994) involving knockout mice carrying mutations in the catalytic domain of TrkA receptor have also provided critical information regarding its role *in vivo*. Moreover striking phenotypic similarities between mice defective either for the receptor or for the ligand represent the most compelling evidence that TrkA receptor mediate most, if not all, of the biological activities of NGF. Indeed mice defective either for NGF or TrkA receptors display severe sensory defects characterised by a complete loss of nociceptive activity. They fail to react to deep pinpricks in their whisker pad and rear pad and exhibit deficiencies in thermoception as they can stay on top of a 60°C hot plate for at least 10s. Neuro-anatomical examination of the NGF- and TrkA-null mice revealed extensive neuronal loss in trigeminal, dorsal root and sympathetic ganglia. In particular in the dorsal root ganglia (DRG) the vast majority of missing neurons corresponds to those of small size, a population known to be NGF-dependent. The sympathetic ganglia are severely shrunken and

contain only a few neurons. From all the data on gene-targeted mice it is worth noting that neurotrophins play a differential role in the peripheral versus the central nervous system. In the peripheral nervous system ablation of neurotrophin or Trk receptor genes results in massive neuronal cell death, whereas the central nervous system neurons of these mutant mice appear to be, for the most part, unaffected in spite of widespread neurotrophin and receptor expression. These observations underline the complex mechanisms involved in the development and survival of the nervous system.

Physiological role of NGF/TrkA interaction

- ***In early embryogenesis and target innervation***

Although the interaction between NGF and TrkA has been very much characterized for its trophic role in preventing programmed cell death during normal development in the peripheral nervous system and in sympathetic neurons, in adults it has also been involved in the regulation of many other specific neuronal functions upon activation of TrkA receptors (Figure 2).

Indeed it is worth noting that the key experiment in the growth factor research field was the demonstration that treatment of newborn rodents with anti-NGF antibodies (as well as inactivation of TrkA or NGF genes) caused the selective elimination of the sympathetic chain (Levi-Montalcini and Booker, 1960; Snider, 1994). Following these findings, a model of target control of neuronal survival and function has been proposed and may be also extended to other neuronal populations and molecules.

- ***In hyperalgesia***

Concerning the role of NGF after development, it was demonstrated to mediate hypersensibilization to noxious heat and mechanical stimuli after

inflammation and lesion leading pain and hyperalgesia (Lewin and Mendell, 1993a).

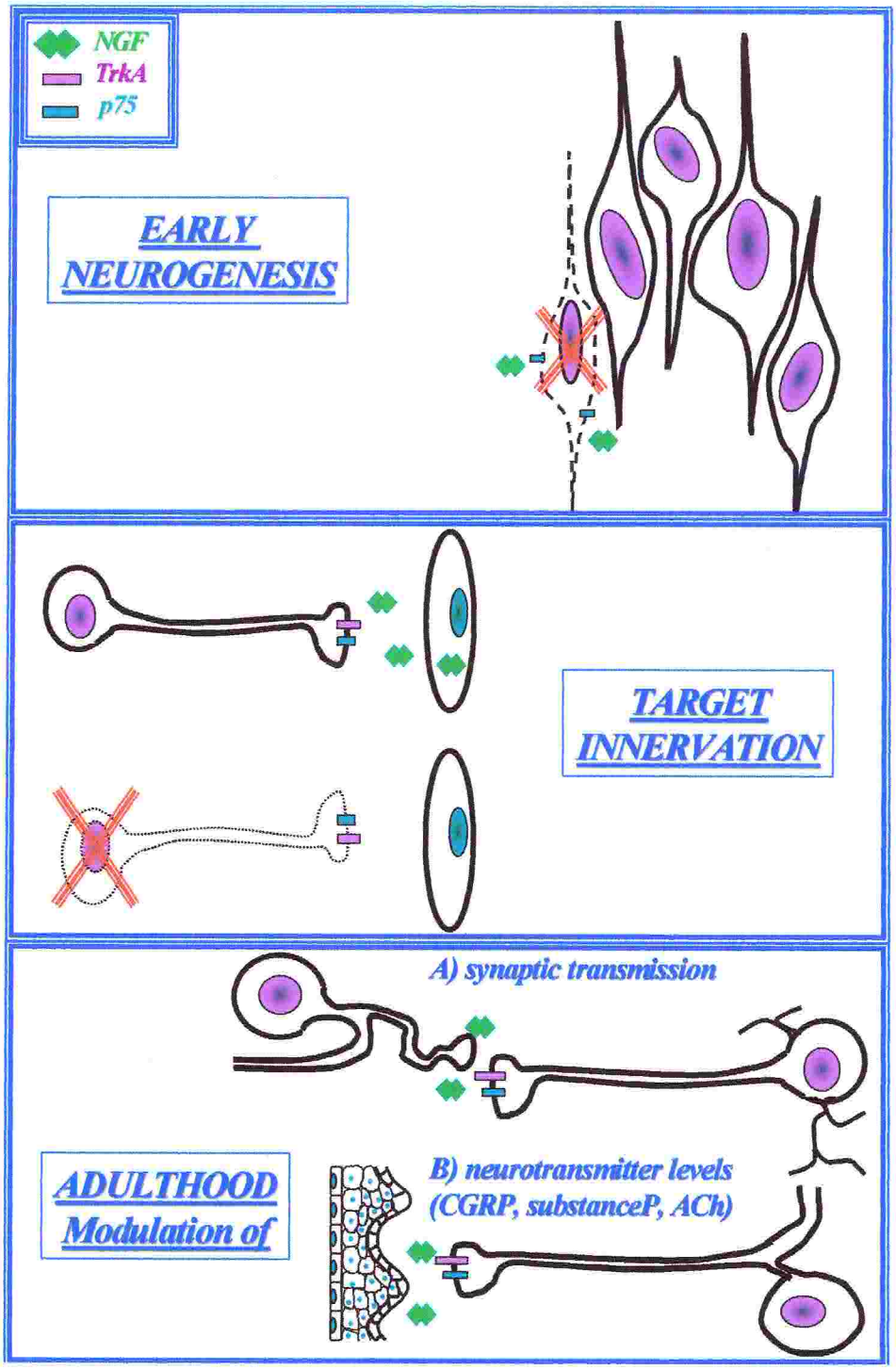


***Figure 2*: NGF physiological functions**

Besides experiments in which hyperalgesia could be induced by NGF (Lewin and Mendell, 1993b) or blocked by specific antibodies injection, it was also possible to detect a marked increase in NGF levels in damage or inflamed tissue (Woolf *et al.*, 1994). Moreover this result could be obtained both *in vitro* (Lindholm *et al.*, 1987) and *in vivo* (Safieh-Garabedia *et al.*, 1995) through cytokines typically involved in tissue damage and inflammatory processes such as interleukin-1β. The basis of this hyperalgesic effect includes the up-regulation of peptide neurotransmitters associated with the detection of potentially painful stimuli (nociceptors), such as substance P and calcitonin gene-related peptide (CGRP) (Donnerer *et al.*, 1992), contained in a subset of NGF-responsive DRG neurons (Verge *et al.*, 1992). Moreover NGF is able to trigger mast cells degranulation, causing the release of histamine and serotonine that sensitize nociceptors (McMahon, 1996).

- ***In neuronal plasticity***

Moreover NGF has also been involved in neuronal plasticity, which relates to changes in neuronal structure and function, caused by environmental stimuli, and therefore is fundamental for postmitotic neurons, since cell division ceases early in the developing nervous system. Growing evidences link morphological changes and electrical activity with neurotrophins function in the CNS (Thoenen, 1995). This hypothesis is based on the early demonstration that the stimulation of a particular pathway in the hyppocampus led to an increased expression of the NGF gene (Gall *et al.*, 1989). Regarding activity-dependent neuronal plasticity, NGF levels have been shown to affect neuronal architecture of sympathetic ganglia, whereas NGF antibodies caused loss of synaptic transmission and contacts (Niå *et al.*, 1978). This was particularly interesting because the primary sites of NGF response are the sympathetic neurons and not the pre-synaptic terminals.

In general the highest transcriptional expression of NGF in the CNS is localized at the level of hyppocampus and cerebral cortex. However excitatory inputs are able to induce, both *in vivo* and *in vitro,* not only a further transcriptional activation (Zafra *et al.*, 1990) but also secretion of NGF along neuronal processes and dendrites (Blöchl and Thoenen, 1996). As a consequence of that it is very likely that NGF causes the rapid release of neurotransmitters not only *in vitro* (Knipper *et al.*, 1993; Blöchl and Sirrenberg, 1996) but also *in vivo*, considering that its injection alters dramatically hyppocampal and cortical electrical activity (Berzaghi *et al.*, 1995). Concerning activity-dependent biosynthesis and release of NGF, it has been demonstrated that the result could be not only a fast action, but also the modulation of long-lasting developmental changes, especially in the visual system. Considering that, in the visual cortex, neurons can receive from each eye different degrees of innervation, following a relay station in the thalamus, it has been shown not only that ocular dominance is the result of activity dependent sorting of thalamic afferents. Furthermore this phenomenon is influenced by the levels of NGF (Maffei *et al.*, 1992), which acts stabilizing synaptic contacts of thalamic afferents. Thus NGF deprivation would leave the cortical network more prone to activity dependent modifications, while exogenus NGF would prematurely stabilize contacts and prevent plasticity (Berardi *et al.*, 1994; Dominici *et al.*, 1994a; Thoenen, 1995; Bonhoeffer, 1996).

Finally, modulation of NGF level has also been involved in activity independent neuronal plasticity regulating the phenotype of neurons without modification in the number of cells. This effect has been demonstrated in the sensory ganglia of rats after the interval in which NGF affects the number of neurons. Indeed reducing NGF levels causes the replacements of C-mechanoheat fibers with a novel type of low

mechanical threshold pressure receptor (Lewin and Mendell, 1994). In a similar way at te level of the skin Aδ afferents, high threshold mechanoreceptor from the sural nerve, are replaced by D-hair type afferents, that respond to low-threshold pressure stimulation of hairs (Ritter *et al.*, 1991).

In this brief review NGF involvement in programmed cell death during normal development is not discussed mainly because it has been shown to be mediated by the interaction of the p75NTR low affinity receptor, in absence of TrkA (Frade *et al.*, 1996).

Implications in diseases states

It is worth noting that several neuronal cell types that are implicated in important disease states express TrkA and therefore respond to NGF, including sensory, sympathetic, and cholinergic neurons.

- ***In neurodegenerative diseases and nervous system injuries***

NGF is considered to be a potential therapeutic agent in the treatment of many neurodegenerative diseases and nervous system injuries (Cuello *et al.*,1996, Mufson *et al.*,1997) since it promotes growth and survival of neurons during development and after neuronal damage. Thus NGF has been suggested not only for the treatment of neuronal damage (Hughes *et al.*, 1997), but also to ameliorate peripheral diabetic neuropathies (Ebadi *et al.*, 1997).

Moreover it has to be stressed that NGF is a potent neurotrophic factor for forebrain cholinergic neurons and one of the earliest changes in Alzheimer's disease is the loss of function of this specific neuronal population, associated with memory loss. For this reason it has been proposed that NGF therapy may delay the onset of Alzheimer's disease (Barinaga, 1994; Lindsay, 1996). Therefore NGF treatment has been

tested in several animal models and it turned out that its administration is able to correct the effects of cholinergic atrophy both in lesioned (Tuszynski *et al.*, 1990; Junard *et al.*, 1990; Hefti, 1986; Montero & Hefti, 1988; Montero & Hefti, 1989) and in aged animals (Smith *et al.*, 1999; Fischer *et al.*, 1987). These encouraging results in animal models led to similar studies also with patients suffering from Alzheimer's disease (Olson *et al.*, 1992; Jonhangen *et al.*, 1998).

Despite the therapeutic potential of NGF, clinical trials featuring this protein have been disappointing (Verrall, 1994; Saragovi and Burgess, 1999) even if some positive effects could be observed. Indeed the treatment with purified mouse NGF was able to increase cerebral blood flow and to decrease nicotinic cholinergic receptor density (Olson *et al.*, 1992; Jonhangen *et al.*, 1998). As a consequence of that only slight improvements in the results of neuropsychological tests could be obtained, but with no any clear cognitive improvement.

There are several causes that can limit the use of NGF as a therapeutic drug. First of all NGF is not able to cross the blood-brain barrier and display very poor diffusion in brain tissue, probably due to interactions with the extracellular matrix. Thus the main drawback of NGF administration is that it requires expensive, invasive surgery and this greatly limits the number of patients that could be treated.

Moreover besides the fact that NGF protein is relatively expensive to produce for medicinal applications, other possible problems can be due to *in vivo* instability (Barinaga *et al.*, 1994) and pleiotrophic effects, caused by activation of signals that were not intentionally targeted (unwanted activation of TrkA receptor, leading pain and hyperalgesia (Lewin and Mendell, 1993a), and/or of low-affinity NGF receptor p75 (Carter and Lewin, 1997)). Finally there are all the inherent drawbacks associated

with the use of polypeptides as drugs (Saragovi et al., 1992) such as problems of ligand clearance and sequestration.

Thus this crucial point could be solved developing small, proteolytically stable molecules with neurotrophic activity, selective for cells expressing TrkA, whose synthesis could be less expensive than NGF purification and whose protocol of administration could be easier and cheaper, if they were able to cross blood-brain barrier.

- ***In certain chronic inflammatory or neuropathic pain states***

On the other hand the fact that high levels of NGF have been shown to produce hyperalgesia in both rodents and humans (Anderson *et al.*, 1995; Lewin *et al.*, 1993) has led to the proposal that NGF antagonists may be useful as analgesics. Such kind of compounds should therefore be very important in the treatment of certain chronic inflammatory or neuropathic pain states.

- ***In several human malignancies***

NGF and TrkA receptor have been implicated as both oncogenic agents and tumor suppressors in various tumors. Among them there are not only neuroectoderm-derived tumors such as medulloblastomas (Revoltella *et al.,* 1980; Bauer *et al.,* 1992), neuroblastomas, glioblastomas (Oelmann *et al.,*1995) , but also non-neuronal carcinomas (Koizumi *et al.,* 1998) such as melanomas (Marchetti *et al.,* 1996), medullary thyroid carcinomas (Goretzki *et al.,* 1987; McGregor *et al,* 1999), pancreatic carcinoid cell lines (Bold *et al.,* 1995), prostate (Djakiew *et al.,* 1991), breast (Tagliabue *et al.*, 2000) and lung carcinomas.

It is worth noting that, although TrkA mediate neuritogenesis and survival of neurons, its activation induce two very different responses in tumor cells. On one, hand since several of these human malignancies express normal TrkA and are NGF dependent or responsive, this

autocrine or paracrine loop may promote and sustain cell growth and survival.

On the contrary it was shown that TrkA receptor display both antimitotic and apoptosis-inducing activity in neural tumor cells. Indeed both in PC12 pheochromocytoma and neuroblastoma cells (Matsushima & Bogenmann, 1993) TrkA showed and antimitotic activity, probably due to increased levels of the cyclin dependent kinase inhibitor p21, at least in PC12 cells. Moreover TrkA activity, upon NGF stimulation, induces cellular apoptosis in TrkA transfected medulloblastoma cells (Muragaki *et al.*, 1997). Similarly rats injected with C6 glioma cells expressing TrkA survived much longer than rats injected with C6 cells expressing kinase-inactive TrkA (Lachyankar *et al.*, 1997), probably because cells expressing wild-type TrkA were less invasive and had a greater rate of apoptosis. As TrkA has never been reported to be apoptotic in postmitotic neurons, this finding may reflect a fatal consequence of a conflict between antiproliferative Trk signaling and the proliferative state of tumor cells. These data are of considerable interest in the neural tumor field, particularly because high level of TrkA expression correlates with favorable prognosis, at least in neuroblastoma (Nakagawara *et al.*, 1993). From this point of view according to the specific biological system within the tumor research field both agonists and antagonists could be employed in the treatment of several types of cancers.

Structure-function studies on NGF/TrkA interaction

As stressed before structure-based drug design of small molecules, which can act as antagonists or agonists of NGF, relies on precise understanding of the interaction of NGF with its receptor TrkA.

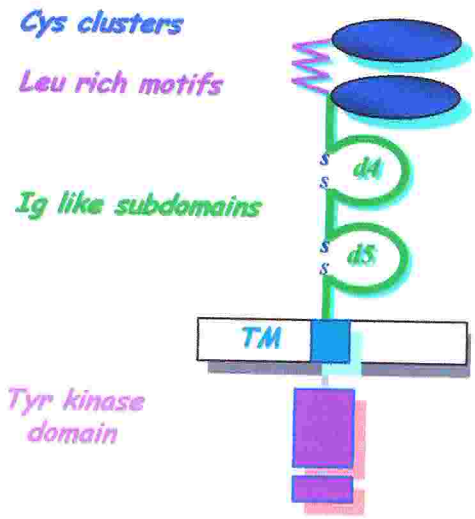


Figure 3: TrkA receptor architecture

This is not so trivial if one considers the complex architecture of the extra cellular portion of TrkA that is composed of five domains (Figure 3). The first three domains consist of a leucine-rich domain (LRM) sandwiched between two cysteine-rich clusters. The fourth and fifth domains (TrkA-d4 and TrkA-d5) are immunoglobulin (Ig)-like domain (Schneider *et al.*, 1991). Several studies have attempted to identify which of these domains is responsible for NGF binding and receptor activation. The results that have been reported are not still conclusive.

Indeed some evidences seemed to define this domain at the level of leucine-rich motif. In particular the second LRM expressed in fusion with maltose binding protein was shown to bind NGF with a similar affinity of the entire extra cellular portion of TrkA and a peptide corresponding to the same LRM inhibited NGF binding to TrkA (Windisch *et al.*, 1995). Moreover the second LRM of TrkB has been shown to bind BDNF, NT3 and NT4/5 but not NGF and naturally occurring TrkB variants lacking two or all three LRMs do not bind BDNF, NT3 or NT4/5 (Ninkina *et al.*, 1997). However further studies on the role of the LRMs in neurotrophin binding showed that their deletion from TrkA does not abolish TrkA autophosphorylation but it changes the differentiative phenotype in neurotrophin responsive cells, resulting in loss of high affinity towards NT3 but not towards NGF (McDonald & Meakin, 1996).

Other evidences have suggested that binding of the neurotrophins to their high affinity Trk receptors is mediated predominantly through interaction with the immunoglobulin-like domains. It was indeed reported (Holden *et al.,* 1997) that the Ig-like domains of TrkA, expressed as recombinant proteins, are able to bind NGF with an affinity similar to that of the native receptor, expressed on the cell surface.

Furthermore NGF was shown to bind cell transiently transfected, expressing a TrkB chimeric protein (obtained replacing both Ig-like domains of TrkB with those of TrkA), with the same affinity as cell expressing wild type TrkA (Perez *et al.,* 1995). Similar results were also obtained replacing both Ig-like domains of TrkC with those of TrkA (Urfer *et al.,* 1995). Moreover it was also shown that expression of truncated TrkA and TrkC receptors, carrying only the second Ig-like domain and tyrosine kinase domain, was sufficient for specific neurotrophin binding and for the autophosphorylation of the truncated TrkC by NT3.

These results strongly support the hypothesis not only that the two immunoglobulin-like domains (domains 4 and 5) of TrkA carry the NGF-binding determinants, but also suggest that TrkA-d5 domain adjacent to the membrane is the dominant element for specific NGF binding. Indeed TrkA-d5 by itself binds to NGF with an affinity similar to that of the native receptor and can antagonize NGF *in vitro* and *in vivo* (D. Dawbarn, unpublished data).

The TrkA-d5/NGF complex (Wiesmann *et al.*, 1999) has been crystallized and its structure elucidated. Furthermore the crystal structures of TrkA-d5 as well as of TrkB-d5 and of TrkC-d5 have been also reported (Ultsch *et al.*, 1999).

It is worth noting that all the structures of the Ig-like domain d5 of the neurotrophin receptors alone form strand-swapped homodimers, in which the amino-terminal A-strand of each molecule is replaced by the equivalent strand of a symmetry-related partner. In this form none of them was able to bind their respective neurotrophin ligands; therefore under the crystallization conditions used all of them were biologically inactive. On the other hand in the co-crystallization experiments TrkA-d5 was present in a monomeric form. Regarding this it is interesting that the starting construct comprised both Ig-like domains, while the resulting crystals contained a complex consisting of a homodimer of NGF and two molecules of TrkA-d5. Therefore a key step in nucleation and/or growth appeared to be proteolysis of the first Ig-like TrkA-d4 during the crystallization process.

Conclusive data (Robertson *et al.*, 2001) were reported to demonstrate that TrkA-d5 alone accounts for all of the contacts with NGF. They showed that this domain binds NGF with an affinity similar to that of the entire extracellular domain, actively sequestering NGF both *in vitro* (preventing NGF induced neurite outgrowth) and *in vivo* (inhibiting NGF induced plasma extravasation). In addition they reported the three dimensional crystal structure of TrkA-d5. Although they obtained crystals of a different space group, they found strand-swapped homodimers as reported above. It is worth noting that unlike previous data, these recombinant proteins were still able to bind NGF with high affinity. Considering that the neurotrophin binding surface is occluded in the strand-swapped homodimer, these results showed that TrkA-d5 is capable of dynamic transitions between inactive dimeric and active monomeric forms. Although no data are available about the propensity to form strand-swapped homodimer in the full-length TrkA receptor where

neighboring domains may limit this mode of oligomerization, such a kind of transitions might suggest a potential mechanism for molecular switching between inactive and active forms of surface receptors.

Structural studies on NGF and TrkA

NGF structure

The first structural information (McDonald *et al.*, 1991) of this interaction was about NGF, a non-covalently linked dimer consisting of two 118-residue polypeptides, each of which contains three intramolecular disulfide bridges (Angeletti & Bradshaw, 1971).

In the structure of murine NGF (Figure 4) the rather elongated physiological dimer is formed by two monomers assembled in a parallel manner and stabilized by a large hydrophobic interface, in which most of the aromatic residues of the molecule are clustered.

The core of the NGF monomer revealed a novel tertiary fold dominated by a pair of two-stranded, twisted antiparallel β-sheets. These β-strands are connected by a number of highly flexible loops, in which most of the variability among the different neurotrophins is located: three β-hairpin loops (L1, L2 and L4) at one end and a reverse turn (L3) at the other.

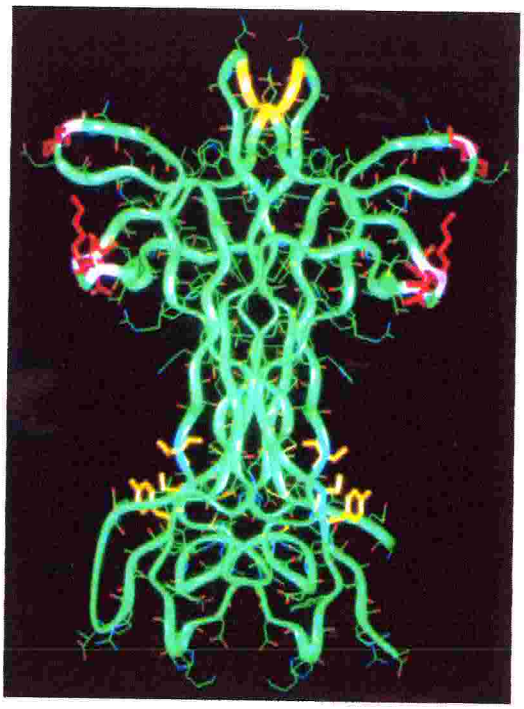


***Figure 4*: NGF structure**

At the level of the latter end, the three disulfide bonds are clustered, with two disulfide bridges and their connecting residues forming a ring

structure through which the third one passes to form a "cysteine-knot" motif (Murray-Rust *et al.*, 1991; Bradshaw, 1993). Both the amino and the carboxy-terminal were disordered in ths structure and so neither of them could be defined.

These structural data provided a structural framework for studies of ligand–receptor contacts using mutant, chimeric and chemically modified neurotrophins (Bradshaw *et al.,* 1994). These studies indicated that several charged residues within three β-hairpin loops are critical for the neurotrophin–p75NTR interaction, whereas a more extended surface is important for neurotrophin–trk receptor interaction (Ibáñez, 1995).

NGF-TrkAd5 structure

As mentioned above the starting point for co-crystallization experiments (Wiesmann *et al.*, 1999) was a construct consisting of domains 4 and 5 together with the linker connecting domain 5 to the transmembrane segment. However, over the long period it took for crystals to appear and to grow, the TrkA protein was partially proteolysed. The resulting crystals contain a complex consisting of one symmetrical NGF homodimer and two copies of only domain 5 of TrkA (residues 280–380) together with the first three linker residues (381–383). The overall shape of this complex resembles a bat with a length of 60 Å and a wing span of 95 Å, its body formed by the NGF dimer and the wings represented by the TrkA-d5 molecules (Figure 5) with the non-crystallographically-related halves of the complex very similar to each other.

Although ligand binding does not seem to induce any conformational changes in TrkA-d5 as it could be seen by superposition onto modelled TrkA-d5 structure in its unbound state (resulting in an r.m.s. difference of only about 0.5 Å for 98 Cα positions), a comparison with the structure of

unbound murine NGF reveals three regions of significant changes in the ligand, some induced by receptor binding.

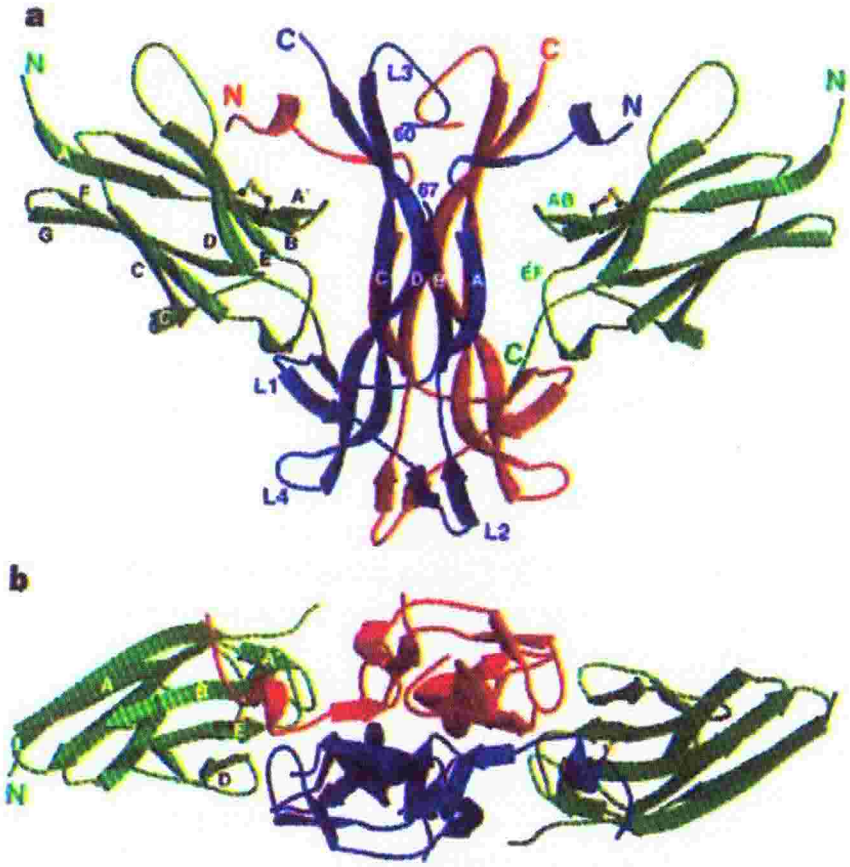


Figure 5: Overall structure of NGF/TrkA complex. Two orientation related by a rotation of 90° about the horizontal axis (NGF monomers in red and blue, TrkA-d5 in green)

Of the local differences, the different conformation of side chain of Trp 99, situated in the interface between the two NGF monomers, is unlikely to be a result of complex formation. The second change is very likely to be due to sequence differences between the human and murine proteins, considering that it involves a solvent exposed amino acidic stretch (61–66), that becomes disordered in the complex, while it is well defined in the murine NGF structures. Finally the most prominent conformational change concerns the amino-terminal residues (2–9), which was shown to

be fundamental for receptor binding (McInnes & Sykes, 1997). These residues are not disordered as observed in all previously reported NGF structures, probably because directly involved in the interface with TrkA-d5.

Overall, about 2,220 $Å^2$ of solvent-accessible surface is buried between NGF and each copy of TrkA-d5. The structure of the TrkA-d5-NGF complex shows that the ligand-receptor interface consists of two patches of similar size (Figure 6).

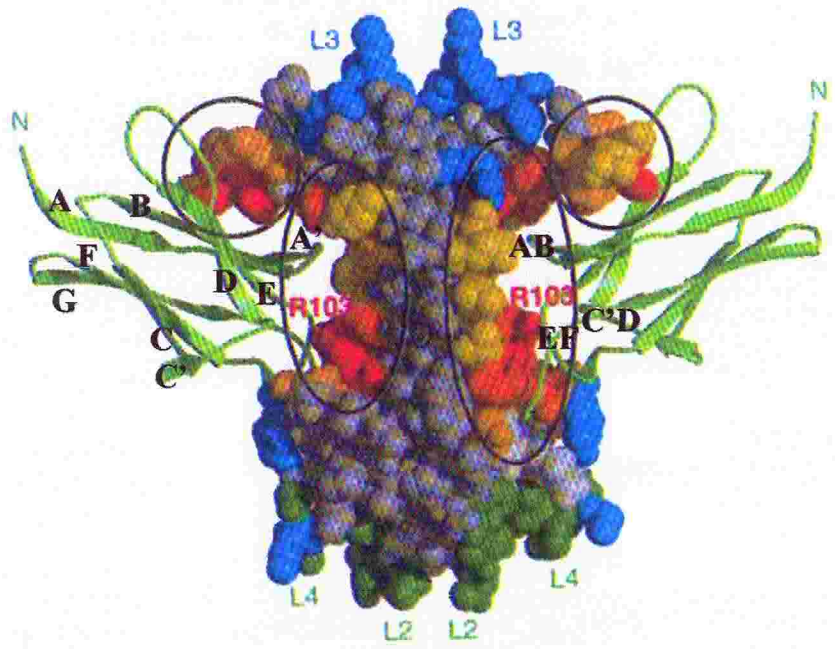


Figure 6: The interface between NGF and TrkA-d5

The first involves one of the β-hairpin loops (L1) and the central β-sheet, that forms the core of the homodimeric NGF molecule, and the loops at the carboxy-terminal pole of TrkA-d5. On TrkA-d5 several different regions are involved (residues from the AB, C'D and EF loops and also from the C-terminus). In particular the residues at the tip of the AB loop, which were responsible for the strand swapping in the structure of unbound TrkA-d5, are in very close contact with NGF. On NGF among

the residues from the four β-strands involved in the interaction surface, the side chain of Arg 103 (strictly conserved in all neurotrophins) protrudes to stack with its aliphatic chain against the phenyl group of Phe 327 of TrkA and to form a hydrogen bond by its guanidinium moiety with the carbonyl oxygen of Asn 349. Considering that most of the residues involved in this patch both on NGF and on TrkA are highly conserved respectively among other neurotrophin and among other Trk receptors, it is very likely that interactions similar to those described are conserved among all neurotrophin–receptor complexes. Therefore this first patch may constitute a conserved binding motif for all family members.

The second patch comprises the amino-terminal residues of NGF (2–13) packing against the solvent-exposed face of 'ABED' sheet of TrkA-d5. In particular among residues 6–9, that adopt a helical conformation upon complex formation, His 4 and Ile 6 appear to be the most important receptor-binding determinants with their side chains almost completely buried in the interface. In particular Ile 6 is embedded into a hydrophobic canyon on the surface of TrkA-d5, formed by the unusual disulphide bridge (between Cys 300 and Cys 350) on the surface of TrkA-d5 with the side chains of Val 294 and Leu 333 lining the walls. Pro 5 and Phe 7 of NGF also contribute to the hydrophobic interactions in this region, while His 4 and Glu 11 form hydrogen bonds to Ser 304 and Arg 347 of TrkA-d5. It is worth noting that the amino-terminal region of the neurotrophins is important for specificity; indeed, a hybrid of NT-3 with the N terminus replaced with that of NGF has high affinity not only for TrkC but also for TrkA (Urfer *et al.*, 1994). Therefore considering that at the level of the second patch the amino acids in the interface share little sequence conservation, both among the ligands and among the receptors,

this interface is very likely to be responsible for specificity among the neurotrophins.

Moreover very recently the structure of the TrkB-d5/neurotrophin-4/5 (NT-4) complex has been solved (Banfield *et al.*, 2001), confirming a consistent pattern of recognition in neurotrophin/ Trk receptor complex formation. Indeed as in the TrkA-d5/NGF interaction, NT-4 forms a crosslink between two spatially distant receptor molecules. The contacts formed in the complex can be divided into a conserved area similar to a region observed in the TrkA-d5/NGF complex and a second site-unique in each ligand-receptor pair-formed primarily by the ordering of the neurotrophin N terminus. This ordering appears to be directed largely by the receptor surface, with the resulting complementary surfaces providing the main determinant of receptor specificity.

TrkAd5 structure

TrkA-d5 alone folds into an I-type of the immunoglobulin superfamily, a type of domain that is often found in trans-membrane receptors and cell adhesion molecules. However it is interesting to underline that only 12 residues fulfill the 20 standard I-set residues: all these variations result in different local features at the C-terminal end of the domain, the most important for ligand binding.

In both reported structural studies (Ultsch *et al.*, 1999; Robertson *et al.*, 2001) it forms a strand-swapped homodimer with the amino-terminal A strand of each molecule forming part of the fold of the adjacent monomer (Figure 7a).

The most reasonable conformation of a folded domain was modeled by changing the position of one residue in the AB loop by a movement of about 3 Å of the residue immediately following the A strand, with the

assumption that no other changes were introduced by the observed strand swapping.

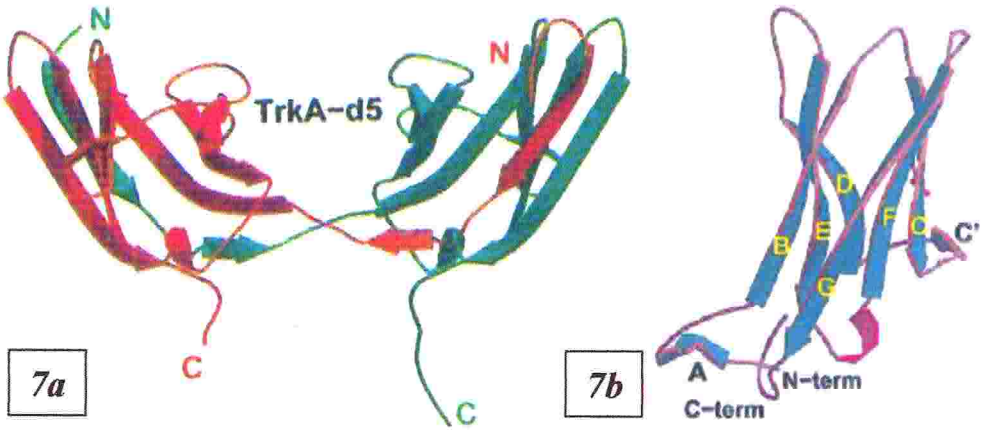


Figure 7: TrkA-d5 overall structure:
a) strand swapped dimers; b) modeled monomer

The crystallographic form of TrkA-d5 (Figure 7b) consists of a β-sandwich formed by two four-stranded β–sheets, one comprising strands A, B, E and D, and the other strands C', C, F and G with a short helical region in the loop connecting strands E and F. Due to a break in the hydrogen-bonding pattern to strand B, strand A can be subdivided in two regions, A and A'. All these strands are between 6 and 14 residues long except strand C' with only two hydrogen bonding to strand C.

The N-terminal end of the molecule comprises the amino-terminal of strand A and three loops (between strands BC with conserved *cis*-proline, between strands DE and FG), while the C-terminal end is formed by the carboxy-terminal end of strand G together with three loops (short remodeled between strands AB, between strands C'D and EF). It is worth noting that TrkA-d5 lacks the usual buried disulfide bond between strands B and F and instead shows an exposed disulfide bond between two conserved cysteines on strands B and E. This structure is consistent with the results of mutagenesis studies (Urfer *et al.*, 1998) by alanine scanning

that identified 17 residues causing a greater than fivefold reduction in ligand binding. Three of them are residues with structural roles (Thr340, Asn356, Asn365) and all the remaining (except Glu339) map in one of the two patches of the binding interface.

Moreover it is interesting to compare TrkA-d5 with TrkB-d5 and TrkC-d5, also reported (Ultsch *et al.*, 1999), especially at the level of the specificity patch. As mentioned above in TrkA-d5 the key feature of this patch is a hydrophobic pocket (Val294, Met296, Pro302, Leu333) with the disulphide bridge at its bottom (Figure 8). In the other two Trk-d5 structures only the disulphide bond at the bottom is conserved while the residues forming the walls are not. In particular in TrkB-d5 three of these respective position are occupied by Thr296, Asp298 and His335 giving rise to a hydrophilic pocket. In the case of TrkC-d5 the pocket is filled by a salt bridge between Arg316 and Glu322 at the top of the disulphide bridge.

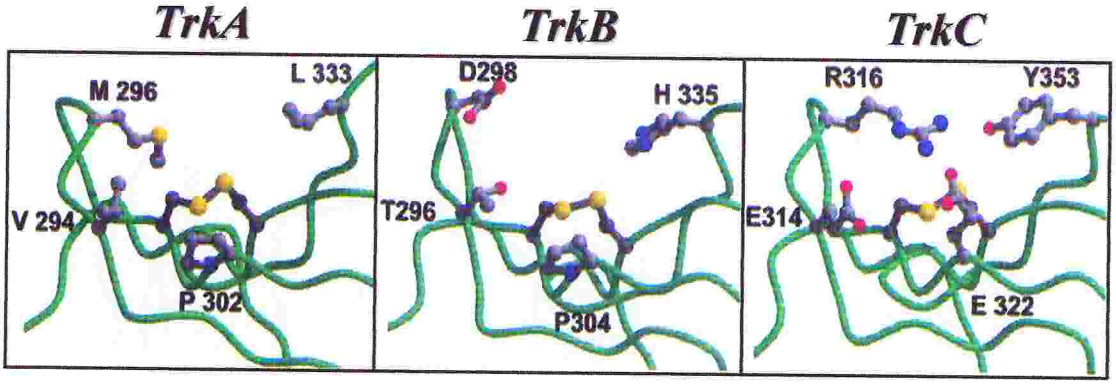


***Figure 8*: Comparison between specificity patches on Trks receptors**

As the crystallized complex includes only the fifth domain and a few of the following linker residues of TrkA, further potential interactions between NGF and other regions of the receptor are not available.

Blocking antibodies: an additional way to study NGF/TrkA interaction

In the absence of complete structural information on the remaining four domains comprising the extracellular portion of TrkA, we sought an additional approach to gain structural information on the receptor-ligand interaction.

We decided therefore to characterize from a structural point of view two neutralizing antibodies that are able to disrupt very efficiently the NGF-TrkA complex formation (Figure 9) in complementary ways.

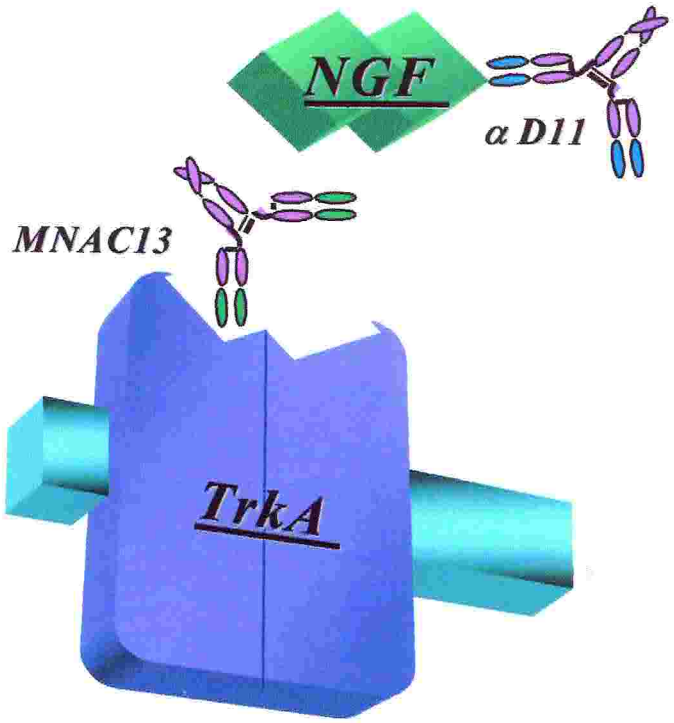


Figure 9: ***Blocking antibodies of NGF/ TrkA interaction***

- ***αD11 antibody***

The monoclonal antibody αD11, raised in rat against mouse NGF (Cattaneo *et al.,* 1988), has been cloned, sequenced and expressed in our laboratory (Ruberti *et al.*, 1993). It is able to recognize NGF molecule derived from different species, suggesting that the epitope is common and highly conserved across evolution. Testing αD11 binding activity towards a panel of NGF mutants, it was possible to map its epitope at the level of loop 41-49 and Lys 95 (Gonfloni, 1995).

The αD11 antibody showed a strong neutralizing activity inhibiting NGF binding to PC12 cells and its biological activity on these cells, as well as in primary sympathetic and sensory neurons. For this strong blocking activity this antibody has been extensively used in several *in vivo* experiments (Maffei *et al.*, 1992; Berardi *et al.*; 1994, Dominici *et al.*, 1994b, Molnar *et al.*, 1997 and 1998) and most importantly in transgenic mice (Ruberti *et al.*, 2000) expressing the neutralizing Mab αD11 was produced by the neuroantibody technique (Cattaneo & Neuberger, 1987; Cattaneo *et al.*, 1988, Piccioli *et al.*, 1995). Because the levels of transgenic anti-NGF antibodies are three orders of magnitude higher in adult than in newborn mice, effective inhibition of NGF actions occurs in adult animals only (Ruberti *et al.*, 2000). Therefore anti-NGF mice provided a unique opportunity to study the consequences of a chronic deprivation of NGF in aged animals. The observed phenotype (Capsoni *et al.*, 2000a) closely resembled that found in Alzheimer disease (Selkoe, 1991; Goedert, 1998), including neurofibrillary tangles, tau hyperphosphorylation, amyloid plaques, massive and widespread neuronal loss, cholinergic deficits in the basal forebrain, and selective behavioral impairments in retention and transfer of spatial memory tasks. The anti-NGF mice represent a comprehensive transgenic model for AD,

displaying a full complement of phenotypic hallmarks of the disease, including (Capsoni *et al.*, 2000b) an AD-linked muscular deficit defined as inclusion body myositis (Vogel, 1998).

- ***MNAC13 antibody***

As αD11 counterpart on the receptor, recently the monoclonal antibody MNAC13 (Cattaneo *et al.,* 1999) has been raised against TrkA extracellular domain using a congenic immunization protocol. After immunization of BALB/C mice using BALB/C-3T3 cells expressing the human TrkA receptor, hybridomas supernatant were screened for the ability to interfere with the neurotrophin-binding activity of TrkA receptor by a functional assay. The MNAC13 antibody inhibits the binding of NGF to TrkA-BALB/C-3T3 as efficiently as the neutralizing anti-NGF antibody αD11.

Considering that this monoclonal antibody is the first anti-TrkA receptor with demonstrated antagonistic properties in biological systems, MNAC13 has been cloned, sequenced and expressed in our laboratory.

It is worth noting that even if this antibody has been raised against native human TrkA, it also recognizes the rat TrkA receptor. MNAC13 antibody is able to bind TrkA receptors exposed on the cell surface and also to soluble forms of its extracellular domain without any cross-reactivity towards the related receptors TrkB and TrkC.

MNAC13 antibody has been shown to inhibit the binding of NGF to human and rat TrkA receptor on cells and is very effective in preventing the functional activation of TrkA by NGF in a variety of systems, including NGF-induced survival and differentiation of PC12 cells. Moreover it was confirmed to be a strong antagonist blocking the

interaction between NGF and TrkA at the level of rat basal forebrain cholinergic neurons *in vivo,* by the method of hybridoma implant. On the other hand by preliminary studies *in vitro* it was shown that the binding of MNAC13 to soluble forms of TrkA receptor is not inhibited by NGF, and vice versa. Therefore the mode of inhibition of the antibody requires an entirely native receptor, most likely because its binding site is very close or coincident to NGF docking site.

Structural studies on antibody molecules

Antibodies can be conveniently thought of as having three main protein regions (Figure 10): two identical antigen binding sites called Fab

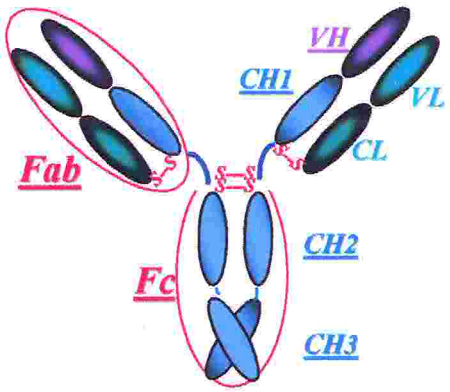


***Figure 10**: Antibody domain*

(the fragment having the antigen binding site) and the protein domain that is involved in immune regulation and is known as the Fc fragment(the fragment that crystallises). The region between the Fab and Fc fragments is called hinge. This segment allows lateral and rotational movement of the two antigen binding domains in order to let them free to interact with a large number of different antigen conformations.

Both the two heavy chain and the light chain polypeptides are identical, but while the first are approximately 55 KDa the other ones are about 25 KDa.

One light chain associates with the amino terminal region of one heavy chain to form an antigen-binding domain. The carboxy-terminal regions

of the two heavy chains fold together to make the Fc domain. The four polypeptide chains are held together by disulphide bridges and non covalent bonds.

The heterogeneity of the variable regions, which provides the structural basis for the large repertoire of binding sites necessary for an effective immune response, does not occur randomly throughout the variable region. Indeed the majority of the variable fragment (Fv) framework region (FW) is well conserved in structure among different antibodies. Large sequence and structural variability in antibodies is restricted to the hypervariable regions, consisting in six loops within the variable domains. Because these loops form the sites of contact with the antigen, they are referred to as CDRs (**c**omplementarity-**d**etermining **r**egions): three of them are located in the light chain (CDR-L1, L2 and L3) and three in the heavy chain (CDR-H1, H2 and H3). Five of them (all except H3) frequently fall into one of between 2 and 6 structural classes, referred as canonical classes and summarised in Table 1 (Chothia & Lesk, 1987; Chothia *et al.*, 1989). Members of each canonical class all have approximately the same backbone conformation, that is determined by the loop length and the presence of a number of key residue, both in the CDRs and the framework (in bold in Table 1).

Therefore to study the structure of the CDRs and of the antigen binding site, in principle crystallisation can be restricted to variable domains (Fv) that are the smaller portion of the antibody which maintains the binding specificity and affinity of the entire antibody.

CANONICAL CDRS CLASSES:	RESIDUE NUMBER:				CDR residues			FRAMEWORK residues			
CDRL1	L1	26	27	28	29	30	31	32	2	25	33	71
	1	S	S	S	V	-	-	S/Y	I	A	L/M	Y
	2	S	Q/G	D/N/S	I/V	-	-		I	A	L	Y/F
	3	S	E	S	L	L	N	F	I	S	L	F
	4	S	Q	S	L	V	H	Y	V	S	L	F

CDRL2	L2	50	51	52	48	64
	1	-	-	S/T	I	G

CDRL3	L3	91	92	93	94	95	96	90
	1	-	-	S/H	-	P	-	Q/N/H
	2	W	T	Y	P	L	I	Q
	3	W	G	R	N	P	-	Q

CDRH1	H1	26	27	28	29	30	31	32	34	94
	1	G	F/Y	T/S/D/I	F/L	S/T	S/D/-	F/Y	M/I/V	R/G
	1'	G	S/D	T/S	I/F	S/T	N/D	D	Y/W	R/N

CDRH2	H2	52a	b	c	53	54	55	71
	1				Y/G	H/D/S	G	
	2	P/T			G/N	S/T	G	A/L
	3	D/P/S			D/G	S/G	S/G	R
	4	N	K	G/P	N/Y	K/N	Y	R

***Table 1:** Canonical classes*

A more classical approach is to crystallise Fab fragments obtained from intact native monoclonal antibodies by a proteolytic cleavage (Figure 11a).

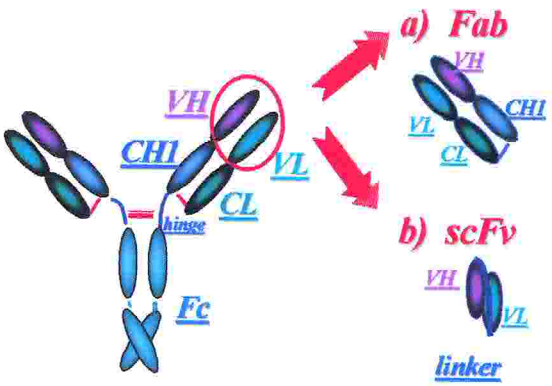


***Figure 11**: Antibody, Fab and scFv*

On the other hand the Fv fragment can be cloned and expressed in a recombinant form as single chain Fv (scFv), in which the two variable chains are joined together by a short engineered linker. This flexible amino acidic sequence is able to prevent dissociation of V_H and

V_L bridging the approximately 35Å between the carboxy terminus of one domain and the amino terminus of the other (Figure 11b).

About this thesis

The antibody molecule has been widely presented as a paradigm of biological individuality; functional and structural individuality and specificity is superimposed on a carefully conserved structure. With such an understanding of the antibody molecule, there is a continuing search for new applications of antibodies in many fields. In particular in this thesis two blocking antibodies MNAC13 and αD11, whose fundamental features are the high affinity and specificity for their antigens, have been exploited as an additional tool to investigate the mode of interaction between a soluble ligand (NGF) and a membrane receptor (TrkA).

Thus considering the crucial importance of NGF-TrkA interaction both in physiological and pathological contests, this thesis is aimed to a complete biochemical and structural characterization of these neutralizing antibodies.

Indeed the determination of their three-dimensional structures may provide deeper insights not only into the molecular basis of the mode of neutralization of the interaction between NGF and TrkA receptor, but also into the molecular basis of the observed high affinity of NGF for TrkA. Taken together, a detailed structural information on the antigenic recognition by the two neutralizing antibodies is expected to aid in the development of analogs as antagonists or agonists of neurotrophins that may have greater affinity or specificity for further experimental and therapeutic applications.

As a first step into the mechanism of TrkA-NGF binding, we present here the characterization of both blocking antibodies from a biochemical and from a structural point of view. Therefore this thesis is organized in two closely intertwined parts.

The first one is aimed to provide further characterization the binding activity of the two antagonists. While αD11 epitope on NGF has already been identified (Gonfloni, 1995), we tried to map MNAC13 binding site on TrkA receptor by two alternative strategies: serial deletions and phage display.

The second part of this work is focused on structural studies performed on both blocking antibodies. Here we report the purification, crystallization and high resolution three-dimensional structure determination of both neutralizing antibodies Fab fragments.

References

Angeletti, R.H.and Bradshaw, R.A. (1971). Nerve growth factor from mouse submaxillary gland: amino acid sequence. *Proc. Natl. Acad. Sci. USA*; **68**, 2417–2420.

Banfield, M. J., Naylor, R. L., Robertson, A. G., Allen, S. J., Dawbarn, D., Brady, R. L. (2001). Specificity in Trk receptor:neurotrophin interactions: the crystal structure of TrkB-d5 in complex with neurotrophin-4/5. *Structure* ; **9**, 1191-1199.

Barbacid, M. (1994). The Trk family of neurotrophin receptors. *J Neurobiol;* **25,** 1386-1403.

Barde, Y.A. (1994). Neurotrophic factors: an evolutionary perspective. *J. Neurobiol.*; **25**, 1329–1333.

Barinaga, M. (1994). Neurotrophic factors enter the clinic. *Science;* **264,** 272-274.

Bauer, J., Margolis, M., Schreiner, C., Edgell, C.-J., Azizkhan, J., Lazarowski, E. & Juliano, R. L. (1992). In vitro model of angiogenesis using a human endothelium-derived permanent cell line: contributions of induced gene expression, G-proteins, and integrins. *J. Cell. Physiol.;* **153**, 437-449.

Berardi N, Cellerino A, Domenici L, Fagiolini M, Pizzorusso T, Cattaneo A, Maffei L. (1994) Monoclonal antibodies to nerve growth factor affect the postnatal development of the visual system. *Proc Natl Acad Sci U S*; **91**, 684-688

Blöchl, A. and Sirrenberg, C. (1996). Neurotrophins stimulate the release of dopamine from rat mesencephalic neurons via Trk and p75Lntr receptors. *J Biol Chem.*; **271**, 21100-21107.

Bold, R. J., Ishizuka, J., Rajaraman, S., Perez-Polo, R., Townsend, C. M., Jr. & Thompson, J. C. (1995). Nerve growth factor as a mitogen for a pancreatic carcinoid cell line. *J. Neurochem.;* **64**, 2622-2628.

Bradshaw, R.A., Murray-Rust, J., Ibáñez, C.F., McDonald, N.Q., Lapatto, R., Blundell, T.L. (1994). Nerve growth factor. Structure/function relationships. *Protein Sci.*; **3**, 1901–1913.

Bonhoeffer, T. (1996). Neurotrophins and activity-dependent development of the neocortex.*Curr Opin Neurobiol.*; **6**, 119-26. Review.

Carter, B.D. and Lewin, G.R. (1997). Neurotrophins live or let die: Does $p75^{NTR}$ decide? *Neuron;* **18,** 187-190.

Capsoni, S., Ugolini, G., Comparini, A., Ruberti, F., Berardi, N., Cattaneo, A. (2000a). Alzheimer-like neurodegeneration in aged antinerve growth factor transgenic mice. *Proc Natl Acad Sci U S A.*; **97**, 6826-6831.

Capsoni, S., Ruberti, F., Di Daniel, E. & Cattaneo, A. (2000b). Muscular dystrophy in adult and aged anti-NGF transgenic mice resembles an inclusion body myopathy. *J. Neurosci. Res.;* **59**, 553-560

Cattaneo, A. & Neuberger, M. S. (1987). Polymeric immunoglobulin M is secreted by transfectants of non-lymphoid cells in the absence of immunoglobulin J chain. *EMBO J.;* **6**, 2753-2758.

Cattaneo, A., Rapposelli, B., Calissano, P. (1988). Three distinct types of monoclonal antibodies after long-term immunization of rats with mouse nerve growth factor. *J Neurochem*; **50**, 1003-1010

Cattaneo, A., Capsoni, S., Margotti, E., Righi, M., Kontsekova, E., Pavlik, P., Filipcik, P. & Novak, M. (1999). Functional blockade of tyrosine kinase A in the rat basal forebrain by a novel antagonistic anti-receptor monoclonal antibody. *J. Neurosci.;* **19,** 9687-9797.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., et al. (1989) Conformations of immunoglobulin hypervariable regions. *Nature*; **342**, 877-883. Review.

Chothia, C. & Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.;* **196**, 901-917.

Crowley, C., Spencer, S. D., Nishimura, M. C., Chen, K. S., Pitts-Meek, S., Armanini, M. P., Ling, L. H., MacMahon, S. B., Shelton, D. L., Levinson, A. D., (1994). Mice lacking nerve growth factor display perinatal loss of sensory and sympathetic neurons yet develop basal forebrain cholinergic neurons. *Cell*; **76**, 1001-1011

Cuello, A. C. (1996). Effects of trophic factors on the CNS cholinergic phenotype. *Prog. Brain Res.;* **109**, 347-358.

Djakiew, D., Delsite, R., Pflug, B. R., Wrathall, J., Lynch, J. H. & Onoda, M. (1991). Regulation of growth by a nerve growth factor-like protein which modulates paracrine interactions between a neoplastic epithelial cell line and stromal cells of the human prostate. *Cancer Res.;* **51**, 3304-3310.

Domenici, L., Cellerino, A., Berardi, N., Cattaneo, A., Maffei, L. (1994a). Antibodies to nerve growth factor (NGF) prolong the sensitive period for monocular deprivation in the rat. *Neuroreport*; **5**, 2041-2044

Domenici, L., Fontanesi, G., Cattaneo, A., Bagnoli, P., Maffei, L. (1994b). Nerve growth factor (NGF) uptake and transport following injection in the developing rat visual cortex. *Vis Neurosci.*; **11**, 1093-1102.

Donnerer, J., Schuligoi, R., Stein, C. (1992). Increased content and transport of substance P and calcitonin gene-related peptide in sensory nerves innervating inflamed tissue: evidence for a regulatory function of nerve growth factor in vivo. *Neuroscience*; 49, 693-698.

Dutar, P., Bassant, M. H., Senut, M. C., Lamour, Y. (1995). The septohippocampal pathway: structure and function of a central cholinergic system. *Physiol Rev.*; **75**, 393-427. Review.

Ebadi, M., Bashir, R.M., Heidrick, M.L., Hamada, F.M., Refaey, H.E., Hamed, A., Helal, G., Baxi, M.D., Cerutis, D.R. and Lassi, N.K. (1997). Neurotrophins and their receptors in nerve injury and repair. *Neurochem Int.;* **30,** 347-374

Fischer, W., Wictorin, K., Bjorklund, A., Williams, L. R., Varon, S., Gage, F. H. (1987). Amelioration of cholinergic neuron atrophy and spatial memory impairment in aged rats by nerve growth factor. *Nature*; **329**, 65-68

Frade, J. M., Rodriguez-Tebar, A., Barde, Y. A. (1996). Induction of cell death by endogenous nerve growth factor through its p75 receptor. *Nature*; **383**, 166-168

Gall, C. M., Isackson, P.J. (1989) Limbic seizures increase neuronal production of messenger RNA for nerve growth factor. *Science*; **245**, 758-761

Goedert, M. (1998). Neurofibrillary pathology of Alzheimer's disease and other tauopathies. *Prog. Brain Res.;* **117**, 287-306.

Goretzki, P. E., Wahl, R. A., Becher, R., Koller, C., Branscheid, D., Grussendorf, M. & Roeher, H. D. (1987). Nerve growth factor (NGF) sensitizes human medullary thyroid carcinoma (hMTC) cells for cytostatic therapy in vitro. *Surgery;* **102**, 1035-1042.

Hefti, F. (1986). Nerve growth factor promotes survival of septal cholinergic neurons after fimbrial transections. *J Neurosci.*; **6**, 2155-2162.

Hefti, F., Hartikka, J., Eckenstein, F., Gnahn, H., Heumann, R. & Schwab, M. (1985). Nerve growth factor increases choline acetyltransferase but not survival or fiber outgrowth of cultured fetal septal cholinergic neurons. *Neuroscience;* **14**, 55-68.

Holden, P. H., Asopa, V., Robertson, A.G., Clarke, A. R., Tyler, S., Bennett, G. S., Brain, S. D., Wilcock, G. K., Allen, S.J., Smith, S. K. & Dawbarn, D. (1997). Immunoglobulin-like domains define the nerve growth factor binding site of the TrkA receptor. *Nat. Biotechnol.;* 15, 668-672.

Hughes, P.E., Alexi, T. and Knusel, B. (1997). Axotomized septal cholinergic neurons rescued by nerve growth factor. *Neuroscience;* **78,** 1037.

Ibáñez, C.F. (1995). Neurotrophic factors: from structure function studies to designing effective therapeutics. *Trends Biotech.*; **13**, 217–227.

Jonhangen, M. E., Nordberg, A., Amberla, K., Backman, L., Ebendal, T., Meyerson, B., Olson, L., Seiger, A., Shigeta, M., Theodorsson, E., Viitanen, M., Winblad, B., Wahlund, L. O. (1998). Intracerebroventricular infusion of nerve growth factor affects in three patients with Alzheimer's disease. *Dementia Geriatr. Cogn. Disorders*; **9**, 246-257.

Junard, E. O., Montero, C. N., Hefti, F. (1990). Long-term administration of mouse nerve growth factor to adult rats with partial lesions of the cholinergic septohippocampal pathway. *Exp Neurol*; 110, 25-38.

Kaplan, D. R. & Stephens, R. M. (1994). Neurotrophin signal transduction by the Trk receptor. *J. Neurobiol.;* **25**, 1404-1417.

Kaplan, D.R., Miller, F.D. (1997) Signal transduction by the neurotrophin receptors. *Curr. Opin. Cell Biol.;* **9**, 213–221.

Knipper M, da Penha Berzaghi M, Blochl A, Breer H, Thoenen H, Lindholm D. (1994). Positive feedback between acetylcholine and the neurotrophins nerve growth factor and brain-derived neurotrophic factor in the rat hippocampus. *Eur J Neurosci* ; **6**, 668-671

Koizumi, H., Morita, M., Mikami, S., Shibayama, E. & Uchikoshi, T. (1998). Immunohistochemical analysis of TrkA neurotrophin receptor expression in human non-neuronal carcinomas. *Pathol. Int.;* **48**, 93-101.

Lachyankar, M. B., Ross, A. H., Litofsky, N. S., Condon, P. J., Quesenberry, P. J., Recht, L. D. (1997). TrkA expression decreases the in vivo aggressiveness of C6 glioma cells. *Cancer Res*; **57**, 532-536.

Lewin, G. R. and Mendell, M. L. (1994). Regulation of cutaneous C-fiber heat nociceptors by nerve growth factor in the developing rat. *J Neurophysiol.*; **71**, 941-949.

Levi-Montalcini, R. (1952). *Ann. N.Y. Acad. Sci.*; **55**, 330-343.

Levi-Montalcini, R., and Booker, B. (1960). Destruction of the sympathetic ganglia in mammals by an antiserum to a nerve growth protein. *Proc Natl Acad Sci USA;* **46**, 384-391

Lewin, G.R., and Mendell, L.M. (1993a). Nerve grow factor and nociception. *Trends Neurosci;* **16**, 353-359

Lewin, G.R., Ritter, A.M., and Mendell, L.M. (1993b). Nerve growth factor-induced hyperalgesia in the neonatal and adult rat. *J Neurosci;* **13**, 2136-2148

Lindholm, D., Heumann, R., Meyer, M., and Thoenen, H. (1987). Interleukin-1 regulates synthesis of nerve growth factor in non-neuronal cells of rat sciatic nerve. *Nature;* **330**, 658-659

Lindsay, R. (1996). Therapeutic potential of the neurotrophins and neurotrophin-CNTF combinations in peripheral neuropathies and motor neuron diseases. *Ciba Foundation Symposium;* **196,** 39-53.

Maliartchouk, S. and Saragovi, H.U. (1997). Optimal NGF trophic signals mediated by synergy of TrkA and p75 receptor-specific ligands. *J. Neurosci.;* **17,** 6031-6037

Marchetti, D., McQuillan, D. J., Spohn, W. C., Carson, D. D. & Nicolson, G. L. (1996). Neurotrophin stimulation of human melanoma cell invasion: selected enhancement of heparanase activity and heparanase degradation of specific heparan sulfate subpopulations. *Cancer Res.;* **56**, 2856-2863.

Maffei, L., Berardi, N., Domenici, L., Parisi, V., Pizzorusso, T. (1992). Nerve growth factor (NGF) prevents the shift in ocular dominance distribution of visual cortical neurons in monocularly deprived rats. *J Neurosci*; **12**, 4651-4662

Matsushima, H. & Bogenmann, E. (1993). Expression of trkA cDNA in neuroblastomas mediates differentiation in vitro and in vivo. *Mol Cell Biol.*; **13**, 7447-7456.

McDonald, N.Q., Lapatto, R., Murray-Rust, J., Gunning, J., Wlodawer, A., Blundell, T.L. (1991) New protein fold revealed by a 2.3 Å resolution structure of nerve growth factor. *Nature;* **354**, 411–414.

McDonald, J. I. S. & Meakin, (1996). Deletion in the extracellular domain of rat trkA lead to an altered differentiative phenotype in neurotrophin responsive cells. *Mol. Cell. Neurosci.;* 7, 371–390.

McGregor, L. M., McCune, B. K., Graff, J. R., McDowell, P. R., Romans, K. E., Yancopoulos, G. D., Ball, D. W., Baylin, S. B. & Nelkin, B. D. (1999). Roles of trk family neurotrophin receptors in medullary thyroid carcinoma development and progression. *Proc. Natl. Acad. Sci. U. S. A.;* 96, 4540-4545.

McInnes, C. & Sykes, B. D. (1997). Growth factor receptors: structure, mechanism, and drug discovery. *Biopolymers;* **43**, 339-366

McMahon, S.B. (1996). NGF as a mediator of innflammatory pain. *Phil Trans R Soc Lond (Biol);* **351**, 431-440

Molnar, M., Tongiorgi, E., Avignone, E., Gonfloni, S., Ruberti, F., Domenici, L., Cattaneo, A. (1998). The effects of anti-nerve growth factor monoclonal antibodies on developing basal forebrain neurons are transient and reversible. *Eur J Neurosci*; **10**, 3127-3140

Molnar, M., Ruberti, F., Cozzari, C., Domenici, L., Cattaneo, A. (1997). A critical period in the sensitivity of basal forebrain cholinergic neurones to NGF deprivation. *Neuroreport*; 8, 575-579

Montero, C. N., Hefti, F. (1989). Intraventricular nerve growth factor administration prevents lesion-induced loss of septal cholinergic neurons in aging rats. *Neurobiol Aging*; **10**, 739-743

Montero, C. N., Hefti, F. (1988). Rescue of lesioned septal cholinergic neurons by nerve growth factor: specificity and requirement for chronic treatment. *J Neurosci.*; **8**, 2986-2999.

Mufson, E. J., Lavine, N., Jaffar, S., Kordower, J. H., Quirion, R. & Saragovi, H. U. (1997). Reduction in p140-TrkA receptor protein within

the nucleus basalis and cortex in Alzheimer's disease. *Exp. Neurol.;* **146**, 91-103.

Muragaki, Y., Chou, T. T., Kaplan, D. R., Trojanowski, J. Q., Lee, V. M. (1997). Nerve growth factor induces apoptosis in human medulloblastoma cell lines that express TrkA receptors. *J Neurosci*; **17**, 530-542

Murray-Rust, J. et al., Bradshaw, R.A. (1993). Topological similarities in TGF-2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure;* **1**, 153–159.

Njå, A. and Purves, D. (1978). The effects of nerve growth factor and its antiserum on synapses in the superior cervical ganglion of the guinea-pig. *J Physiol.*; **277**, 55-75.

Ninkina, N., Grashchuck, M., Buchman, V. L., Davies, A. M. (1997). TrkB variants with deletions in the leucine-rich motifs of the extracellular domain. *J Biol Chem*; **272**, 13019-13025

Oelmann, E., Sreter, L., Schuller, I., Serve, H., Koenigsmann, M., Wiedenmann, B., Oberberg, D., Reufi, B., Thiel, E. & Berdel, W. E. (1995). Nerve growth factor stimulates clonal growth of human lung cancer cell lines and a human glioblastoma cell line expressing high-affinity nerve growth factor binding sites involving tyrosine kinase signaling. *Cancer Res.;* **55**, 2212-2219.

Olson, L., Nordberg, A., von Holst, H., Backman, L., Ebendal, T., Alafuzoff, I., Amberla, K., Hartvig, P., Herlitz, A., Lilja, A., Lundqvist, H., Langstrom, B., Meyerson, B., Persson, A., Viitanen, M., Winblad, B., Seiger, A. (1992). Nerve growth factor affects 11C-nicotine binding, blood flow, EEG, and verbal episodic memory in an Alzheimer patient (case report). *J Neural Transm Park Dis Dement*; **4**, 79-95.

Perez, P., Coll, P. M., Hempstead, B. L., Martin-Zanca, D., Chao, M. V. (1995). NGF binding to the Trk tyrosine kinase receptor requires the extracellular immunoglobulin-like domains. *Mol. Cell Neurosci.*; **6**, 97-105.

Revoltella, R. P. & Butler, R. H. (1980). Nerve growth factor may stimulate either division or differentiation of cloned C1300 neuroblastoma cells in serum-free cultures. *J. Cell. Physiol.;* **104**, 27-33.

Ritter, A. M., Lewin, G. R., Kremer, N. E., Mendell, L. M. (1991). Requirement for nerve growth factor in the development of myelinated nociceptors in vivo. *Nature*; **350**, 500-502

Robertson, A. G. S., Banfield, M. J., Allen, S. J., Dando J. A., Tyler, S. J., Bennett, G. S., Brain, S. D., Mason, G. G. F., Holden, P. H., Clarke, A. R., Naylor, R. L., Wilcock, G. K., Brady, R. L. & Dawbarn, D. (2001). Identification and structure of the Nerve Growth Factor binding site on TrkA. *Biochem. Biophys., Res. Comm.*; **282**, 131-141.

Ruberti, F., Capsoni, S., Comparini, A., Di Daniel, E., Franzot, J., Gonfloni, S., Rossi, G., Berardi, N., Cattaneo, A. (2000). Phenotypic knockout of nerve growth factor in adult transgenic mice reveals severe deficits in basal forebrain cholinergic neurons, cell death in the spleen, and skeletal muscle dystrophy. *J Neurosci.*; **20**, 2589-2601.

Safieh-Garabedia, B., Poole, S., Allchorne, A., Winter, J., and Woolfe, C.J. (1995). Contribution of Interleukin-1-beta to the inflammation-induced increase in nerve growth factor levels and inflammatory hyperalgesia. *Br J Pharmacol;* **115**, 1265-1275

Saragovi, H.U. and Burgess, K. (1999). Small molecule and protein-based neurotrophic ligands: agonists and antagonists as therapeutic agents. *Expert Opin Ther Patents;* **9,** 737-751.

Saragovi, H.U., Greene, M.I., Chrusciel, R.A., and Kahn, M. (1992). Loops and secondary structure mimetics: Development and applications in basic science and rational drug design. *Biotechnology;* **10,** 773-778.

Selkoe, D. J. (1991). The molecular pathology of Alzheimer's disease. *Neuron;* **6**, 487-498

Schneider, R. & Schweiger, M. (1991). A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogenic trk and trkB tyrosine kinase receptors. *Oncogene;* **6**, 1807-1811

Sieler, M. and Schwab, M. E. (1984). Specific retrograde transport of nerve growth factor (NGF) from neocortex to nucleus basalis in the rat. *Brain Res.*; **300**, 33-39.

Smeyne, R. J., Klein, R., Schnapp, A., Long, L. K., Bryant, S., Lewin, A., Lira, S. A., Barbacid, M. (1994). Severe sensory and sympathetic neuropathies in mice carrying a disrupted Trk/NGF receptor gene. *Nature*; **368**, 246-249.

Smith, D. E., Roberts, J., Gage, F. H., Tuszynski, M.H. (1999). Age-associated neuronal atrophy occurs in the primate brain and is reversible by growth factor gene therapy. *Proc Natl Acad Sci U S A*; **96**, 10893-8.

Snider, W.D. (1994). Functions of the neurotrophins during nervous system development: What the knockouts are teaching us. *Cell;* **77**, 627-638

Tagliabue, E., Castiglioni, F., Ghirelli, C., Modugno, M., Asnaghi, L., Somenzi, G., Melani, C. & Menard, S. (2000). Nerve growth factor cooperates with p185(HER2) in activating growth of human breast carcinoma cells. *J. Biol. Chem.;* **275**, 5388-5394.

Thoenen, H. (1995). Neurotrophins and neuronal plasticity.*Science*; **270**, 593-598. Review.

Tuszynski, M.H., Amaral,D G., Gage, F. H. (1990). Nerve growth factor infusion in the primate brain reduces lesion-induced cholinergic neuronal degeneration. *J. Neurosci.*; **10**, 3604-3614.

Ultsch, M. H., Wiesmann, C., Simmons, L. C., Heinrich, J., Yang, M., Reilly, D., Bass, S. H. & de Vos, A. M. (1999). Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J. Mol. Biol.;* **290**, 149-159.

Urfer, R., Tsoulfas, P., Soppet, D., Escandon, E., Parada, L. F., Presta, L. G. (1994). The binding epitopes of neurotrophin-3 to its receptors trkC and gp75 and the design of a multifunctional human neurotrophin. *EMBO J.* **13**, 5896-5909.

Urfer, R., Tsoulfas, P., O'Connell, L., Shelton, D. L., Parada, L. F. & Presta, L. G. (1995). An immunoglobulin-like domain determines the specificity of neurotrophin receptors. *EMBO J.;* **14**, 2795-2805.

Urfer, R., Tsoulfas, P., O'Connell, Hongo, J. A., Zhao, L. G., Presta, L. G. (1998). High resolution mapping of the binding site of TrkA for nerve growth factor and TrkC for neurotrophin-3 on the second immunoglobulin-like domain of the Trk receptors. *J Biol Chem.*; **273**, 5829-5840.

Verge, V. M., Merlio, J. P., Grondin, J., Ernfors, P., Persson, H., Riopelle, R. J., Hokfelt, T., Richardson, P. M. (1992). Colocalization of NGF binding sites, trk mRNA, and low-affinity NGF receptor mRNA in primary sensory neurons: responses to injury and infusion of NGF. *J Neurosci*; **12**, 4011-4022

Verrall, M. (1994). Lay-offs follow suspension of clinical trials of protein. *Nature;* **370,** 6.

Vogel, H. (1998). Inclusion body myositis-a review. *Adv. Anat. Pathol.;* **5**, 164-169

Wiesmann, C., Ultsch, M. H., Bass, S. H. & de Vos, A. M. (1999). Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor . *Nature;* **401**, 184-188.

Windisch, J.M., Marksteiner, R. and Schneider, R. (1995). Nerve growth factor binding site on TrkA mapped to a single 24-amino acid leucine-rich motif. *J Biol Chem;* **270,** 28133-28138

Windisch, J. M., Auer, B., Marksteiner, R., Lang, M. E. & Schneider, R. (1995). Specific neurotrophin binding to leucine-rich motif peptides of TrkA and TrkB. *FEBS Lett.;* **374**, 125-129.

Windisch, J. M., Marksteiner, R., Lang, M. E., Auer, B. & Schneider, R. (1995). Brain-derived neurotrophic factor, neurotrophin-3, and neurotrophin-4 bind to a single leucine-rich motif of TrkB. *Biochemistry;* **34**, 11256-11263.

Woolf, C.J., Safieh-Garabedian, B., Crilly, P., and Winter, J. (1994). Nerve growth factor contributes to the generation of inflammatory sensory hypersensitivity. *Neuroscience;* **62**, 327-331

Zafra, F., Hengerer, B., Leibrock, J., Thoenen, H., Lindholm, D. (1990). Activity dependent regulation of BDNF and NGF mRNAs in the rat hippocampus is mediated by non-NMDA glutamate receptors. *EMBO J*; **9**, 3545-3550.

PART 1:

BIOCHEMICAL CHARACTERIZATION and in vitro FUNCTIONAL STUDIES of MNAC13 ANTIBODY

CHAPTER 2:Epitope mapping by serial deletions

Introduction

To characterize protein-protein interaction and in particular in the case of antibody-antigen binding, an important point is trying to identify the precise site of the interaction itself.

Indeed proteins contain a large number of antigenic determinants or epitopes, that cover most regions, if not the entire protein surface (Benjamin *et al.*, 1984). Inherent in the concept of epitopes (coined by N. Jerne in 1960 to mean surface configurations, single determinants, structural themes, immunogenic elements, antigenic patterns, haptenic groups, specific areas) is the point that they are present on the surface of native folded proteins, while linear peptide sequences seen after protein unfolding may also be located in the inside the native molecule (and in this case are referred as "cryptotopes"). This concept reflects the difference in antigen recognition between B and T cells: T-lymphocytes recognize processed peptides in association with MHC (major histocompatibilty complex) molecules, whereas the antibody recognition is exquisitely specific for native conformation. As a consequence, epitopes may be composed not only by a single continuous length of the polypeptide chain (sequential epitopes) but also by several polypeptide segments distant in the primary sequence, but contiguous on the surface of the folded native protein (conformational or discontinuous epitopes).

Numerous studies (Goodman, 1989) have reported that epitopes on native proteins consist of short segments (around 4-7 amino acid residues) that

can be mimicked or mapped utilizing synthetic peptides of a similar length.

On the other hand all reported X-rays structures of antibody-protein antigen complexes have characterized structural epitopes that occupy large areas (from 680 to 880 $Å^2$) comprising from 15 to 22 amino acid residues belonging to several polypeptide segments distant in the primary sequence, but continuous on the surface of the folded native protein. Among these residues, only a subset contributes to most of the free binding energy, as it has been shown by molecular modelling and site-directed mutagenesis studies on protein antigens and antibodies. These contact residues that can be seen as the functional epitope are in many cases scattered over two or three discontinuous polypeptide segments. For this reason scanning for antibody binding with overlapping sets of peptides covering the antigen sequence have proven only partially successful.

Considering that, in the characterization of TrkA-MNAC13 interaction two approaches have been pursued to identify which region of the ligand is specifically recognized by the antibody and are reported in this and in the following chapter. A first method to map the localization of the epitope is to construct serial deletion mutants of the ligand and analyzing variation in binding affinity to the antibody.

In particular as previously reported TrkA extracellular domain (ECD) is quite complex and can be subdivided in three principal regions, whose folding is likely to be independent from each other: so a first step is to identify which of these domains is involved in MNAC13 interaction (Figure 1). Then by finer deletions of this domain it should be possible to map a smaller region carrying the epitope, which can be both linear and conformational.

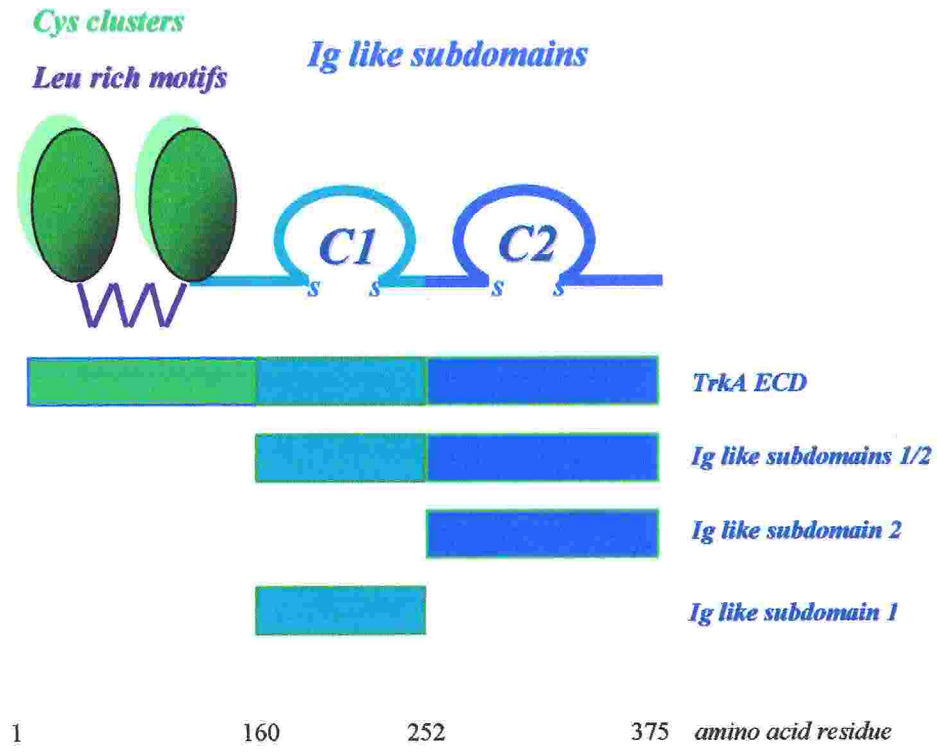


***Figure 1*: TrkA ECD subdomains**

Materials and methods

Cloning of TrkA subdomains in pMAL2C

To express single TrkA subdomains in fusion with MBP, they had been amplified by PCR using a set of oligonucleotides shown in Figure 2, carrying restriction sites at the 5' end to allow a subsequent directional cloning in the fusion vector.

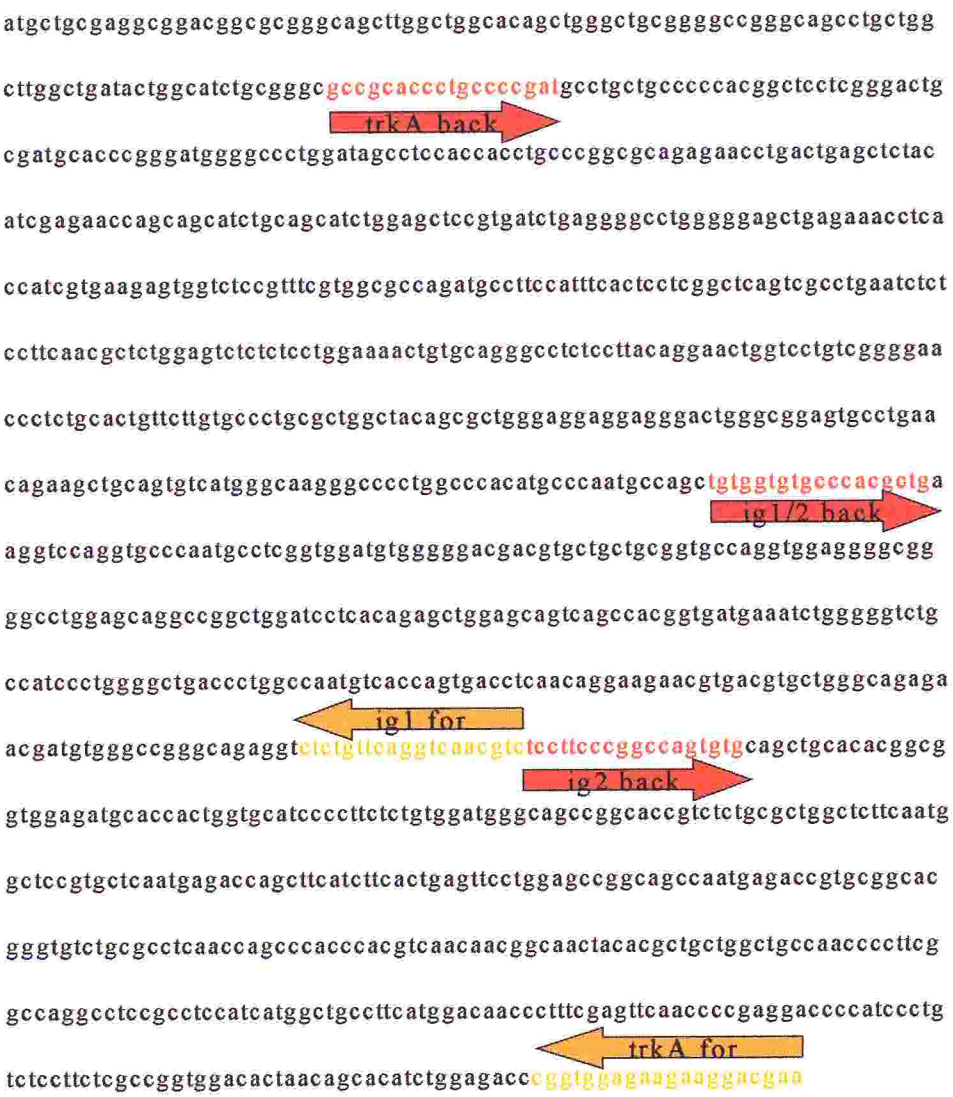


***Figure 2**: trka ECD nucleotidic sequence and PCR primers*

The obtained PCR products are summarized in Table 1.

PCR FRAGMENTS	***PRIMERS***
TrkA ECD	*trkA back/ trkA for*
ig like domains 1-2	*ig1-2 back /trkA for*
ig like domain 1	*ig1-2 back /ig 1 for*
ig like domain 2	*ig2 back /trkA for*

***Table 1*: PCR amplification of trkA subdomains**

After PCR amplification and removal of primers and enzyme by the PCR purification kit (Qiagen), each PCR product was digested by *EcoRI XbaI* for 2 hours at 37°C.

After another step of purification of the digested fragments by PCR purification kit, each of them had been cloned directionally *EcoRI XbaI* in the vector pMAL 2C (and in second instance also in the vector pMAL 2P) digested with the same restriction enzymes, dephosphorilated and purified, by ligation reactions at 22°C for 2 hours. Each ligation mix was used to transform electrocompetent DH5αF' by elettroporation. To isolate positive clones, a screening by PCR was performed directly on bacterial colonies. Then in order to control that no point mutations had been introduced during PCR amplification, two of each constructs were sequenced using the sequencing primer *MalE:* 5' CAG ACT GTC GAT GAA GCC 3'.

Cloning of TrkA ig1 like subdomain deletion mutants in pMAL

A similar procedure was used to express single ig1 like subdomain deletion mutants in fusion with MBP: they had been amplified by PCR

using a set of oligonucleotides shown in Figure 3, carrying the same restriction sites at the 5' end.

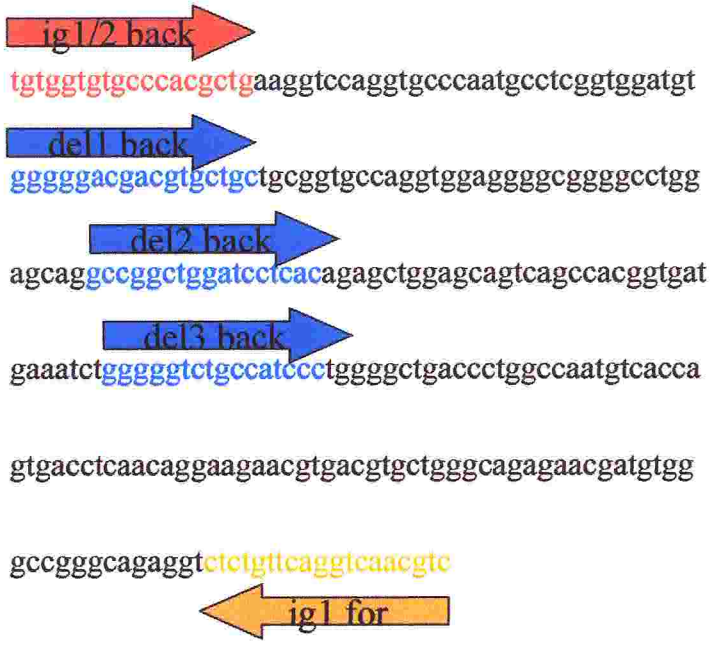


Figure 3 : ***ig1 like subdomain nucleotidic sequence and PCR primers***

The resulting PCR fragments, summarized in Table 2, had been cloned directionally in both vectors pMAL 2C and pMAL 2P as reported above.

PCR FRAGMENTS	***PRIMERS***
deletion 1	*del1 back /ig 1 for*
deletion 2	*del2 back /ig 1 for*
deletion 3	*del3 back /ig 1 for*

Table 2: ***PCR amplification of serial deletion mutants***

of ig like subdomain 1

After restriction digestion, ligation and transformation, positive clones had been first isolated by a PCR screening on bacterial colonies and then sequenced using the sequencing primer *MalE*.

Protein expression and extraction.

Each bacterial clone was first incubated shaking over night at 30°C in 2xTY Amp with 5% glucose and then, once diluted 1:100 in 2xTY Amp with 0.5% glucose, grown at 37°C to reach an OD_{600nm} of 0.7. To induce expression of each fusion protein, IPTG 0.5mM was added and left at 30°C for 5 hours. Then the bacteria were pelleted by centrifugation at 3500 rpm for 15 min 4°C.

Protein extraction from bacterial cells

To extract fusion proteins from bacterial cells, pellets were resuspended in 500µl of B-PER reagent (Pierce), vortexed and centrifuged at 13000 rpm for 5 min at 4°C to separate soluble proteins in the supernatant from the insoluble ones in the pellets. To purifiy inclusion bodies, each pellet was resuspended in 500µl of B-PER with lysozyme at a final concentration of 200µg/ml, vortexed and 1 ml of 1:10 diluted B-PER were added. After centrifugation for 10 min at 13000 rpm at 4°C, the pellet was resuspended in 1 ml of 1:10 diluted B-PER, vortexed and centrifugated for 10 min at 13000 rpm at 4°C: this step was performed two more times before resuspending the purified inclusion bodies in PBS.

Protein extraction from bacterial periplasm

For a periplasmic extraction, pellets were resuspended in 1/40 of total growth volume of PPB buffer (200mg/ml sucrose, 1mM EDTA, 30Mm Tris HCl pH 8) and incubated in ice for 20 min. After centrifugation at 5000 rpm for 15 min at 4°C to separate the supernatant (osmotic shock preparation) the pellets were resuspended in the same volume of Mg_2SO_4 5mM and incubated in ice for 20 min. Then both the osmotic shock preparation and periplasmic preparation were spun for 10 min at 13000

rpm at 4°C, pooled together and dialyzed over night at 4°C against PBS using Spectra-Por Membrane 12/14K (Spectrum).

Refolding procedure

After purification inclusion bodies had been resuspended in PBS and solubilized adding 8M urea. Trying to refold each fusion protein, they had been dialized using a Spectra-Por Membrane 12/14K (Spectrum) against PBS with decreasing concentration of urea and βmercapto ethanol in the following 6 steps:

PBS UREA 6.0M βmercaptoethanol 5.0mM

PBS UREA 4.0M βmercaptoethanol 3.3mM

PBS UREA 2.0M βmercaptoethanol 1.6mM

PBS UREA 1.0M βmercaptoethanol 0.8mM

PBS UREA 0.5M βmercaptoethanol 0.4mM

PBS

In each of these steps the proteins were left 2 hours at 4°C to allow a complete osmotic exchange.

ELISA assay

96 well plates were coated with 100 µl per well of the ligand (NGF or MNAC13 antibody) diluted in coating buffer (50mM Na carbonate pH 9.6) at a final concentration of 10 mg/ml and left over night at 4°C. After 3 washes with PBST (PBS Tween20 0.1%) and 3 washes with PBS plates were blocked by 200 µl of PBS (3% non fat dry milk) for 1 hour at room temperature. After 3 washes with PBST and 3 washes with PBS first 50 µl per well of PBS (6% non fat dry milk) and then the same volume of each bacterial lysate or supernatant were added and incubated 2 hours at room temperature. Then plates were washed 3 times with PBST and 3

times with PBS and incubated with the primary antibody, an anti-MBP antibody diluted 1:5000 in PBS (3% non fat dry milk) for 1 hour at room temperature. After the same washing procedure of the previous step the secondary antibody, an anti-rabbit antibody peroxidase conjugated, diluted 1:2000 in PBS (3% non fat dry milk) was added and incubated for 1 hour at room temperature. Finally after 3 washes with PBST and 3 washes with PBS 70 µl of substrate solution (diluted TMB) were added. After few min at room temperature in the dark the reaction was blocked adding the same volume of 0.1M H_2SO_4 and the intensity of each colorimetric signal was measured at 450mn by an Elisa Reader (*Spectra*).As a positive control TrkA immunoadhesin (TrkA ECD Camel), expressed by Baculovirus system, was employed.

Results

In order to express a number of serial deletion mutants of TrkA ECD, it has to be taken in account the crucial point of very low solubility when expressed in a recombinant way in bacteria (Holden *et al.*, 1997).

To try to overcome this problem a fusion vector (pMAL) was employed. It drives the expression of the cloned sequence in fusion with Maltose Binding Protein (MBP) at the amino terminal. In this way it has been shown that the folding of the carrier protein can favour the correct folding of the fusion partner, increasing in many cases the solubility of the fusion protein.

Moreover periplasmic localization of recombinant protein may help to achieve a further increase in the solubility, even if the expression levels are strongly reduced. For this reason in the first step of cloning trkA subdomains only pMAL2C was employed to express in fusion with MBP the TrkA ECD and its subdomains reported in Figure 2.

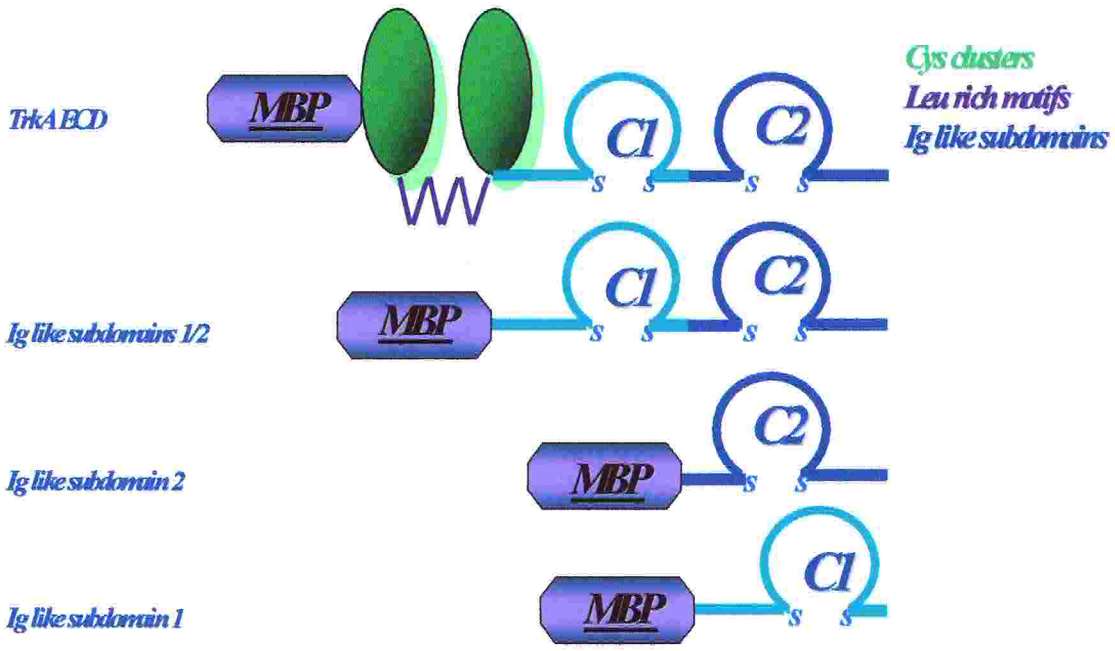


***Figure 2:* TrkA ECD and TrkA subdomains fusion proteins**

For each of the first three constructs, after cloning and sequencing, four different colonies were grown and induced for protein overexpression. After extraction from bacterial cells both the soluble and insoluble fraction were analyzed by SDS PAGE, followed by Coomassie staining (Figure 3B).

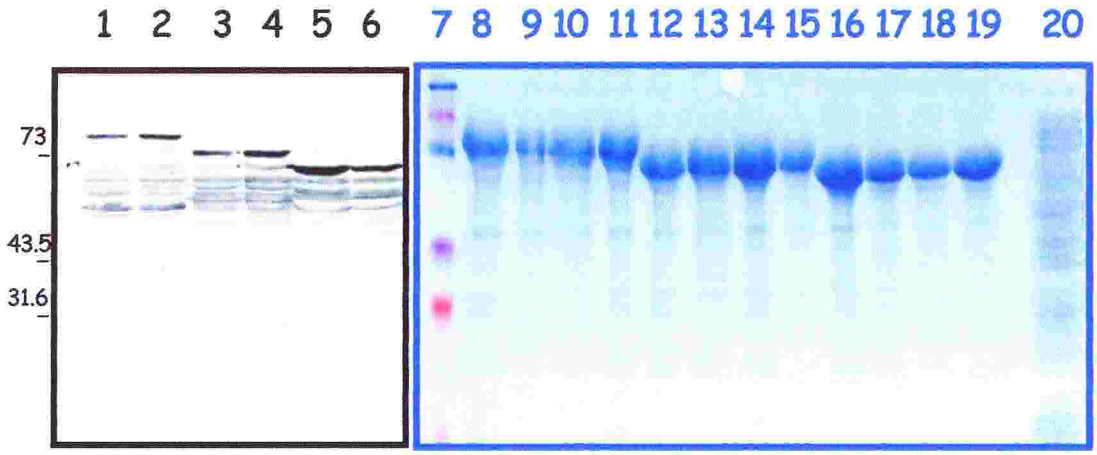


Figure 3A: Western blot anti MBP. Lane 1-2: TrkA ECD MBP (clones 3 and 4) soluble fractions; Lane 3-4: Ig like subdomains1/2 MBP (clones 9 and 10) soluble fractions; Lane 5-6: Ig like subdomain2 MBP(clones 15 and 16) soluble fractions.

Figure 3B: SDS PAGE and Coomassie staining. Lane7: MW; Lane 8-11: TrkA ECD MBP (clones 1 to 4) purified inclusion bodies; Lane 12-15: Ig like subdomains1/2 MBP(clones 5 to10) purified inclusion bodies; Lane 16-19: Ig like subdomain2 MBP(clones 13 to 16) purified inclusion bodies; Lane 20: before induction.

Here SDS PAGE and Coomassie staining of soluble fractions is not reported because the amounts of recombinant proteins were negligible and can be clearly seen only by Western Blot (Figure 3A). Therefore from these preliminary experiments of over-expression it was possible to conclude that no improvements in TrkA subdomains solubility could be

achieved even with a fusion partner and favorable induction conditions (low temperature and concentration of inducing agent). In order to try to obtain properly folded proteins after solubilization with urea 8M, a simple protocol of refolding by dialysis was performed, diluting progressively the concentration of denaturing agent. To check that in this way all these fusion proteins refolded in a proper way, a control of biological activity in vitro was performed testing their ability to bind NGF.

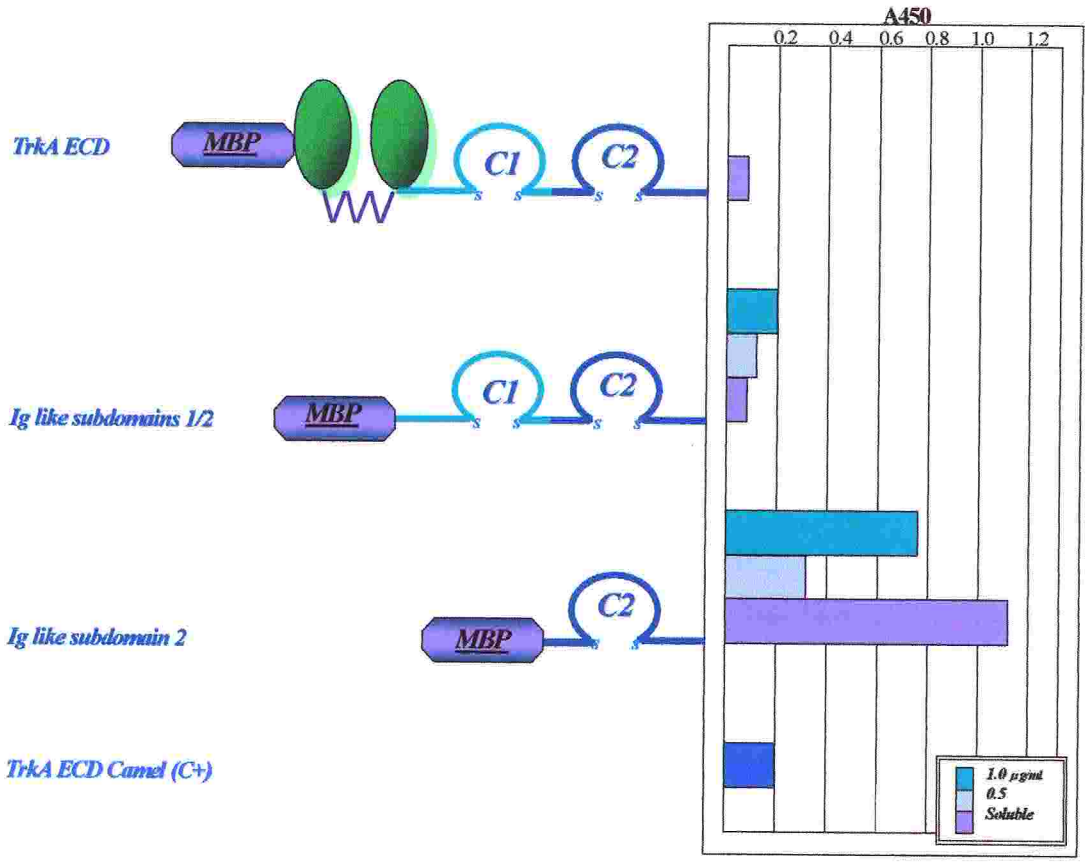


***Figure 4:* Elisa to test NGF binding**

In Figure 4 results of an Elisa experiment are reported: after NGF coating the ability to interact of soluble fractions and purified refolded insoluble fractions for each fusion protein was tested.

subdomains is involved in the interaction with MNAC13 immunoglobulin. As shown in Figure 5 no interaction could be detected between MNAC13 and TrkA Ig2 like subdomain as suggested by previous results of competition experiments between MNAC13 and NGF for TrkA ECD binding (Cattaneo et al., 1999). On the other hand a strong binding activity towards the immunoglobulin was detected for Ig1-2 like subdomains, suggesting that MNAC13 specifically recognizes a region at the level of Ig1 like subdomain. Interestingly it was possible to detect the interaction between MNAC13 and refolded TrkA ECD even if it has been shown to be unable to bind NGF. A possible explanation could be that refolding process was only partial and could be more effective at the level of MNAC13 binding site that NGF binding site. Another possible hypothesis may be that MNAC13 can interact with its ligand even if it is not completely in a native state, considering that this antibody can be used also in Western Blot (unpublished data): so the interaction can occur also with a partially refolded protein.

Therefore to check that MNAC13 binding site is really mapped on TrkA Ig1 like subdomain and to try to identify a smaller region necessary and

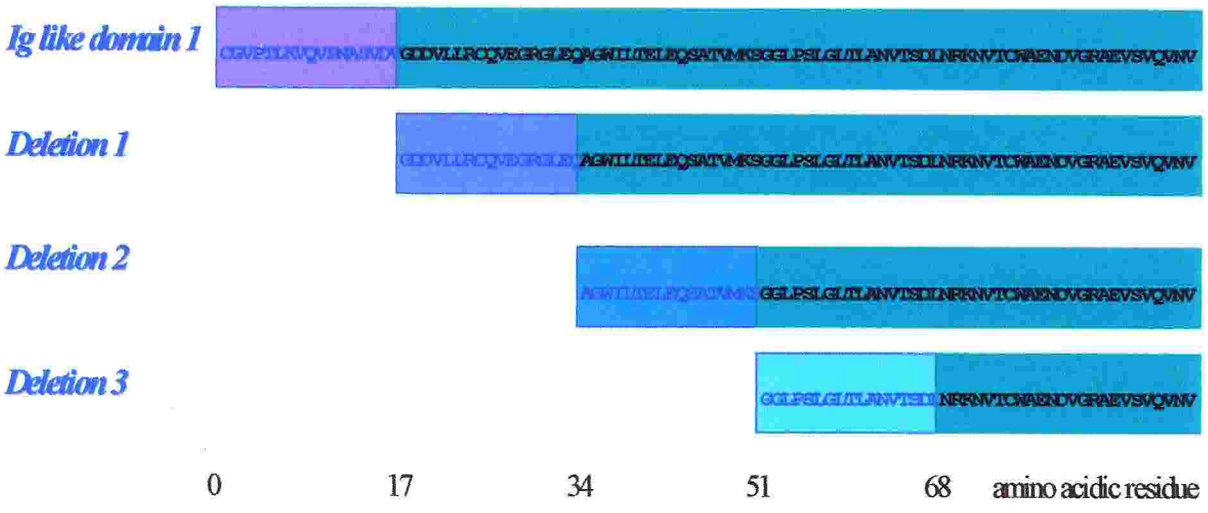


***Figure 6:* Ig1 like subdomain and its serial deletions**

sufficient for this interaction, a second step of cloning was performed.
This time both pMAL2C and pMAL2P were employed to express in fusion with MBP the TrkA ECD Ig1 like subdomain and its serial deletions reported in Figure 6; in the meanwhile also the previous three fusion proteins were cloned into pMAL2P.
As expected periplasmic expression of all of the fusion proteins was quite low and could be detected only by Western Blot (Figure 7).

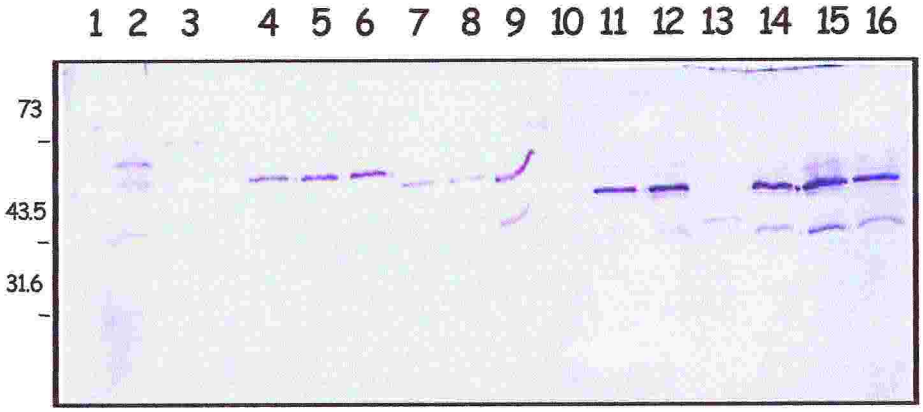


Figure 7: Western blot anti MBP. Lane 1: TrkA ECD MBP periplasmic extract; Lane 2: Ig like subdomains2 MBP periplasmic extract; Lane 3: Ig like subdomain1-2 MBP periplasmic extract; Lane 4-6: Ig like subdomain1 MBP (clones 8-10) periplasmic extracts; Lane 7-9: Ig like subdomain1 *Deletion*1 MBP (clones 1-3) periplasmic extracts; Lane 10-12: Ig like subdomain1 *Deletion2* MBP(clones 4-6) periplasmic extracts; Lane 13-15: Ig like subdomain1 *Deletion3* MBP(clones 7-9) periplasmic extracts.

Two clones for each construct were tested by an Elisa assay with MNAC13 coating to verify that the antibody interacts with TrkA ECD at the level of Ig like subdomain1 (as suggested by previous Elisa results) and to try to identify a smaller region of this domain necessary and sufficient for the interaction itself.

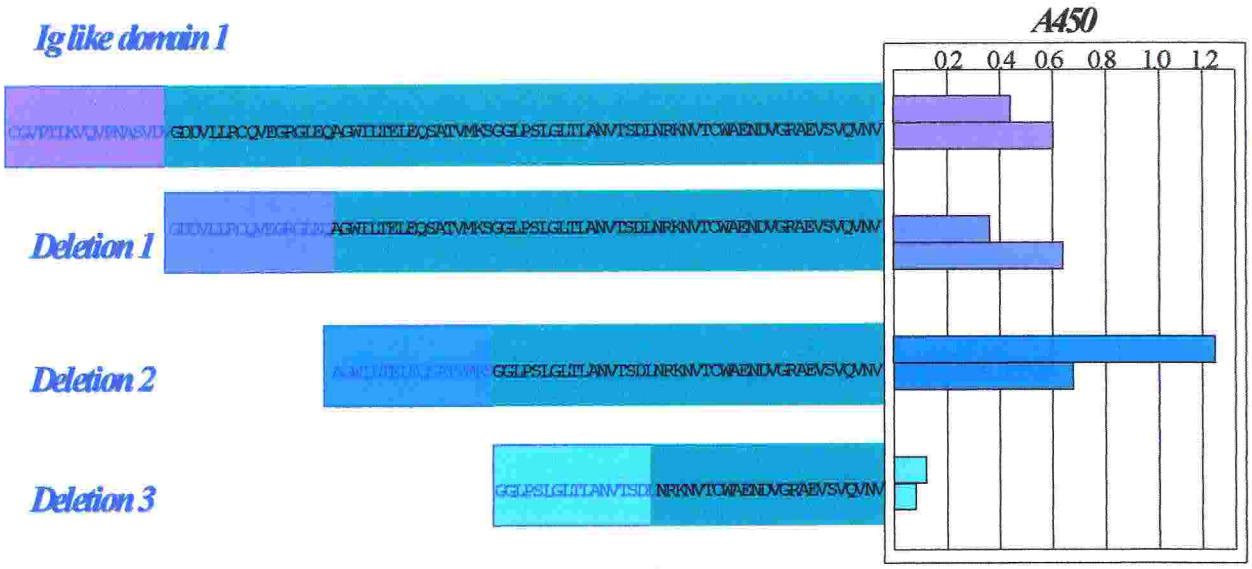


***Figure 8:* Elisa assay to map MNAC13 binding**

As shown in Figure 8 it was possible to confirm that MNAC13 recognizes Ig like subdomain1.

Moreover MNAC13 binding activity is retained only towards the first two constructs obtained by serial deletions (starting from the amino terminal of Ig like subdomain1) but is almost completely lost in the last smaller one (*Deletion3*). These data seemed to suggest that MNAC13 epitope was mapped inside the amino acidic sequence spanning between amino acid 34 and 51 (AGWILTELEQSATVMKS) of Ig like subdomain1, considering that this region distinguished *Deletion2* from *Deletion3*.

Discussion

In order to map the localization of the epitope recognized by MNAC13 on TrkA extracellular domain (ECD) a pool of serial deletion mutants of the ligand were constructed and expressed to analyze correlation between portions of amino acidic sequences of TrkA ECD and binding affinity to the antibody.

To perform this kind of study a preliminary problem is expression of correctly folded recombinant proteins to be analyzed. Indeed many heterologous proteins that are overexpressed lead to the formation of insoluble protein aggregates known as inclusion bodies. Current estimates are that 15-20% of human gene constructs expressed in E. coli are soluble, 20-40% form as inclusion bodies and the remainder do not express significantly or are degraded. Unfortunately also TrkA ECD protein was reported to be quite insoluble if expressed in bacterial hosts.

Anyway it has been reported that inclusion body formation can be minimized by adopting several modifications in expression conditions, such as induction at low temperatures, lowering the concentration of inducing agent and expressing the protein as a fusion with a "solubilizing partner". Following these hints all TrkA ECD constructs were expressed in fusion with MBP obtaining quite good level of intracellular expression but any improvement in solubility. Taken these results in account two strategies have been followed. First of all considering that inclusion body derived proteins can be advantageous because are produced in high yields and are generally protected from proteolytic degradation, a refolding procedure was performed. Unfortunately from biological activity data (i.e. binding activity towards NGF and MNAC13) it was clear the refolding efficiency was inversely correlated to the complexity of the

protein: indeed it seemed to be very effective for single domain Ig like subdomain2 and quite low or at least partial for TrkA ECD.

On the other hand periplasmic secretion can increase recombinant protein solubility; so each recombinant fusion protein was also expressed using a plasmidic vector carrying periplasmic secretion signal. By this system anyway even if the expression of all the recombinant constructs was quite similar and in principle no folding problems should occur making them suitable to be compared in terms of binding activity, the expression level was extremely poor.

However relying on this bacterial expression system, it has been possible to analyze MNAC13 binding activity to different TrkA ECD regions even if only qualitatively. First of all it was possible to compare MNAC13 interaction with the three principal subdomains that constituted TrkA ECD: from this more general and "macroscopic" point of view Ig like subdomain1 was identified as necessary and sufficient for MNAC13 recognition. Then inside this subdomain finer deletions have been expressed and compared in terms of binding activity towards MNAC13 in order to try to map a smaller region carrying the epitope, which can be both linear and conformational. The whole set of data that have been obtained seemed to exclude that MNAC13 recognition took place at the most amino terminal portion of Ig like subdomain1 (amino acid 1 to 34: CGVPTLKVQVPNASVDVGDDVLLRCQVEGRGLEQ. Moreover it might be likely to conclude that MNAC13 epitope was localized at the level of the sequence spanning between amino acid 34 and 51 (AGWILTELEQSATVMKS) of Ig like subdomain1, considering that this region was present in *Deletion2* but not *Deletion3* together with the fact that MNAC13 lost its binding activity in the case of *Deletion3*. However an alternative explanation of these results could be that *Deletion3* recombinant protein was too small in terms of number of amino acidic

residues to permit a proper folding of the deleted subdomain and this could be particularly likely if MNAC13 epitope was a conformational one.

References

Goodman J. D. (1989). In antigenic deyerminants and immune regulation, E. Sercarz, ed. (Basel: Karger), pp. 1-22

Holden, P. H., Asopa, V., Robertson, A. G., Clarke, A. R., Tyler, S., Bennett, G. S., Brain, S. D., Wilcock, G. K., Allen, S. J., Smith, S. K., Dawbarn, D. (1997). Immunoglobulin-like domains define the nerve growth factor binding site of the TrkA receptor. *Nat Biotechnol*; **15**, 668-672

Sachdev, D., Chirgwin, J. M. (1998). Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expr Purif* ; **12**, 122-132

Wulfing, C., Pluckthun, A. (1994). Protein folding in the periplasm of Escherichia coli. *Mol Microbiol* ; **12**, 685-692

CHAPTER 3: Epitope mapping by phage display

Introduction

Besides the method reported in the previous chapter, several alternative approaches could be followed to map in a finer way which region of the ligand is specifically recognized by the antibody.

In particular any peptide domain can in principle be used as a carrier for oligopeptides to be selected for binding to specific ligands. Moreover if the carrier peptides are linked to their genes, affinity purification of the fusion proteins will result in the selection and cloning of their coding sequences. In this way carrier peptides become exposition vectors that are analog of chemically synthesized epitopes but they have the key advantages of clonability and replicability.

For this purpose several outer membrane proteins or proteins, that form bacterial thread-like organelles, have been used to present oligopeptides to the exterior of the bacterial cell (Charbit *et al.*, 1988; Agterberg *et al.*, 1989, Freudl, 1989; Hedegard & Klemm, 1989).

Moreover turning from bacteria to phage increases by a factor of 10^3 the number of clones that can be handled in each experiment and has the further advantage of easier storage and construction of large peptide libraries. This kind of large library, called epitope library, is a particular application of phage display technology originated from the pioneering work of G. Smith (Scott & Smith, 1990; Cesareni, 1992; Smith & Scott 1993). They can display ten millions of different peptide epitopes: among them it is possible to select for binding to an antibody or other ligand protein by affinity purification of reactive phage from the library,

propagating individual phage clones and sequencing the inserted region in their DNA to find out the amino acidic sequence of the displayed peptides.

The maximum possible diversity of different peptide sequences or shapes displayed by a given library is determined by the length of the peptide insert. In practical terms, for libraries containing inserts of 8 amino acids or greater, the total complexity of the library is limited not only by theoretical constraints, but by practical and methodological considerations which make it hard to derive libraries with a total complexity of much above 10^{11}.

To develop such phage epitope libraries there are two different approaches based on the properties of the two component of filamentous phage coat (Figure 1). First of all it is possible to insert coding sequences amino terminus of the product of gene III (pIII protein) without any interference with phage infectivity (Scott & Smith, 1990; Cwirla *et al*, 1990; Devlin *et al*, 1990). Indeed pIII, present in 3-5 copies at one end of the phage filament, is involved in phage adsorption and so some insertions that disturb its conformation may result in phage particles that are not infective.

On the other hand other libraries are based on the product of gene VIII (pVIII protein), the major coat protein of phage f1 (Greenwood et al., 1991, Felici et al., 1991): during the process of phage assembly 2700 copies polymerize forming a helical array around the chromosomal DNA core. Also in this case it has been shown that it is possible to display up to 6 amino acids at its amino terminus without any interference with translocation across the membrane or phage assembly resulting in exposition at higher density on the phage capsid.

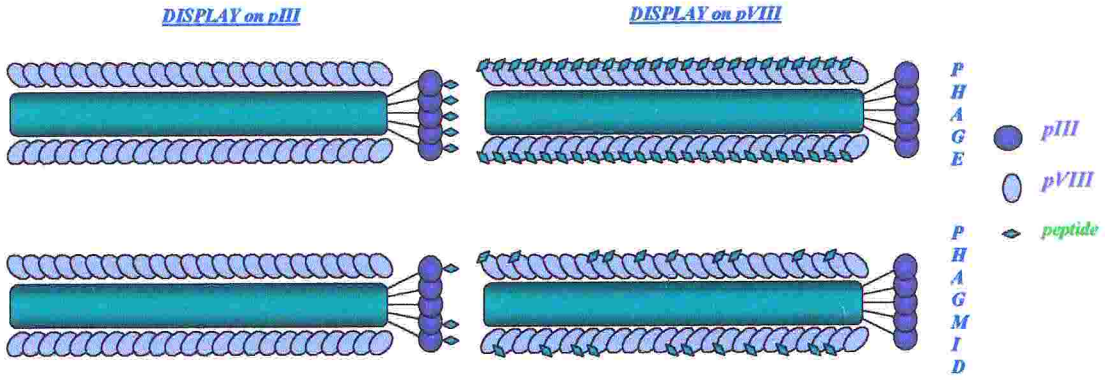


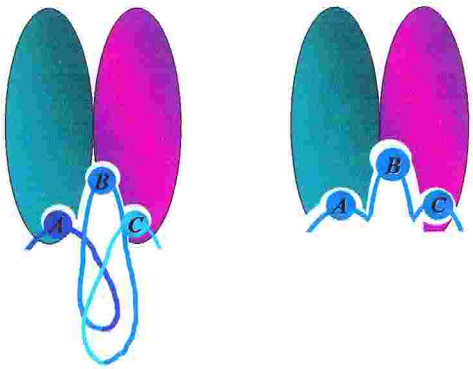
***Figure 1*: Phage display libraries**

Otherwise longer peptides cause some interference between the closely packed N-terminally extended subunits. As a consequence wild type pVIII must be co-expressed with the recombinant one. This can be obtained using phagemids, that are vectors containing besides the recombinant pVIII the packaging signal of Ff. Phagemids are able to drive very high level expression of recombinant protein but are not infectious, since they require the presence of helper phages for replication and packaging. In this conditions phagemids are preferentially packaged because the packaging signal of helper phages has been disabled and it is then possible to assemble hybrid capsids where the recombinant pVIII is dispersed in an otherwise wild-type coat. This mean that phagemid preparations are both phenotypically and genotypically heterogeneous. A survey of peptides of 9 amino acids of random sequence has shown that at last 80% of them can be successfully translocated across the membrane and presented on phage coats when inserted after the 5^{th} amino acid of the mature pVIII. The density of the peptide on the phage surface, which somehow depends on the peptide sequence, varies between 10^3/phage (more than 1/3 of the total number pVIII subunits) and less than 10^2/phage (Cesareni, 1992).

Screening such random peptide libraries has become an important tool for the evaluation of protein-protein interaction (Burritt *et al.*, 1996) and in

particular has led to the identification of many epitopes recognized by monoclonal antibodies (Sibille & Strosberg, 1997; Stephens *et al.*, 1995).

It is worth noting that by this approach it is possible to identify not only continuous (linear) epitopes but also mimics of discontinuous conformational dependent epitopes, composed of residues distant in the primary sequence but adjacent in the folded structure (Figure 2).



A) B)

***Figure 2:** Discontinuos (A) and linear (B) epitope*

Therefore in order to select binding peptides for antiTrkA MNAC13 monoclonal antibody, a phagmidic library exposing random peptides of 9 amino acids fused to pVIII (kindly provided by prof. G. Cesareni) was employed. Indeed considering that this library has been successfully tested to recover, from this random collection of peptides, the ones that bind a monoclonal antibody raised against oligopeptide 163-171 of interleukin 1β, it has been shown to be suitable to identify peptides similar to linear epitopes or whose structure mimic discontinuous or conformational epitopes.

As an alternative to classical method of panning against an antibody that has been immobilized on a surface (Scott & Smith, 1990), the library can be reacted with the antibody in solution after affinity capture of the antibody onto protein A and/or protein G sepharose beads (Figure 3).

In addition to requiring substantially less antibody per experiment than surface panning, solution panning can result in improved accessibility of the antigen binding site to phage displayed peptides, as well as avoiding partial denaturation of the antibody on the plastic surface. Usually

fortuitous selection of peptide sequences that specifically bind protein A or protein G can be avoided alternating rounds of panning between protein A-sepharose and protein G-sepharose. Unfortunately mouse IgG_1

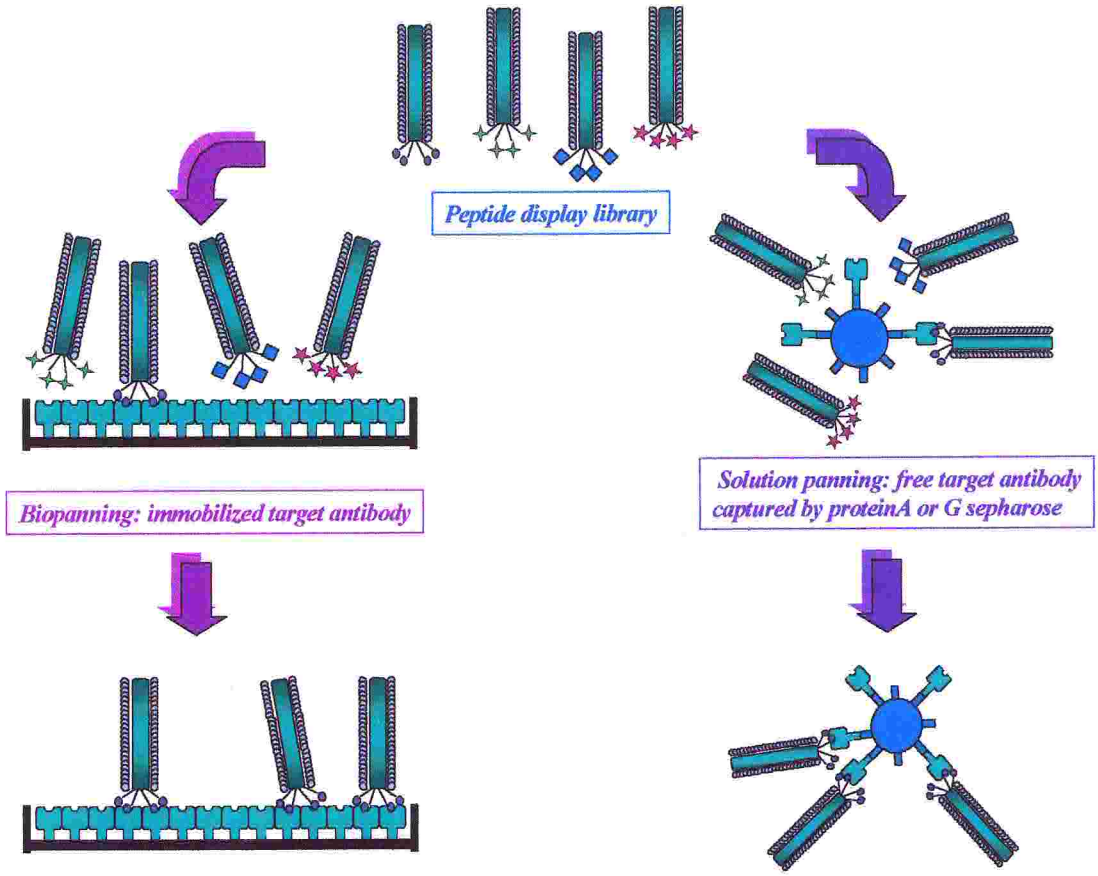


***Figure 3**: Solution panning and biopanning procedure*

shows quite low affinity towards protein A, so protein G has been used in the first rounds and only one cycle of selection was performed using protein A in basic environment to increase binding affinity. After several rounds of affinity enrichment, the selected phages are eluted, amplified and sequenced. Considering that phages are exposition vectors, each coding sequence and peptide are physically linked, and so the amino acidic sequence can be easily deduced from DNA sequences.

Materials and methods

Phagmidic library

A phagmidic library exposing random peptides of 9 amino acids fused to pVIII (kindly provided by prof. G. Cesareni) used in this work. It has been obtained cloning in PC89 (BamHI EcoRI) a mixture of completely degenerate sequence of 26 nucleotides (Felici *et al.*, 1991). It was possible to evaluate insertion efficiency by color phenotype because of the presence of lacZ gene downstream gVIII and of two unique restriction sites flanking an in-frame amber mutation. In this way they could see that 40% of the $9.4x10^7$ indipendent transformants formed blue colonies, indicating that they contained an oligonucleotide insertion without any in-frame nonsense codon.

Solution panning

To the antibody monoclonal antibody on the surface of the resin, 0.1 ml of protein G sepharose, washed twice with TTBS (50mM Tris pH 7.5, 150mM NaCl, 0.5% Tween20), was incubated at room temperature with 15 μg/ml of MNAC13 antibody for 2 hours with rotation. After two brief washes by TTBS the resin was blocked first by TTBS with BSA 10 mg/ml for 2 hours at 4°C and then by 2.5 10^{10} pfu/ml of uv-killed phage particles in TTBS with BSA 1 mg/ml for 4 hours at 4°C with rotation. Uv-inactivated blocking phage particles, defective in replication, have been obtained by irradiation with a germicidal lamp at a distance of 60 cm for 15 min.

Then the resin was incubated with the pVIII phagmidic library (10^{11} pfu/ml in TTBS with BSA 1 mg/ml) over night at 4°C with rotation.

The unbound supernatant (U) was stored and after 10 washes for 10 min with TTBS bound phage particles (E) were eluted by incubation with 0.9 ml of elution buffer (100mM HCl, BSA 1 mg/ml) for 10 min at 37°C with strong agitation and quickly neutralized by 150 µl Tris 1.5M, pH 8.8. At the end of each of the three round of selection performed input phage, unbound (U) and eluted (E) ones have been titred.

To alternate a round of panning using protein A-sepharose, the procedure was the same, except that TTBS pH was raised to 8.5 to increase binding affinity of MNAC13 monoclonal antibody to protein A.

Phage storage and amplification

Eluted phage particles were incubated with freshly grown bacteria (a phage: bacteria ratio of 100:1) for 30 min at 37°C standing with occasional agitation to perform the infection event. Then bacteria were plated and after over night incubation at 37°C were scraped by adding 1.5 ml of LB and stored in glycerol stocks at - 80°C.

To amplify phage particles after each round of selection 10 µl of bacteria that have been scraped and carefully resuspended were grown in 2xTY Amp at 37°C to an OD_{600nm} of 0.5. Then helper phages were added in a phage:bacteria ratio of 20:1 and left at 37°C for 30 min standing with occasional agitation. After the superinfection event the excess of helper phages was removed by centrifugation at 3500 rpm for 15 min 4°C and the bacterial pellet was resuspended in 2xTY Amp Kan and grown over night at 37°C.

To concentrate amplified phages this culture was spun at 3500 rpm for 15 min at 4°C. To the supernatant $1/5^{th}$ of its volume of PEG solution (20% PEG6000, 2.5M NaCl) was added and incubated on ice for 1 hour. After

centrifugation at 4000 rpm for 15 min 4°C pelleted phages are resuspended in 1.0 ml of PBS.

Phage ELISA assay

The overall procedure was very similar to the one previously reported (pag 49) with the following modifications:

- Coating: ligand (MNAC13 antibody) diluted at a final concentration of 10 mg/ml.
- Incubation: 50 µl per well of each amplified phage clone and 50 µl per well PBS (6% non fat dry milk).
- Secondary antibody: anti-geneIII antibody peroxidase conjugated diluted 1:5000.

Phage immunoscreening

Phage particles were spotted on nitrocellulose filter. After 5 min at 60°C the filters were blocked in PBS (5% non-fat dry milk) for 2 hours at room temperature with gentle agitation and then incubated over night at 4°C with MNAC13 monoclonal antibody (at a final concentration of 1 □g/ml in PBS, 5% non-fat dry milk). After 10 washes for 10 min in PBST at 4°C with gentle agitation the filters had been finally incubated with alkaline-phosphatase conjugated anti-mouse or anti-rat antibody for 4 hours at 4°C with gentle agitation, washed 10 times for 10 min in PBS (0.1% Tween 20); then the substrate solution (100mM Tris pH 9.6, 100mM NaCl, 5mM $MgCl_2$, 0.15 mg/ml BCIP, 0.3 mg/ml NBT) was added and left at room temperature.

Sequencing

To sequence phage clones that showed a strong signal both in phage immunoscreening and in phage ELISA assay corrisponding single bacterial colonies were grown over night at 37°C in 5ml 2xTY (Amp 10% Glucose) and phagmidic DNA was sequenced using as a primer oligonucleotide R156: 5' AAC CGA TAT ATT CGG TCG CTG AGG C 3' with the Epicentre Sequitherm Excel II kit and analyzed on a Likor 4000L automatic sequencer.

Results

For the first round of selection anti-TrkA MNAC13 monoclonal antibody was reacted with 10^{11} phage library particles in solution after affinity capture of the antibody protein G sepharose beads.

After amplification of selected phage particles, a second cycle was performed in the same conditions.

Then to avoid fortuitous selection of peptide sequences that specifically bind protein G, a round of panning with protein A sepharose was performed on the amplified eluted phages. A final round of selection was performed using again protein G. In the last two round of selection the concentration of the selector monoclonal antibody was decreased of an order of magnitude (from 100 to 10nM) to select only stronger binders and to lower background phages. The titers of unbound (U) and eluted (E) phages for each round of selection are reported in Table 1:

	I round		*II round*		*III round*		*IV round*	
	U	*E*	*U*	*E*	*U*	*E*	*U*	*E*
Anti-TrkA MNAC13	$2x10^9$	$3x10^5$	$7x10^5$	$5x10^7$	$5x10^6$	$2x10^6$	$4x10^5$	$7x10^7$

***Table 1**: Titers (pfu/ml) of unbound (U) and eluted (E) phages for each round of selection against anti-TrkA MNAC13*

The data concerning the second round of selection are consistent with a positive selection process. To confirm this hypothesis a preliminary phage immunoscreening was performed after amplification of single phage clones, obtaining a 66% of positive clones able to bind MNAC13 antibody as shown in panel A) of Figure 4.

As expected the third cycle, performed by protein A, showed a decrease in the number of retained phages due to the loss of binders that interact with protein G instead with the selector antibody. In the final round by protein G the amount of specific phage particles increased again by a positive selection process.

After these four rounds of selection retained phage particles were used again to infect bacteria. Single bacterial clones carrying the selected phagmid were grown in 96 well plates and superinfected by helper phages to produce and amplify each selected phage particles. In parallel each bacterial clones was also plated in presence of X-gal IPTG to control the real presence of a coding peptidic sequence cloned upstream pVIII. Indeed if no oligonucleotide has been inserted in the phagmidic vector or if it carries an in-frame stop codon (both of these possibility could have happened during the construction of the library due to the fact that the oligonucleotidic sequences were completely random), the vector is unable to drive not only expression of recombinant pVIII but also of the downstream β-galactosidase gene. So by color phenotype it is possible to distinguish aspecific background phages, that are always present for their selective advantage in packing during amplification due to the absence of a recombinant pVIII protein, whose folding is less stable than the wild type one. As shown in panel B) of Figure 4 after four rounds of selections the presence of these background phages is quite relevant considering that only 23% of eluted phages effectively carried a peptide. Then further experiments were concentrated only on phage particles obtained from blue colonies indicating that they contained an oligonucleotide insertion without any in-frame nonsense codon.

So 42 phage clones obtained from blue colonies were tested both by phage immunoblotting and phage ELISA assay (panel C) and D) of Figure 4). These two approaches are somehow complementary. Indeed

while in phage ELISA assay binding of phage particles can occur at the level of any portion of the selector antibody (as in the case of solution biopanning itself), in the case of immunoblotting theoretically only the correctly orientated antibody molecules can be detected and so peptides that interact with the antigen binding site should be selected.

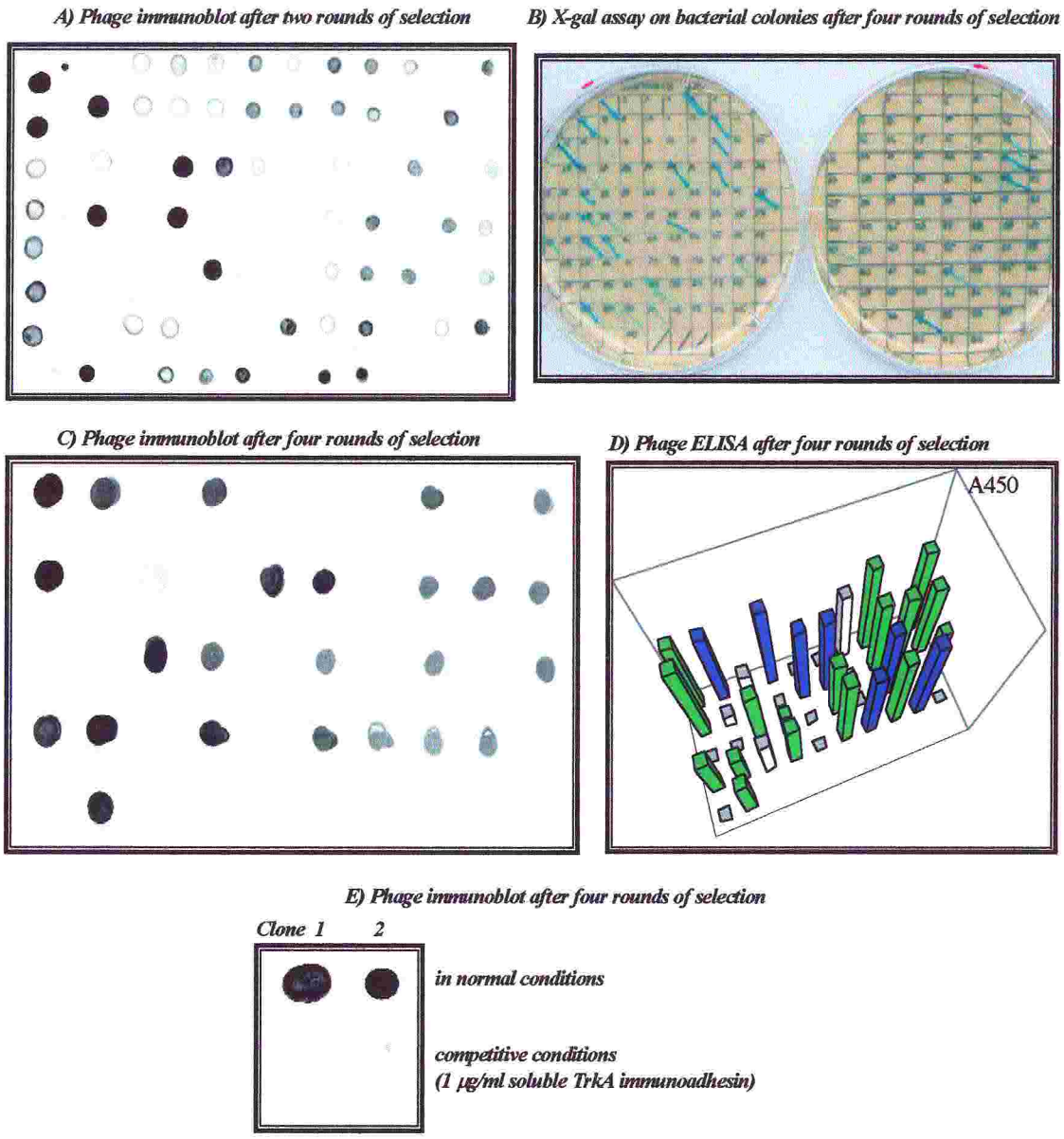


***Figure 4*: Phage immunoscreening and phage ELISA**

Positive phagmidic clones in both phage immunoblotting and phage ELISA (around 57% of eluted phages effectively carrying a peptide) were

analyzed by DNA sequencing, resulting in two main dominating phage populations (reported in Table 2) with the best affinity for anti-TrkA MNAC13 monoclonal antibody:

CLONES	NUCLEOTIDIC SEQUENCE	PEPTIDIC SEQUENCE
1, 8, 10, 11, 18, 19, 20, 23, 24, 26, 30, 31, 32, 34, 36, 38, 42	GAATTCGGTTTGCCCGTTAATGATGAAATACCGGAT	EFGLPVNDEIPD
2, 4, 15, 16, 28, 37, 39	GAATTCTACTATTGGCTCACATTATTCGGGGGGGAT	EFYYWLTLFGGD

Table 2: Phagmidic clones selected against anti-TrkA MNAC13

Finally as a further specificity control an immunoscreening, in which the binding between the selector antibody and both phagmidic populations was competed by the soluble antigen molecule (TrkA immunoadhesin), was performed. As shown in panel E) of Figure 4, one clone for each phagmidic population (Clone 1 representative for the most abundant population -around 71%- and Clone 2 for the less abundant one –around 29%-), was tested, obtaining a clearly positive competition and thus proving that both selected phages interact to the antigen binding site of the selector antibody.

Discussion

As a second approach to strictly map the shorter epitope recognized by the antiTrkA MNAC13 monoclonal antibody, a phagmidic library, exposing random nonapeptides on pVIII, was employed following a solution panning procedure.

In comparison to phagic libraries on pVIII, that are phenotypically and genotypically homogeneous, this phagmidic one is characterized by a higher complexity in the displayed random peptide sequences joined to the fact that this exposition vector is able to polymerize at high density. Unfortunately a disadvantage of this library is the fact that the number of exposed peptides is not homogeneous and so it is not possible to compare directly affinity of individual phage clones towards the selector antibody. However a lower number of displayed peptides should increase the importance of affinity versus avidity during the selection of binders.

In general the result of a phage library selection is a collection of peptides that shows affinity for the selector antibody and should be (even if not in all the cases) characterized by a common consensus amino acidic sequence. In the ideal cases of linear epitopes this consensus sequence share high level of similarity with a region of the antigen and so the site that is specifically recognized by the antibody binding site can be directly recognized and mapped. Anyway in many other situations no clear homology can be identified between the consensus sequence and the antigen one or no consensus sequence can be inferred from the selected peptides. A hypothesis to rationalize this case is that probably the recognized epitopes are not linear, but discontinuous or conformational and so many of the selected peptides can be mimotopes.

In particular in this case the result is even more complex, considering that the selection towards stronger binders was so heavy that two main

dominating sequences could be identified at the end of the four rounds of selection. Moreover it was striking the fact that no clear consensus could be identified between them. Regarding this result, it is worth noting that such a strong enrichment of these two apparently unrelated peptides happened quite early during the selection procedure. Indeed already after two rounds of selection the great majority of the binders carried one of these two peptides (data not shown), with the same relative abundance. These earlier stages of selection were characterized by the sporadic presence of other different clones, that have been fortuitously selected and so have been eliminated alternating rounds of panning between protein A sepharose and protein G sepharose.

Analyzing both aminoacidic sequences of the two peptides none of them seemed to closely resemble to any linear region on the extracellular domain of TrkA receptor, even if both of them were able to be competed by TrkA interaction to the antibody. These data seemed to suggest a conformational dependence of the epitope recognized by MNAC13 antibody, considering the lack of any evident consensus between the two selected peptides and also of any clear linear sequence homology with TrkA. It is quite interesting to underline that similar result were previously obtained for epitope mapping of αD11 antibody by phage display (Gonfloni, 1995). Indeed the selected phages did not reveal any evident consensus or resemblance to any linear region on NGF molecule. They were characterized by a consensus HD and by the presence of basic residues (R). The importance of these charged residues for the interaction was confirmed by results from epitope mapping by site-directed mutagenesis, that identified in residues K95 and L2 loop of NGF as necessary and sufficient for αD11 recognition.

Therefore focusing on the results reported in the previous chapter on epitope mapping by serial deletions, these data might be combined

turning out with a different possible outcome. Indeed Figure 5 shows a comparison between these peptides and the shorter region of TrkA extracellular domain that was still recognized by MNAC13 antibody, named deletion 2. Regarding the two peptides together the specific aminoacidic sequence (in bold), that have been fished out, also additional four residues of the amino terminal region of pVIII have been considered. In this quite rough alignment it was possible to identify a resemblance in the chemical environments of these regions, that was even more evident between TrkA sequence and the first peptide, representative of the predominant population of the selected peptides (Peptide 1).

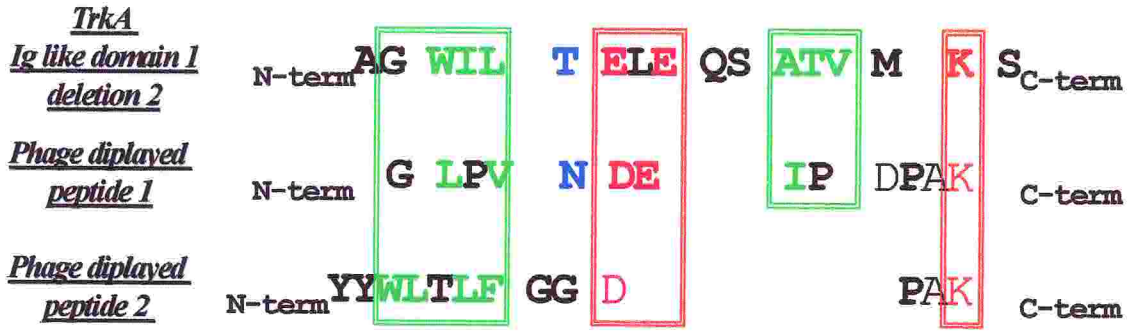


***Figure 5:** Sequence alignment among peptides fished out both by serial deletions (TrkA Ig like domain 1- deletion2) and by phage display (Phage displayed peptide 1 and 2)*

Figure 5 puts in evidence by colored boxes these aminoacidic sequences. Starting from the amino terminal it shows

- a first hydrophobic region (first green box), followed by an hydrophilic residue;
- a central acidic cluster (first red box);
- a second hydrophobic region (second green box),
- a conserved lysin (second red box) at the carboxi terminal end, that belongs to the first residues of pVIII.

Regarding the presence of Pro residues in the selected phage epitopes it is very likely that they are not directly involved in the interaction itself, considering that Pro residues are not hydrogen bond donors because they lack of the amide proton, replaced by a CH_2 group. Indeed the most convincing explanation of their importance is probably linked to their structural meaning. Indeed the backbone conformation is very restricted not only at the level of the Pros themselves (due to a quite strict range of possible ϕ dihedral angles around -65°), but also at the level of preceding residues (due to the bulk of the N-CH_2 group). Therefore their presence is often index of helix or β-sheet break. Moreover sequences, particularly rich in Pro (four or more), adopt the so-called polyproline II helix (Pawson, 1995). This extended structure with three residues per turn has been shown to be involved in many binding processes. A possible explanation can be that binding to a more rigid region cause a smaller loss in entropy than the interaction to a more flexible sequence. It is worth noting that this particular structure is well tolerated by the amino terminus of pVIII, which is itself a single a helix. Regarding this, although the inserted amino terminus has been shown by solid state NMR to be able to adopt an independent structure, epitopes, highly enriched after several cycles, share the propensity to show Pro in conserved positions (Cesareni, unpublished data). Thus Pro residues may provide to the epitope the structural characteristics involved both in the stabilization, the recognition and the interaction between the epitope and the antibody. From this point of view data obtained by phage display instead of identifying the precise linear epitope, recognized by MNAC13 antibody, can be informative in the identification of the chemical environment and of the crucial residues involved in MNAC13 binding and recognition.

References

Agterberg, M., Adriaanse, H., Tijhaar, E., Resink, A., Tommassen, J. (1989). Role of the cell surface-exposed regions of outer membrane protein PhoE of Escherichia coli K12 in the biogenesis of the protein. *Eur J Biochem*; **185**, 365-370

Burritt, J. B., Bond, C. W., Doss, K. W., Jesaitis, A. J. (1996). Filamentous phage display of oligopeptide libraries. *Anal Biochem.*; **238**, 1-13.

Cesareni, G. (1992). Peptide display on filamentous phage capsids. A new powerful tool to study protein-ligand interaction. *FEBS Lett.*; **307**, 66-70.

Charbit, A., Molla, A., Saurin, W., Hofnung, M. (1988). Versatility of a vector for expressing foreign polypeptides at the surface of gram-negative bacteria. *Gene*; **70**, 181-189.

Cwirla, S. E., Peters, E. A,, Barrett, R.W., Dower, W. J. (1990). Peptides on phage: a vast library of peptides for identifying ligands. *Proc Natl Acad Sci U S A*; **87**, 6378-6382.

Devlin, J. J., Panganiban, L. C., Devlin, P. E. (1990). Random peptide libraries: a source of specific protein binding molecules. *Science*; **249**, 404-406.

Felici, F., Castagnoli, L., Musacchio, A., Jappelli, R., Cesareni, G. (1991). Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *J Mol Biol.*; **222**, 301-310.

Freudl, R. (1989). Insertion of peptides into cell-surface-exposed areas of the Escherichia coli OmpA protein does not interfere with export and membrane assembly. *Gene*; **82**, 229-236.

Gonfoni, S. (1995). Recombinant antibodies as structural probes for neurotrophins.

Greenwood, J., Willis, A.E., Perham, R. N. (1991). Multiple display of foreign peptides on a filamentous bacteriophage. Peptides from Plasmodium falciparum circumsporozoite protein as antigens. *J Mol Biol.*; **220**, 821-827.

Sibille, P. & Strosberg, A. D. (1997). A FIV epitope defined by a phage peptide library screened with a monoclonal anti-FIV antibody. *Immunol Lett.;* **59**, 133-137.

Hedegaard, L. & Klemm, P. (1989). Type 1 fimbriae of Escherichia coli as carriers of heterologous antigenic sequences. *Gene*; **85**, 115-24.

Pawson, T. (1995). Protein modules and signalling networks. *Nature*; **373**, 573-580

Scott ,J. K., Smith. G. P. (1990). Searching for peptide ligands with an epitope library. *Science*; **249**, 386-390

Smith, G. P. & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.*; **217**, 228-257.

Stephen, C. W., Helminen, P., Lane, D. P. (1995). Characterisation of epitopes on human p53 using phage-displayed peptide libraries: insights into antibody-peptide interactions. *J Mol Biol* ; **248**, 58-78

PART 2:

STRUCTURAL CHARACTERIZATION of TWO NEUTRALIZING ANTIBODIES

CHAPTER 4: scFv expression, purification and characterization

Introduction

As mentioned before in order to study the structure of the antigen binding site of an antibody, in principle crystallisation can be restricted to variable domains (Fv), the smaller portion of the antibody which maintains the binding specificity and affinity of the entire antibody.

Instead of following the more classical approach of trying to crystallise Fab fragments obtained from intact native monoclonal antibodies by a proteolytic cleavage (Figure 1a), we decided to focus our attention first on the recombinant form of single chain Fv (scFv).

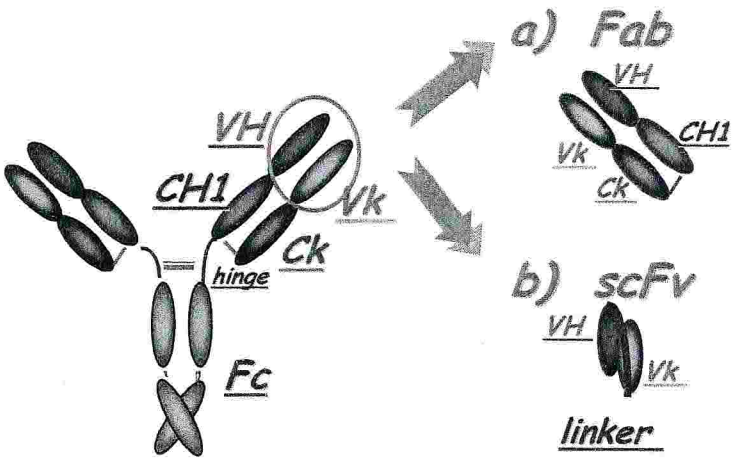


***Figure 1*: Antibody, Fab and scFv**

To obtain these small recombinant molecule the Fv fragment is cloned and expressed in heterologous systems, where the two variable chains are joined together by a short engineered linker. This flexible amino acidic sequence ranges from 14 to 25 amino acids and is able to prevent

dissociation of V_H and V_L bridging the approximately 35Å between the carboxy terminus of one domain and the amino of terminus the other (Figure 1b).

Considering that both MNAC13 and αD11 had been obtained in the recombinant form of scFv fragments (Cattaneo *et al.,* 1997), in first instance αD11 scFv had been chosen as the first target for crystallisation studies.

Materials and methods

Insect cell culture

Sf9 (*Spodoptera frugiperda*) cells were amplified at 27°C in TMNFH medium (Sigma) supplemented by 10% FCS, Glutammin (292 µg/ml), Gentamicin (25 µg/ml) and grown in flasks in monolayers.
High Five cells were grown in flasks in monolayers at 27°C in Excell medium supplemented by Glutammin (292 µg/ml), Gentamicin (25 µg/ml). Viral *AcNPV* (*Autographa californica* nuclear polyhedrosis virus) genome wilde type was included in Baculogold transfection Kit (Pharmingen Inc.).

Cloning of αD11 scFv in pAcGp67B

To express αD11 scFv by Baculovirus system, it had been cloned directionally in the transfer vector pAcGp67B: 5 µg of both the vector and a plasmid carrying αD11 scFv for eukaryotic expression were digested NcoI NotI over night at 37°C. After purification from agarose gel of the restriction band corresponding to the molecular weight of αD11 scFv, it had been incubated at 22°C for 2 hours in presence of ligase and of digested, dephosphorilated and purified transfer vector in 2:1 molar ratio. 1µl of ligation mix was used to transform electrocompetent DH5αF' by elettroporation. To isolate positive clones, a screening by PCR was performed directly on bacterial colonies. A second step of sub cloning was performed to insert at the 3' end of αD11 scFv a tag carrying the *c-myc* immuno-detection tag and a polyhistidine purification tag. This nucleotidic region had been amplified by PCR using the following oligonucleotides:

C-myc for: 5' GAC ACT AAG ATC TGA ATT CTT ATT AAT GGT GAT 3'

C-myc back : 5' CAA ATC TTG TGC GGC CGC C 3'

carrying restriction sites at the 5' ends for a directional cloning in frame a the 3' end of αD11 scFv sequence. After PCR amplification and removal of primers and enzyme by PCR purification kit (Qiagen), the PCR product was digested *NotI BglII* for 2 hours at 37°C. After another step of purification of the digested fragment by PCR purification kit, it had been directionally cloned *NotI BglII* in the vector pAcGP67B carrying αD11 scFv, digested with the same restriction enzymes, dephosphorilated and purified, by ligation reaction at 22°C for 2 hours in molar ratio 1:3 (vector:insert). After elettroporation of electrocompetent DH5αF' with 1µl of ligation mix, a screening by PCR was performed to isolate positive clones that have been sequenced using the primer pAcGP 5'CAG TCA CAC CAA GGC TTC 3' by pAcGP 5'CAG TCA CAC CAA GGC TTC 3' with the Epicentre Sequitherm Excel II kit and analyzed on a Likor 4000L automatic sequencer.

Moreover this second step of subcloning was performed in parallel to modify the transfer vector pAcGP67B inserting *c-myc* immuno-detection tag and a polyhistidine purification tag. Therefore a transfer vector suitable for scFv cloning and expression by Baculovirus system was constructed and named pAcGP67B+*myc/hisTAG.*

Changing αD11 scFv linker

The first PCR was performed to amplify separately αD11 V_L and V_H using respectively the following pairs of nucleotidic primers:

$V_k$for 5' GAA GTT ATG GTC GAC CCT CCG GAA CGT TTC AGC TCC AGC TTG G 3'

$V_k$back 5' AGC AAG CGG CGC GCA TGC CGA YAT TCA GAT GAY DCA GTC 3'

$V_H$for 5' GAT TGG TTT GCC GCT AGC TGA GGA GAC GGT GAC TGA GGT 3'

V_H back 5' TTA TCC TCG AGC GGT ACC CAG GTG CAG CTG AAG SAS TC 3'

After PCR amplification and removal of primers and enzyme by PCR purification kit (Qiagen), the PCR products were further on amplified separately using respectively the following two sets of longer primers:

V_kPTLfor 5' ACC GCT CGA GGA TAA CTT CGT ATA GTA TAC ATT ATA CGA AGT TAT GGT CGA CCC TCC 3'

V_kPT1back 5' CGC TGG ATT GTT ATT ACT CGC AGC AAG CGG CGC GCA TGC C 3'

V_kPT2back 5' TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC 3'

V_HPT1for 5' CCA GGC CCA GCA GTG GGT TTG GGA TTG GTT TGC CGC TA 3'

V_HPT2for 5' TGG TGA TGG TGA GTA CTA TCC AGG CCC AGC AGT GGG TTT G 3'

V_HPTLback 5' GGA GGG TCG ACC ATA ACT TCG TAT AAT GTA TAC TAT ACG AAG TTA TCC TCG AGC GGT A 3'

Then primers and enzyme were removed by PCR purification kit (Qiagen) to perform the final step of assembly: the two fragments obtained from the second PCR step were amplified together using a

second set of external set of primers (*V_kPT2back* and *V_HPT2for*). The resulting single PCR product, which carried αD11 V_L/V_H joined together by the new longer linker, was purified by PCR purification kit (Qiagen) and digested by *BssHII* and *NheI* restriction enzymes for the final directional cloning as well as the 5μg of the vector pDan5 for 2 hours at 37°C.

Then after vector dephosphorilation and another step of purification of the digested PCR fragment and vector by PCR purification kit, ligation reaction has been set up at 22°C for 2 hours in molar ratio 1:3 (vector:insert) to clone αD11 scFv with the new linker directionally *BssHII NheI*. After elettroporation of electrocompetent DH5αF' with 1μl of ligation mix, a screening by PCR was performed to isolate positive clones that have been sequenced using the primer SEQH: 5'TGG TGA TGG TGA TGG TGA GT 3'.

Then a further PCR was performed to insert at each end compatible cloning sites for pAcGp67B+*myc/hisTAG* using the following pair of oligonucleotidic primers:

VHBVfor 5' TTT TTG TGC GGC CGC TGA GGA GAC AGT GAC TGA 3'

VkBVback 5' TCC CGG GCC ATG GGA GAC ATC CAG ATG ACA CAG 3'

After purification the PCR product and 5μg of the vector 5 pAcGp67B+*myc/hisTAG* were digested using *NcoI/NotI* restriction enzymes; the plasmid was dephosphorilated and together with the digested PCR fragment have been purified and incubated in molar ratio 1:3 (vector:insert) at 22°C for 2 hours to perform a ligation reaction. After elettroporation of electrocompetent DH5αF' with 1μl of ligation

mix, a screening by PCR was performed to isolate positive clones that have been sequenced using the primer pAcGP.

Transfection and selection of αD11 scFv recombinant baculovirus

1 μg of transfer vector carrying αD11 scFv was co-transfected with 0.25 μg of linearized viral DNA, using calcium phosphate-mediated co-precipitation method modified for insect cells, according to the protocol suggested by Baculogold transfection kit.

After 6 days, cell surpernatant was recovered and plaque assay was performed to isolate single viral clones. Infecting with serial dilutions transfection supernatant a constant number of *Sf9* cells (2 10^6) and then overlaying them with agarose, it was possible to isolate single plaques after one week. Single viral clones were recovered from isolated plaques and used to infect 10^6 *Sf9* cells. After one more week it was possible to collect cell supernatant (named P0), containing viral particles whose title is around 10^8 PFU/ml.

Amplification of αD11 scFv recombinant baculovirus

The selected positive viral clone was amplified by two sequential infections to increase the amount of virus particles keeping constant the title.

For a first step of amplification 10^6 *Sf9* cells were infected with a low molteplicity (MOI=5) to avoid unwanted recombination between viral genomes: after 72 hours it was possible to collect viral particles with the cell supernatant (named P1); In a second step 3 10^6 *Sf9* cells were infected with the same low molteplicity: cell supernatant containing viral particles (named P2) was collected after 72 hours. Each infection was

incubated at 27 °C for 1 hours and then viral inoculum was removed and TNM-FH insect medium was added.

Protein expression by baculovirus system

To express recombinant protein 150 10^6 High five cells were infected with high molteplicity of infection (MOI=20) to maximize protein production using amplified viral particles contained in P2 cell supernatant.

Cell supernatant containing expressed recombinant proteins was collected after 72 hours to perform protein purification.

Purification of αD11 scFv by metal ion affinity chromatography

After collection and clarification by centrifugation at 4°C first for 10 min at 1000 rpm and then at 4000 rpm for 20 min, the pH of the supernatant was raised adding 1M NaOH to reach a value around 8 and NaCl was added to a final concentration of 300mM. Then it was mixed with 1ml of Ni NTA sepharose (QIAGEN) washed twice by PBS. After a batch incubation at 4°C for 2 hours rotating, the resin was spun for 10 min at 800 rpm and washed in batch with 50 ml of WB1 buffer (50mM Na Phosphate pH 8, 250mM NaCl). Then it was packed in a column and extensively washed with WB1 and in second instance with WB2 buffer (50mM Na Phosphate pH 8, 500mM NaCl). To remove aspecific binders the column was successively washed with low concentration of imidazole contained in EL1 buffer (WB2 buffer + 15mM imidazole). Finally elution took place by competition by high concentration of imidazole using EL2 buffer (WB2 buffer + 100mM imidazole).

Then all fractions were pooled, dialyzed over night at 4°C across a Spectra-Por Membrane 12/14K (Spectrum) against 10mM Tris pH 8.0, 50mM NaCl and concentrated using Centriprep 50K (Amicon)

concentrators. The exact amounts of purified protein were determined by Lowry Assay (Bio-Rad).

Western blot

Protein samples were loaded on discontinuous 10%SDS PAGE, separated by electrophoresis and transferred on nitrocellulose filters by electroblotting.

The filters were blocked by MPBS (PBS, 5% non fat dry milk) for 1 hour at room temperature and then incubated with the primary antibody MAb9E10, which recognizes *c-myc* tag, over night at 4°C.

After 3 washes by PBS the filters were incubated for one hour at room temperature with the secondary antibody, a goat anti-mouse peroxidase conjugated diluted 1:2000 in MPBS. Finally after extensive washings by PBST (PBS Tween20 0.1%) and by PBS, ECL system was employed to detect the presence of recombinant proteins.

ELISA assay and Competitive ELISA assay

The overall procedure of ELISA assay was very similar to the one previously reported (pag 49) especially concerning the washing procedures and details about blocking step, with the following modifications:

- Coating: NGF diluted at a final concentration of 10 mg/ml.
- Incubation: 50 µl per well of PBS (6% non fat dry milk) and 50 µl of cell supernatants.
- Primary antibody: MAb9E10 diluted 1:1000.
- Secondary antibody: anti-mouse antibody peroxidase (Dako) conjugated diluted 1:2000

As a positive control TrkA immunoadhesin (TrkA ECD Camel), expressed by Baculovirus system, was employed.

On the other hand the overall outline of competitive ELISA assay consisted in the same protocol of ELISA assay following the following procedure. First of all for NGF coating titration serial dilutions of NGF were tested in the coating step and the incubated with 0.1μg/ml of purified αD11 scFv diluted in MPBS.

In the following ELISA assay NGF coating was performed at a final concentration of 3.5μg/ml and serial dilutions of purified αD11 scFv in MPBS were incubated. Finally also in the competitive ELISA assay NGF coating was performed at a final concentration of 3.5μg/ml. After blocking step, two series of wells were incubated with 100μl per well of serial dilutions of αD11 scFv, treated with subtilisin, in MPBS for 2 hours at room temperature. Then after the same washing procedure previously reported the two series of wells have been incubated with 100μl per well of purified αD11 scFv at final concentrations respectively of 50 and 100 ng/ml in MPBS for 2 hours at room temperature.

Results

αD11 scFv expression by baculovirus system and purification

In order to express and purify αD11 scFv suitable for crystallization studies, Baculovirus expression system was employed considering the quite high level of expression that can be obtained. The overall procedure is shown in Figure 2.

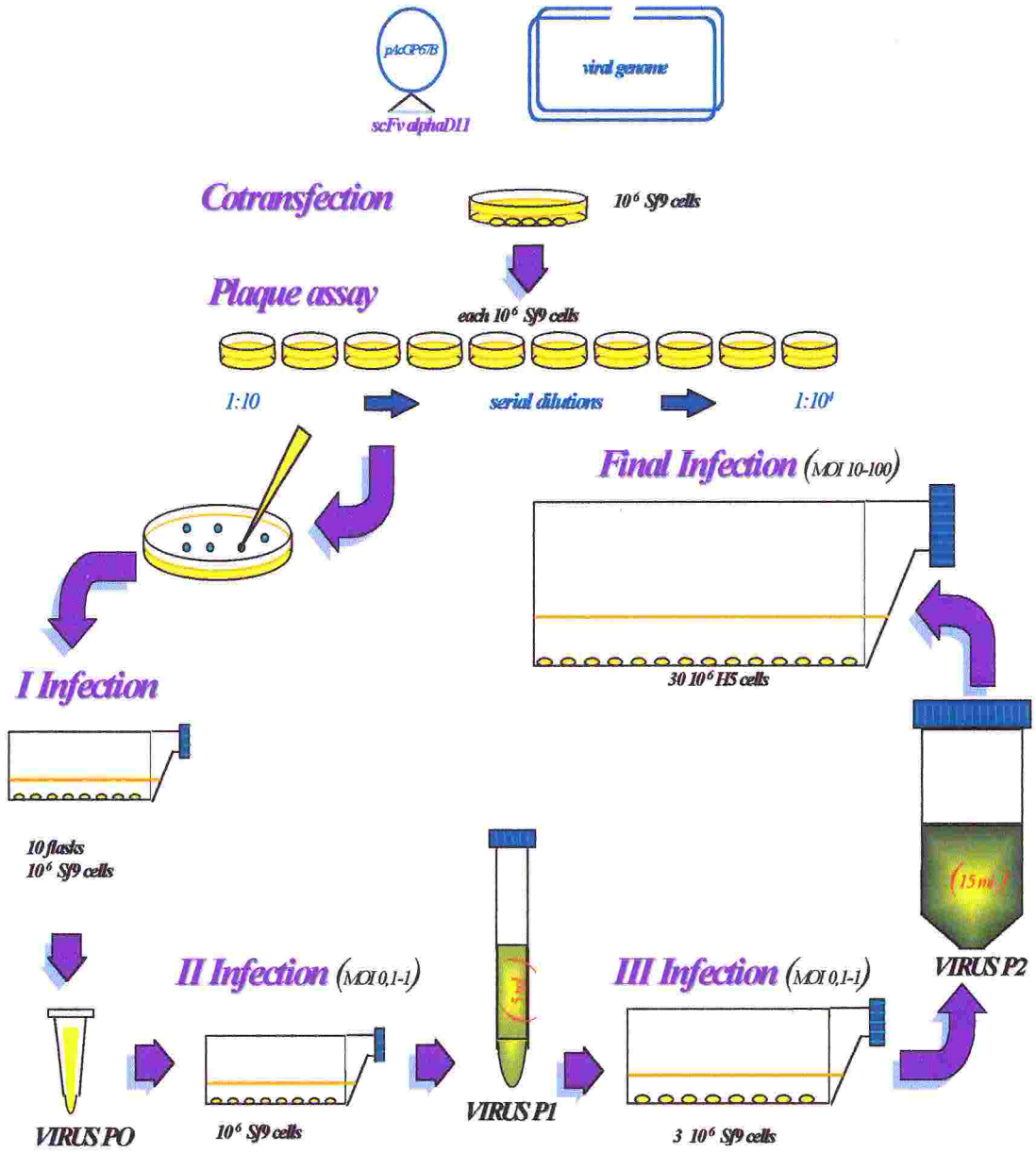


***Figure 2:* Baculovirus expression system**

αD11 scFv coding sequence has been cloned by restriction in the transfer vector pAcGP67B which carries, upstream of the polylinker, the secretion signal of a viral glycoprotein GP67 that is efficiently recognized by insect cells.

The carboxy terminal tag of αD11 scFv, that was lost during this sub-cloning step, was re-amplified by PCR and subcloned at the 3' end of αD11 scFv coding sequence. In this way it was possible to detect protein expression through the presence of *c-myc* immunodetection tag and to purify it by metal ion affinity chromatography by poly histidine tag.

To produce recombinant virus genome transfer vector carrying αD11 scFv was co-transfected with a modified type of baculovirus genome, which does not code for viable virus, using Baculogold System. This system relies on insect cell machinery that is able to mediate recombination between homologous sequences of transfer vector and viral DNA reconstituting viable recombinant viruses.

Therefore after 6 days cell supernatant was recovered and a Western blot was performed to detect the presence of the protein and as a consequence of the recombinant virus both in the supernatant, considering that the recombinant protein had a secretion signal at its amino terminal, and in cell lysates (Figure 3).

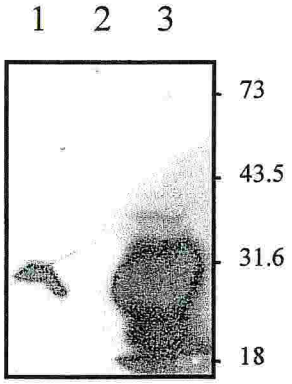


Figure 3: Western blot anti *c-myc* tag. Lane 1: supernatant of transfected cells; Lane 2: molecular weights ; Lane 3: transfected cell lysates

Then plaque assay was performed to isolate single recombinant viruses that have been picked up from single plaques and used to infect *Sf9* cells. After one more week cell supernatants (named P0) were collected and analyzed together with cell lysates by Western blot (Figure 4) to verify the presence of the protein.

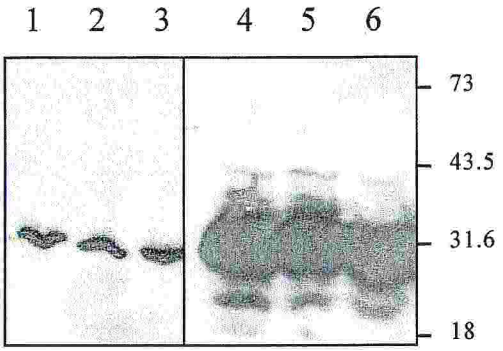


Figure 4: Western blot anti *c-myc* tag. Lane 1-3: cell supernatants (clones 3 and 4) soluble fractions; Lane 4-6: cell lysates (clones 9 and 10) soluble fractions.

Moreover to verify binding activity of αD11 scFv all the supernatants were also tested by ELISA resulting in strong positive signals >1.5 (data not shown).

Then one of the positive viral clones was amplified to increase virus titer and amounts as reported previously; finally it was used to express αD11 scFv infecting High Five cells, which have been shown to be more efficient in protein expression and secretion.

Finally it was possible to purify αD11 scFv using metal ion affinity chromatography (IMAC) and eluted by competition using high concentration of imidazole. All fractions across the elution peak were collected and assessed for homogeneity by Coomassie blue staining after separation by SDS-PAGE electrophoresis (Figure 5). After dialysis and concentration it was possible to obtain samples in a concentration range between 4 and 7 mg/ml with a purity higher than 99%, as controlled by SDS-PAGE electrophoresis.

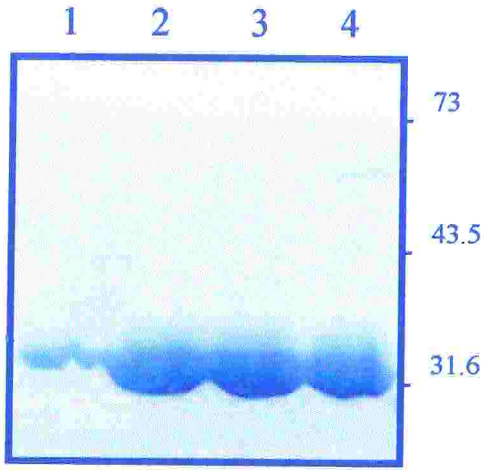


Figure 5: SDS PAGE and Coomassie staining of fractions across elution peak in αD11 scFv purification.

αD11 scFv characterization by size exclusion chromatography

Even if purified αD11 scFv samples seemed to be suitable for crystallization in terms of concentration and purity, not promising results were obtained from a preliminary array of crystallization trials set up at 16°C by the hanging-drop method using both the sparse-matrix sampling method (Jancarik & Kim, 1991), with Crystal Screen I and II (Hampton Research; Laguna Niguel, CA, USA), and Low Ionic Strength Screen (Hampton Research; Laguna Niguel, CA, USA), a grid screen method based on PEG3350.

Considering these preliminary negative results and what reported in a review (Essig *et al.*, 1993) that underlies that the principal obstacle to crystallization is reversible aggregation of the intact scFv proteins upon concentration to conditions suitable for crystallisation, gel filtration chromatography was performed to analyse αD11 scFv aggregation state. Therefore size exclusion chromatography was performed on a Superdex G75 column (Pharmacia) using an FPLC system (Pharmacia). After an extensive equilibration with Tris HCl 100mM pH 8.0, NaCl 150mM and calibration of the column by a cocktail of molecular weight markers 200 µl of sample were loaded. As it can be noticed analyzing the resulting gel

filtration profile shown in Figure 6 a quite low fraction of αD11 scFv is present in a momomeric state. Moreover it is worth noting that the greater fraction of αD11 scFv population was composed by monomers or dimers but by higher multimeric complexes. Considering these results it can be concluded that it is very likely that the high level of heterogeneity in αD11 scFv aggregation state adversely affects αD11 scFv crystallisation.

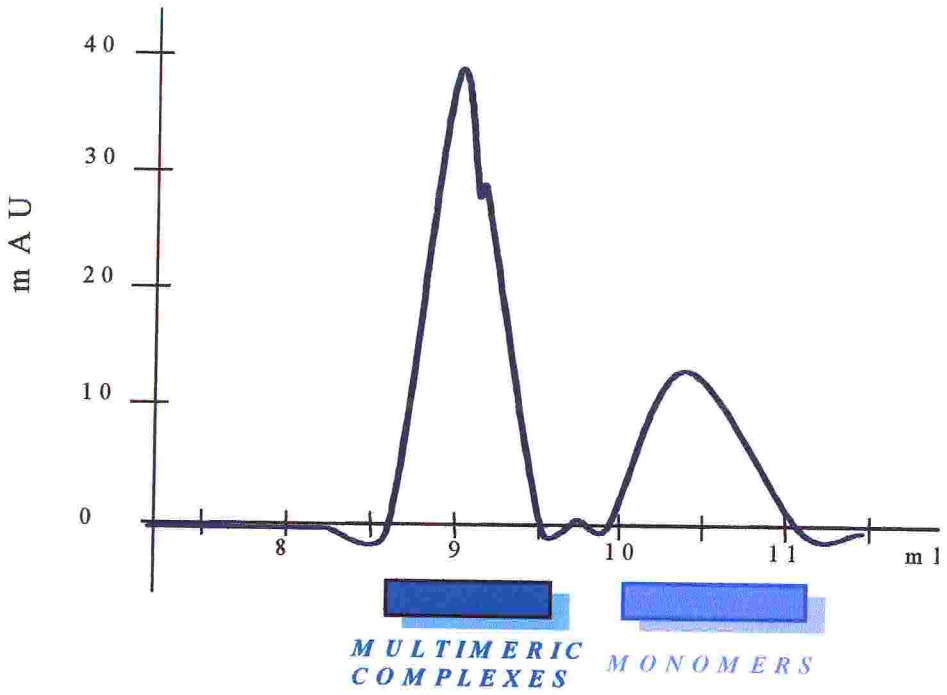


Figure 6: ***αD11 scFv gel filtration profile***

To overcome the problem of aggregation there are several approaches reported in the literature trying to obtain an homogeneous population of scFv molecules: two of them have been applied to improve αD11 scFv crystallization chances.

αD11 scFv linker clipping by subtilisin

In the literature (Essig *et al,* 1993) proteolytic clipping of the scFv linker has been reported to be a generally useful technique for the crystallization of these molecules, preventing the formation of multivalent species at high protein concentrations, and reducing steric strain from the linker on the V_H/ V_L interface. To verify this working hypothesis, αD11 scFv linker (whose aminoacidic sequence is GGGGSGGGGSGGGGS) was subjected to proteolytic cleavage with subtilisin in a molar ratio $1:10^4$ with the substrate at 20°C for 4 and 8 hours, monitoring the process by SDS PAGE electrophoresis and Coomassie staining (Figure 7).

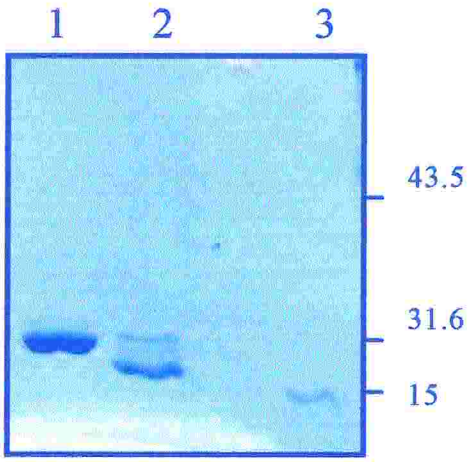


Figure 7: SDS PAGE and Coomassie staining of αD11 scFv subtilisin digestion. Lane 1: undigested αD11 scFv; Lane 2: αD11 scFv incubated with subtilisin for 4 hours; Lane 3: αD11 scFv incubated with subtilisin for 8 hours.

Moreover size exclusion chromatography was performed again as previously described to compare the profile before and after linker clipping.

As reported in Figure 8 after 6 hours of incubation with subtilisin αD11 scFv population was composed essentially by monomers and Fv fragments, with a very low fraction of dimers while higher aggregation complexes were not present anymore.

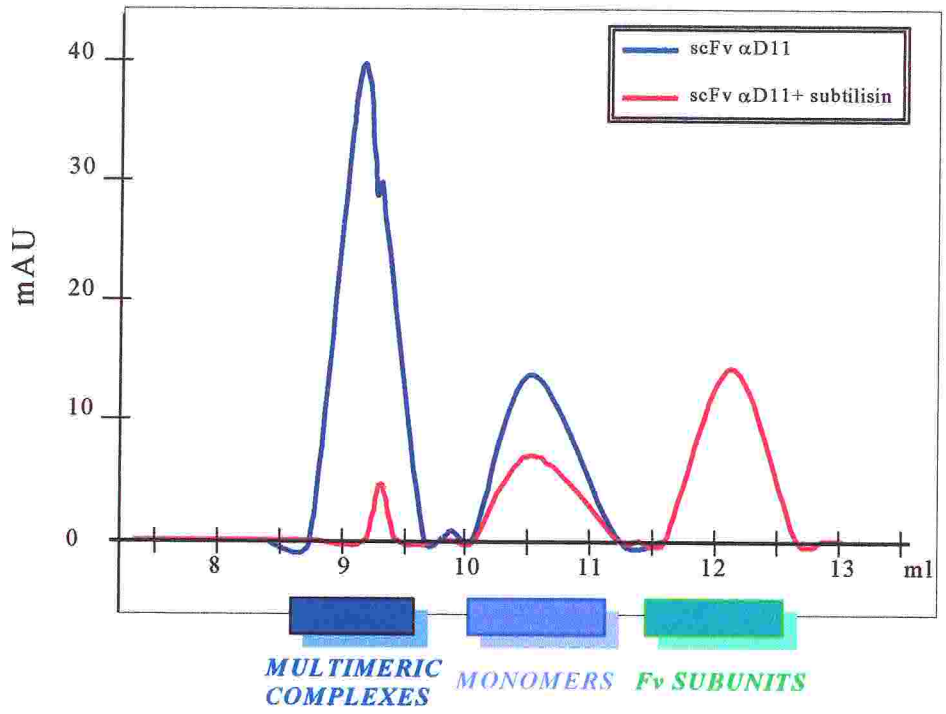


Figure 8: ***Comparison of αD11 scFv gel filtration profile before and after linker clipping***

To check that linker clipping did not affect biological activity of αD11 scFv, ELISA assays were performed: not surprisingly direct assay failed to detect any significant interaction. This result was consistent with the fact that Western Blotting with the *c-myc* detection antibody (data not shown) did not detect the V_L domain even if V_H and V_L of the Fv could be clearly seen by SDS-PAGE. Taken together these data indicated a consistent removal of the *c-myc*/his tag extension by protease treatment. Therefore competitive ELISA was performed to verify that NGF binding activity was conserved in Fv fragments. Figure 9 shows preliminary data used later to set up competitive ELISA assay. In particular from NGF coating titration shown in Figure 9A the value corresponding to the 66% of the maximum was estimated around 3.5 μg/ml. Then using such NGF concentration for coating, αD11 scFv was titrated in order to find out the concentration value corresponding to 50% of binding (Figure 9B).

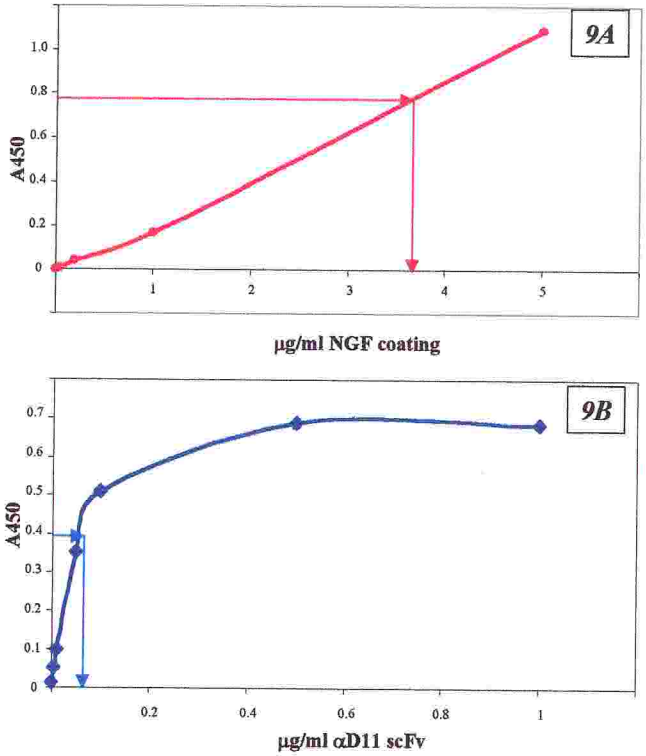


Figure 9A: NGF coating titration with αD11 scFv incubation (0.2 µg/ml)
9B: αD11 scFv titration with NGF coating (3.5 µg/ml)

These conditions were quite far from a saturation situation and moreover NGF molecules (and so αD11 binding sites) were limiting.

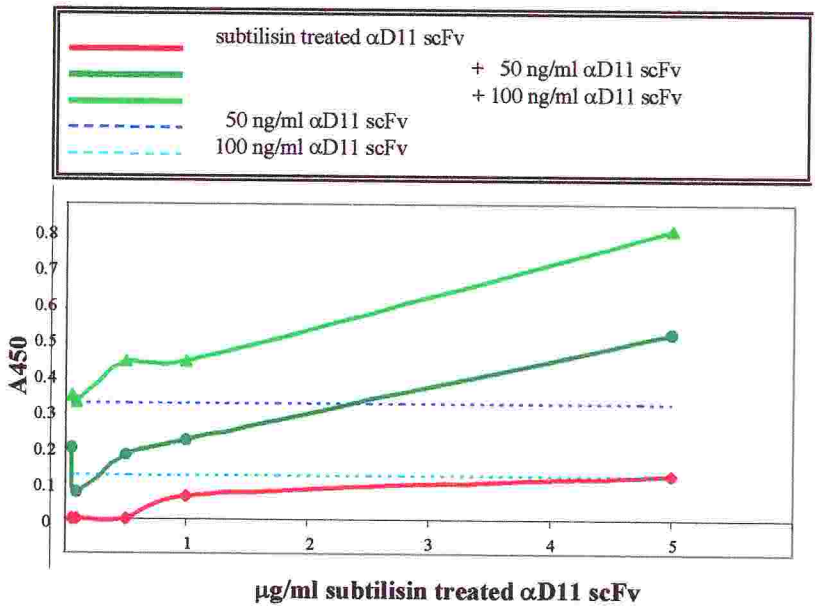


Figure 9C: Competitive ELISA assay

Anyway results, obtained in such conditions from a competitive ELISA Assay and summarized in Figure 9C, were quite unexpected.

Indeed after incubation with increasing concentrations of subtilisin cleaved αD11 scFv, adding constant concentrations of untreated αD11 scFv two opposite results could be hypothesized. If Fv fragments did not retain any binding activity towards NGF, the signal should have been constant regardless Fv concentration and quite similar to the one indicated by dashed lines in Figure 9C. On the contrary if Fv fragments interacted with NGF, they should have progressively saturated accessible NGF sites and so should have competed with untreated αD11 scFv for binding resulting in a progressive lowering of the signal as well as increasing Fv fragments concentration. But as shown in Figure 9C the signal increased in a quite linear way with Fv fragments concentration: this result not only demonstrated that subtilisin treated αD11 scFv maintain NGF binding activity, but also suggested that it somehow increased αD11 scFv avidity towards NGF.

Considering these data subtilisin treatment was employed to try to solve the problem of micro-heterogeneity in αD11 scFv samples, but unfortunately it was not feasible to purify Fv fragments from αD11 scFv digested samples. Indeed if the time of subtilisin treatment was over 6 hours, secondary sites of cleavage reduced of a great deal protein concentration; on the other hand by this time of incubation the fraction of Fv fragments population was too low. So as previously reported (Hoedemaeker *et al.*, 1997) subtilisin was added directly inside the hanging drops together with the αD11 scFv sample (in a molar ratio $1:10^4$ with the substrate) setting up a vast array of crystallization trials, similar to the preliminary one. Also in this case no positive results could be obtained.

αD11 scFv linker change

An alternative method to overcome the problem of aggregation in scFv crystallization has been reported (Essig *et al,* 1993) to be purification of dimeric complexes. But analyzing αD11 scFv size exclusion profile, the fraction of dimers to purify was too low especially in comparison to multimeric complexes population for this approach to be feasible. In order to try to force scFv equilibrium towards monomers and dimers, decreasing considerably multimeric complexes fractions, a possible approach is to increase the length of the linker sequence. So this previous linker 15 aminoacids long (GGG GSG GGG SGG GGS) was replaced by a longer (20 aminoacids long) one, whose sequence is SGGSTITSYNVYYTKLSSSGT. This new longer linker has been successfully used in the expression of many scFv, giving rise to very high fractions of mononeric form.

The overall procedure performed to change αD11 scFv linker relies on a three step PCR and is summarized in Figure 10. The first PCR was performed to amplify separately the two variable chains choosing the pairs most similar to αD11 V_L/V_H 3' and 5' ends from a pool of oligonucleotidic primers representative of mouse V_L/V_H 3' and 5' ends. In the second PCR step these amplified fragments were further on amplified using two sets of longer primers:

the external ones (*V_kPT1back* and *V_HPT1for*) to insert at each end restriction sites (*BssHII, NheI*) for the final directional cloning;

the internal ones (*V_kPTLfor* and *V_HPTLback*) carrying the two complementary strands of the new linker sequence.

Finally in the assembly step the two fragments obtained from the second PCR step were amplified together using a second set of external primers

($V_kPT2back$ and $V_HPT2for$). The resulting single fragment carried αD11 V_L/V_H joined together by the new longer linker with *BssHII* and *NheI* restriction sites respectively at 5' and 3' end for the final directional cloning in pDan5. After screening by PCR to isolate positive clones and sequencing to control that no point mutations have been introduced by PCR steps, a further PCR was performed to insert at each end compatible cloning sites for pAcGp67B+*myc/hisTAG*. After directional cloning *NcoI/NotI*, selection and sequencing of positive

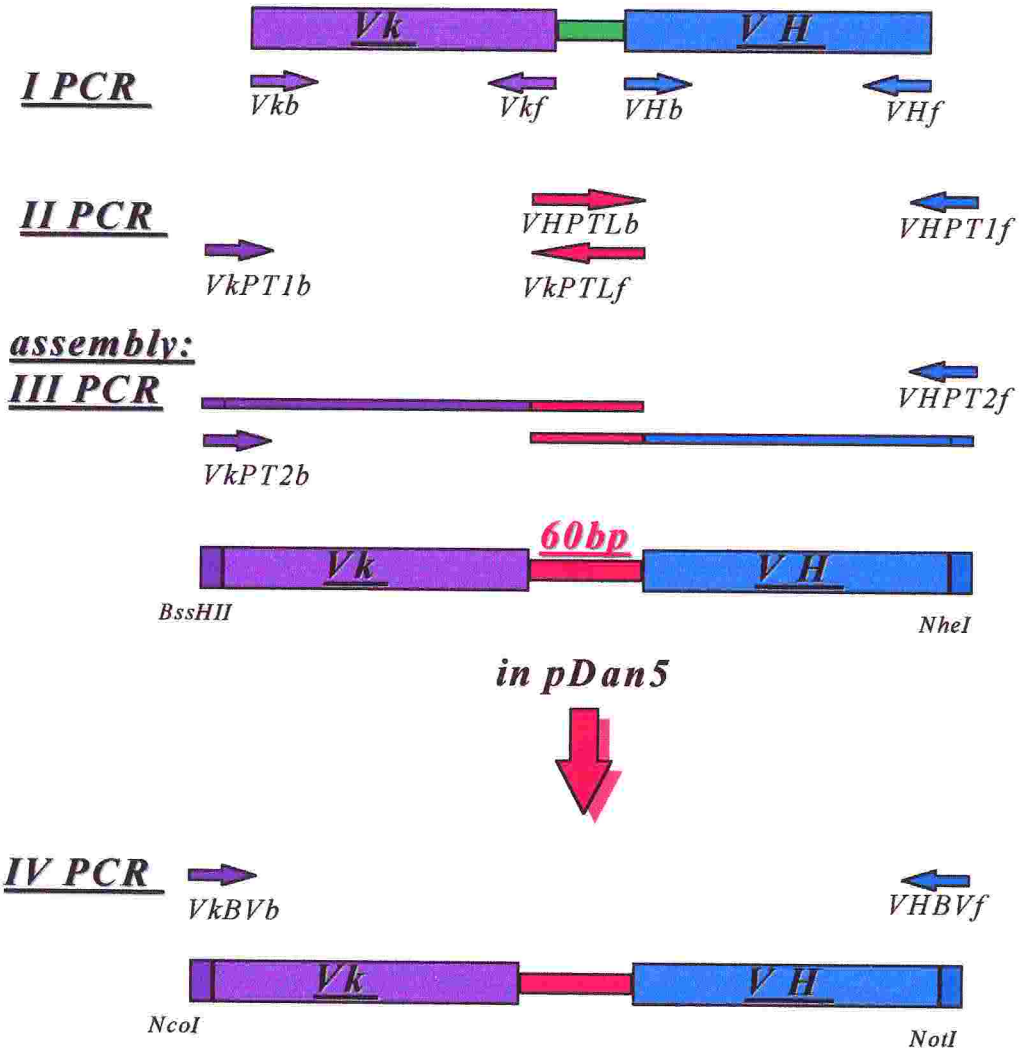


Figure 10: scFv linker change procedure

clones, recombinant Baculovirus were obtained and used to express αD11 scFv longer linker as previously described. Then αD11 scFv longer linker was purified by ion metal affinity chromatography and the purified sample was analyzed by size exclusion chromatography to check if expected variation in αD11 scFv aggregation state had indeed taken place. Figure 11 shows the comparison of gel filtration profiles of αD11 scFv carrying longer and shorter linker. Even if an improvement in αD11 scFv solubility could be achieved considering the increased fractions of monomeric and dimeric complexes, however multimeric complexes population was still too high.

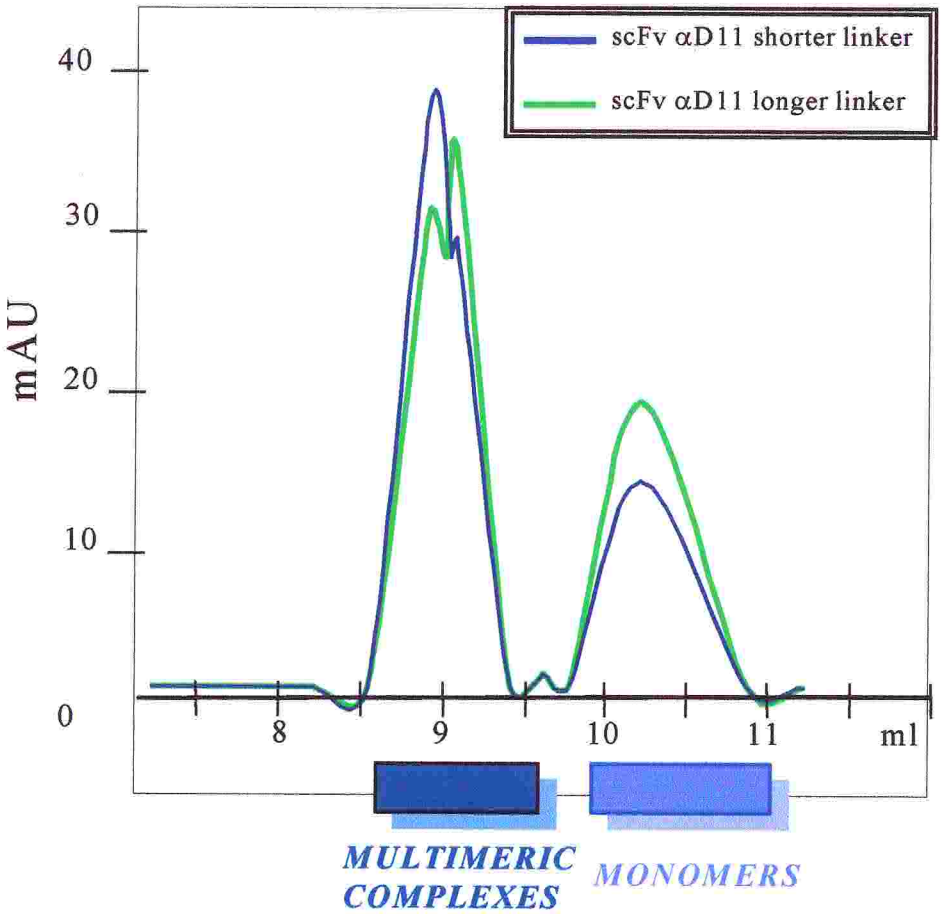


Figure 11: ***Comparison of αD11 scFv gel filtration profile before and after linker change***

Therefore also in such conditions the approach of purifying enough dimeric complexes for crystallization studies was not feasible.

Discussion

Baculovirus system was chosen to express αD11 scFv considering that it guarantees high level of expression of recombinant protein and so allows purifying suitable amounts to set up crystallization trials. Indeed this system is based on highly expressed very late gene promoters of *Autographa californica* Nuclear Polyhedrosys Virus. In particular polyhedrin is the gene of choice to be engineered as an expression vector because it has been shown to be non-essential for viral replication in culture and moreover its promoter is strong and strictly temporally regulated. Considering the large size of viral DNA, an indirect approach is necessary to modify viral genome. This can be done indirectly cloning αD11 scFv nucleotidic sequence in a so called transfer vector in which portions of viral genome spanning the polyhedrin gene have been inserted and manipulated, maintaining promoter and UTR that are important for high level of expression. Then to transfer the gene of interest to viral genome, recombinant transfer vector and manipulated viral DNA are co-transfected in insect cells: inside these host cells the sequences flanking the gene of interest recombine with the homologous sequences in viral genome reconstituting viable recombinant baculoviruses. After isolation and amplification of single clones of recombinant baculovirus it was possible to set up an expression and purification system leading to high amount of αD11 scFv with a high level of purity, suitable in principle for crystallization studies.

Nevertheless preliminary screening of crystallization conditions did not give any positive indication. A possible explanation of these

disappointing results is the fact that in scFv crystallization the principal obstacle has been reported (Essig *et al,* 1993) to be reversible aggregation of the intact scFv proteins upon concentration to conditions suitable for crystallisation. Indeed usually one of the fundamental requirements for crystallisation studies is not only high purity bur also homogeneity in the population of target molecules and so this multimeric equilibrium adversely affects scFv crystallisation. Considering that aggregation seems to be primarily linker dependent, it is unlikely that it is due to aspecific interactions between scFv molecules. On the contrary aggregation may result from specific interactions, in which linker joined V_H and V_L associate not with each other, but with the appropriate domain of neighbouring molecules. Therefore the interface between V_H and V_L would still be a standard one even if these two domains were not connected to each other by a linker. Moreover this kind of interaction would result not only in dimer formation, but also in higher multimeric complexes (Figure 12).

To verify this explanation αD11 scFv gel filtration profile was analyzed. Therefore even if αD11 scFv obtained by metal ion affinity chromatography was shown to be homogeneous by SDS-PAGE, it was demonstrated not only to exist in distinct aggregation states but also to be particularly prone to equilibrate into different multimeric forms.

Two main approaches to overcome this obstacle have been reported.

A first approach that has been followed is linker clipping, using subtilisin, a proteolytic enzyme that preferentially cleaves at the level of the linker of scFv. Linker clipping obviously destroys the multimeric nature of any complexes present, since all V_H/V_L interfaces are believed to be identical, regardless of multimerization state. Indeed the result of size exclusion

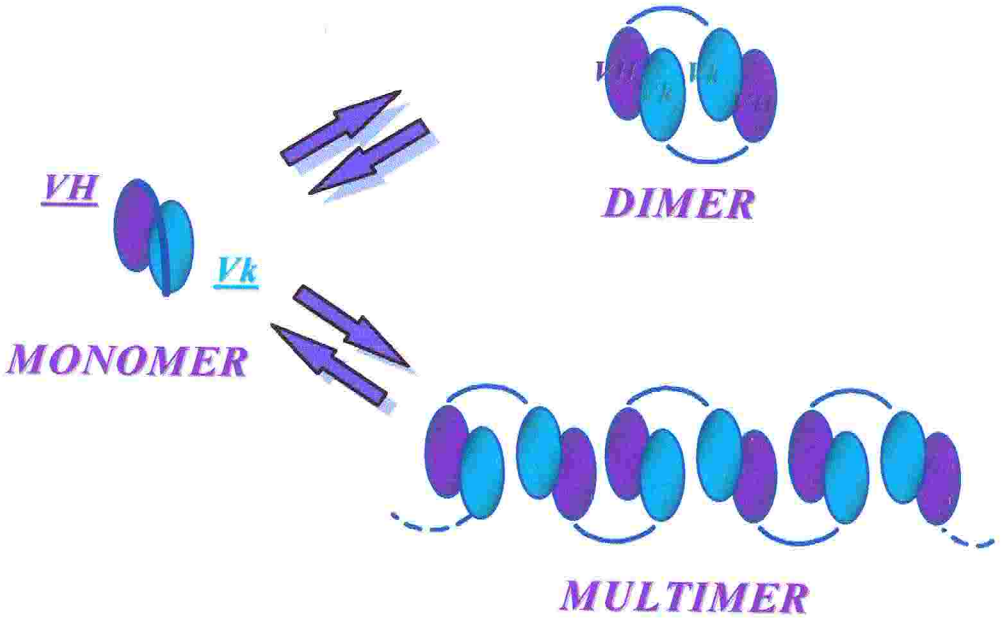


Figure 12 : Hypothesis for scFv aggregate formation

chromatography shows a profile consisting only in monomers and Fv fragments, with a very low contamination of dimeric complexes. Considering this result together with competitive ELISA Assay data, a second array of crystallization trials was set up adding the enzyme directly to the drops, but no improvements in the outcome of crystallization trials could be seen.

The alternative method relies in the purification of dimeric complexes. However considering αD11 scFv size exclusion profile, the fraction of dimers to purify was quite low especially in comparison to multimeric complexes population. It is worth noting that in general scFv aggregation state is not dependent only on the peculiar combination of V_L/V_H but also to some extent on type and length of linker sequence connecting the two variable domains. Indeed increasing the length of the linker sequence it has been shown (Desplancq *et al.*, 1994) to be possible to force scFv equilibrium towards monomers and dimers, decreasing considerably multimeric complexes fractions. By a three steps PCR procedure it was

possible to insert a longer linker between V_L and V_H αD11 scFv. Anyway comparing gel filtration profiles of αD11 scFv with longer and shorter linker, multimeric complexes fraction was not only still present but also too high to set a purification system for dimers, even if an improvement in both monomeric and dimeric populations could be achieved.

On the other hand from preliminary experiments (data not shown) neither MNAC13 scFv resulted to be suitable for crystallisation studies. Indeed although it was possible to obtain good level of expression both in bacteria and by baculovirus system, the output from the purification step (based on metal affinity chromatography) was extremely low. This problem was probably due to a particular folding of the carboxy-terminal region of MNAC13 scFv, resulting in the partial masking of the his-tag.

References

Desplancq, D., King, D. J., Lawson, A. D., Mountain, A. (1994). Multimerization behaviour of single chain Fv variants for the tumour-binding antibody B72.3. *Protein Eng*; **7**, 1027-1033.

Essig, N.Z., Wood, J. F., Howard, A. J., Raag, R., Whitlow, M. (1993) Crystallization of single-chain Fv proteins. *J Mol Biol*; **234**, 897-901.

Hoedemaeker, F. J., Signorelli, T., Johns, K., Kuntz, D. A., and Rose, D. R. (1997).

J Biol Chem; **272**, 29784-29789.

CHAPTER 5: Fabs Purification

Considering that the recombinant forms of both the two antibodies were shown to be unsuitable for crystallisation studies, a more classical approach was employed to try to obtain structural information on their antigen binding sites. Indeed as widely reported in the literature the more common way to solve the structure of an antibody is to crystallise its Fab fragments.

So to achieve this goal, antibodies can be fragmented by controlled partial digestion with several proteolytic enzymes. In particular papain recognizes its principal sites that are found on the amino-terminal side of the disulfide bonds holding the two chains together. Therefore cleavage with papain releases two antigen binding domains and one Fc fragment as shown in Figure 1.

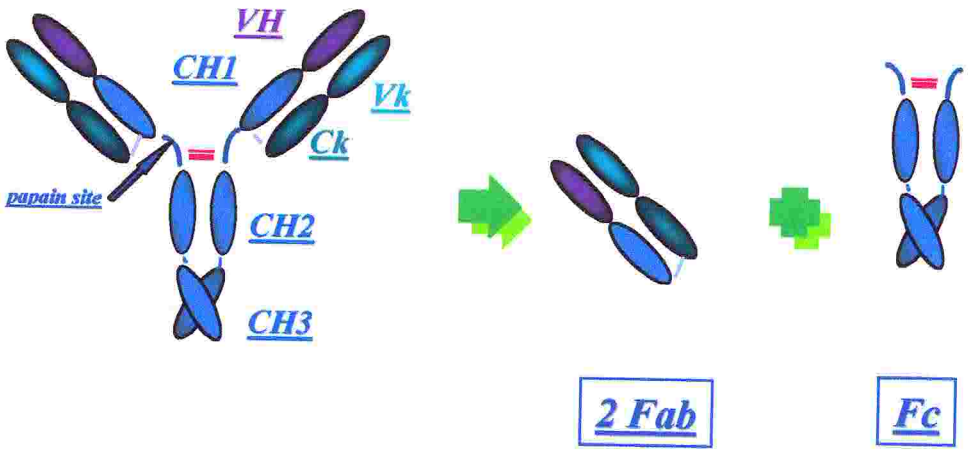


***Figure 1*: Papain cleavage**

Therefore it is possible to obtain suitable samples for structural studies starting from monoclonal antibodies expressed by *in vitro* tissue culture of hybridomas, then purified by salting out combined with antibody

affinity chromatography. Once proteolytically digested by papain, there are several biochemical methods to purify Fab fragments from intact immunoglobulins and Fc portions. Moreover these procedures can be scaled up in order to obtain high amounts of protein samples with a high level of purity and homogeneity that are the main requirements for crystallization experiments.

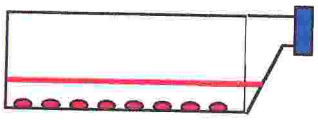
Biochemical methods and results

Purification of Mabs

The MNAC13 IgG1 and αD11 IgG2a immunoglobulin were obtained from hybridomas by *in vitro* tissue culture. After extract clarification by centrifugation for 10 min at 1000 rpm both antibodies were purified from supernatant, by salting out, slowly adding 29% ammonium sulfate at 4°C. After an incubation over night at 4°C the supernatants are spun twice at 5000 rpm at 4°C for 5 min: pellets are resuspended in 1/20 of the starting volume of PBS and then dialyzed over night at 4°C across a Spectra-Por Membrane 12/14K (Spectrum) against PBS.

Then to purify both immunoglobulin antibody affinity chromatography was performed using a Protein G Sepharose column (Pharmacia) considering that mouse IgG_1 and rat IgG2a show higher affinity for protein G than protein A. After equilibration of 1 ml of Protein G Sepharose with Tris HCl 10mM pH 7.2, NaCl 150mM sample are first spun at 4000 rpm at 4°C for 10 min and then loaded on le column by a peristaltic pump. After extensive washes with the equilibration buffer pure Mab were eluted by low pH buffer (10mM HCl) and suddenly neutralized with small volumes of Tris HCl 1.5M pH 8.8. All Mab fractions were pooled, concentrated using Centriprep 50K (Amicon) concentrators and dialyzed over night at 4°C across a Spectra-Por Membrane 12/14K (Spectrum) against 10mM sodium phosphate pH 7.0, 20mM EDTA. The overall procedure is summarized in Figure 2.

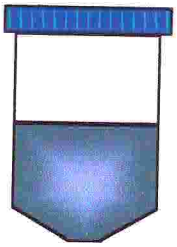
Expression by hybridomas



by in vitro tissue culture
supernatant collected

1. Salting out procedures

based on selective precipitation
by ***ammonium sulphate 28.75%***



2. Antibody affinity chromatography

based on ***protein G Sepharose***
acid elution and neutralization
15mg pure Mab
from 1l of culture supernatant

***Figure 2**: Mab purification procedure*

Papain cleavage

To obtain Fab fragments the proteolytic enzyme papain was employed because it cleaves intact immunoglobulins at the level of the hinge between C_{H1} and C_{H2}, as shown in Figure 1.

To find out the best digestion conditions an enzymatic time course was performed, treating with immobilized papain (Pierce) (at an enzyme: substrate ratio of 1:15) in 13mM Cys at 37°C shaking in a waterbath.

The time course of the proteolytic cleavage was monitored by examining aliquots by separation on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie staining, as reported in Figure 3.

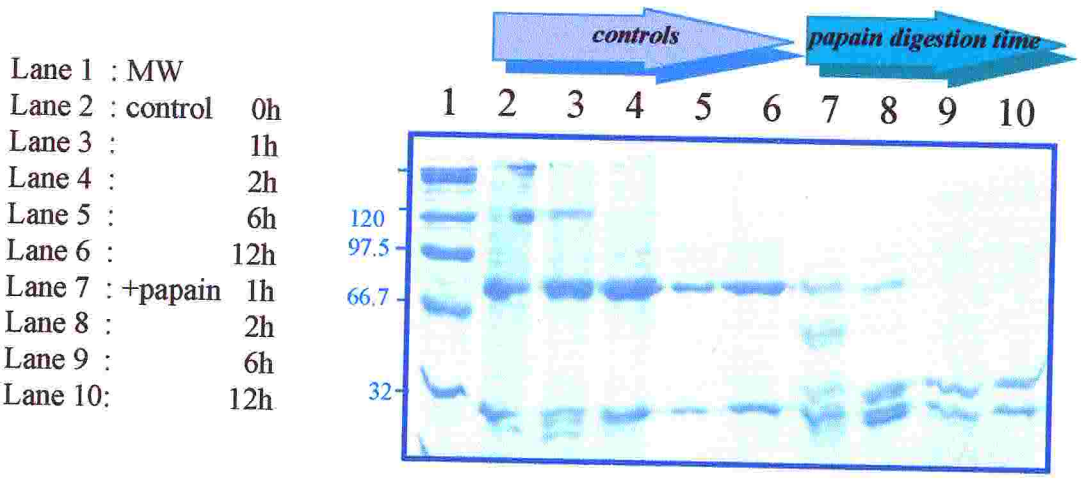


Figure 3: Papain cleavage time course

Considering that after 6 hours of incubation the proteolytic cleavage was almost complete, each Mab protein solution (10mg/ml) was treated in the same digestion conditions for 5 h at 37°C shaking in a waterbath. Bead-immobilized papain is particularly convenient for Fab preparation: indeed at the end, to remove, the protease samples are spun first at 4°C for 15 min at 3000 rpm and then 4°C for 15 min at 14000 rpm. The resulting

digest was dialyzed over night against each specific buffer used in the further ion exchange chromatography purification step.

Purification of the FabMNAC13

After papain digestion the resulting sample was dialyzed at 4°C over night against Tris HCl 100mM pH 8.0.

The removal of Fc fragments and uncleaved IgG1 was achieved by passage through a DEAE-Sephacel (Pharmacia) column equilibrated with the same buffer. FabMNAC13 was collected in the flow through while Fc fragments and a fraction of uncleaved IgG1 were eluted by 200mM NaCl.

After this first step of purification FabMNAC13 was concentrated using Centricon 30K (Amicon) concentrators: to separate Fab fragments from the quite low amount of uncleaved IgG1 that were still present size exclusion chromatography on a Superdex G75 column (Pharmacia) was performed using an FPLC system (Pharmacia). After equilibration with two bed volumes of Tris HCl 100mM pH 8.0, NaCl 150mM at a flow rate of 0.7 ml/min and calibration of the column by a cocktail of molecular weight markers 200 µl of sample were loaded and fractions across the 50 KDa peak were collected and assessed for homogeneity by Coomassie blue staining after separation by SDS-PAGE electrophoresis (Figure 4A). Considering that no contaminating band was still present, all the fractions were pooled and concentrated using Centricon 30K (Amicon) concentrators obtaining sample in a concentration range between 5 and 15 mg/ml with a purity higher than 99%, as shown by SDS-PAGE electrophoresis and Western Blot analysis (Figure 4B).

The exact amounts of purified protein were determined by Lowry Assay (Bio-Rad). The overall procedure is shown in Figure 5.

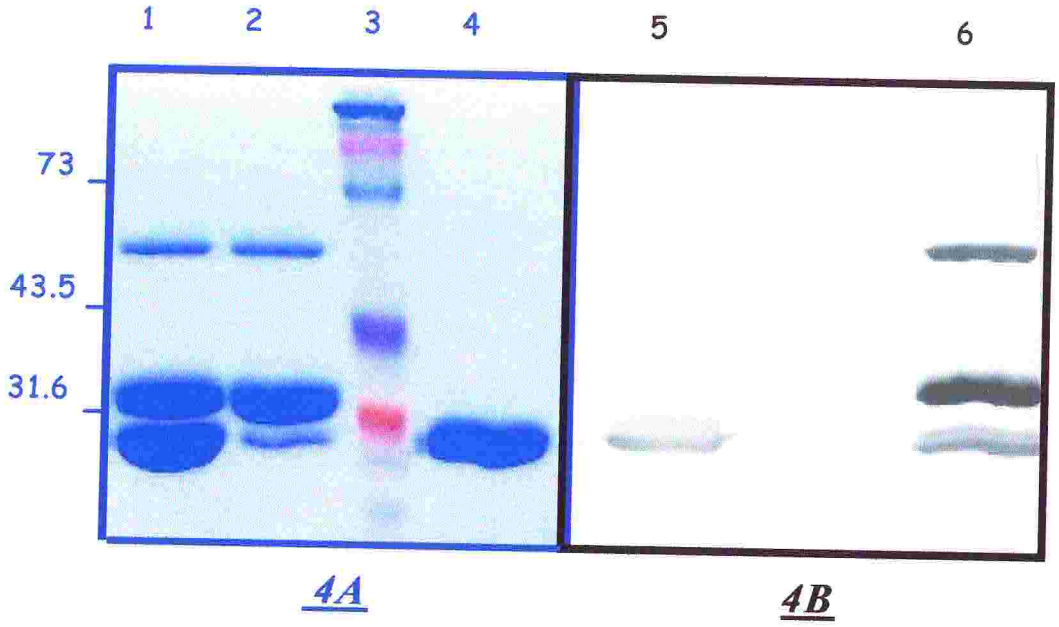


Figure 4A: SDS-PAGE and Coomassie staining. Lane 1: digested sample, Lane 2: DEAE bond Fc and Mab eluted by Tris 100mM pH 8.0 NaCl 200Mm, Lane 3: MW, Lane 4: purified Fab after gel filtration. ***Figure 4B***: SDS-PAGE and Western blot by anti-mouse antibody conjugated with alcaline phosphatase. Lane 5: purified Fab after gel filtration, Lane 6: digested sample.

Purification of the FabαD11

The overall procedure to purify FabαD11 is very similar to the one applied for FabMNAC13 purification and shown in Figure 5.

Indeed the same two purification steps are performed: first ion exchange chromatography to remove Fc fragments and a fraction of uncleaved IgG2a, then size exclusion chromatography to increase purity level of FabαD11 sample removing left traces of intact Mab.

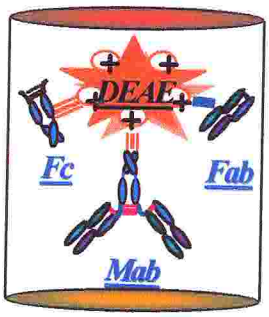
So after papain cleavage the resulting digest was extensively dialyzed at 4°C over night against sodium phosphate 10mM pH 7.8 and loaded on a DEAE-Sephacel (Pharmacia) column equilibrated with the same buffer.

Ion Exchange Chromatography

***DEAE** sephacell (anion exchange resin)*

MNAC13

100mM *Tris* **pH 8.0**



αD11

10mM *Na Phosphate*
pH 7.8

Fab ***pI higher*** than Fc and Mab's

weaker Fab interaction

Fab collected in the flow through

Gel filtration Chromatography

SuperdexG75

100mM *Tris* **pH 8.0**, 150mM *NaCl*

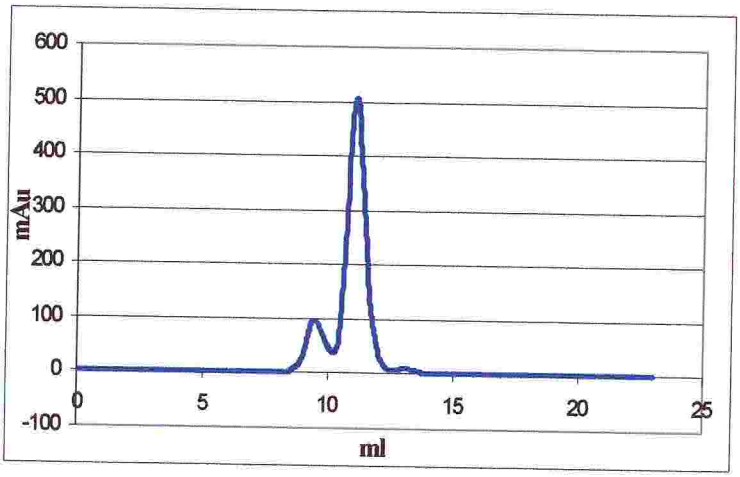


***Figure 5**: Fab purification procedure*

FabαD11 was collected in the flow and through concentrated using Centricon 30K (Amicon) concentrators, while Fc fragments and uncleaved IgG2a were eluted by sodium phosphate 200mM pH 6.8.

In this case this single purification step was sufficient to eliminate also all contaminating undigested IgG2, as it could be assessed by SDS-PAGE and Coomassie blue staining (Figure 6). After concentration using Centricon 30K (Amicon) concentrators it was possible to obtain sample in a concentration range between 10 and 15 mg/ml with a purity higher than 99%. Exact amounts of purified protein were determined by Lowry Assay (Bio-Rad).

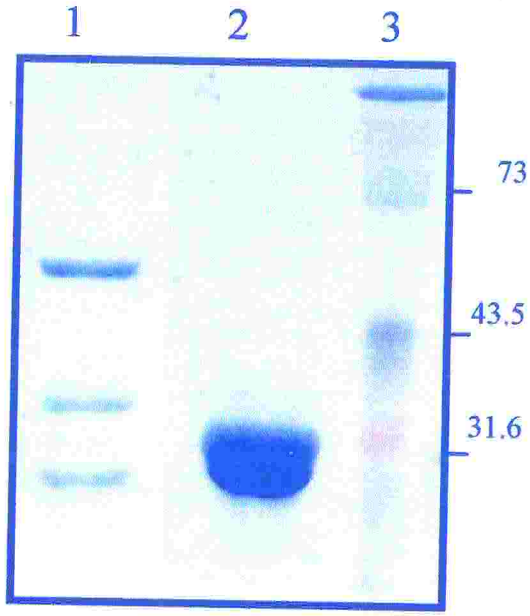


Figure 6: SDS-PAGE and Coomassie staining. Lane 1: digested Mab:, Lane 2: purified Fab, Lane 3: MW

Discussion

In order to perform a structural study of the antigen binding sites of the two blocking antibodies, the classical approach of focusing on their Fab fragments has been followed. As a consequence of that the starting molecules are the corresponding monoclonal antibodies, which are obtained by hybridomas by *in-vitro* tissue culture. By this expression system is possible to collect the antibodies in the tissue culture supernatant with a concentration around 50μg/ml. Then the purification protocol of both immunoglobulins is articulated in two following steps.

First selective precipitation by salting-out procedures using ammonium sulphate, then antibody affinity chromatography using Protein G Sepharose that strongly binds to the Fc region of IgG.

Therefore to obtain Fab fragments purified IgG are subjected to controlled proteolytic digestion using papain, which splits the molecules in two Fab fragments and an Fc fragment in the hinge region. In particular bead-immobilised enzyme is useful for Fab preparation as the protease can be quickly removed by centrifugation.

The result of a typical proteolytic digest of an antibody consists in several components (native undigested antibodies, Fab fragments and Fc fragments). In order to achieve separation of the required antibody fragment from the other contaminants a wide variety of methods have been reported in the literature.

A first approach that has been unsuccessfully followed is based on antibody affinity chromatography using Protein G or Protein A: this is essentially a passive purification method where Fab fragments do not bind to the ligand while Fc fragments and intact antibody do. Following this approach both Protein G and Protein A have been employed and the same disappointing results were obtained using a specific kit (Pharmacia) developed to perform this task.

A second method that has been performed with partial positive results is ligand affinity chromatography: as opposite of the previous one it allows to specifically bind Fab fragments relying on the strength of the interaction with the ligand itself, which is coupled to a column by CN Br activated sepharose. So both NGF and TrkA immunoadhesin have been immobilized on sepharose to perform ligand affinity chromatography. While very low binding capacity was shown by TrkA sepharose, NGF resin allowed purifying quite good amounts of αD11 Fab; unfortunately

the life of this column was too low to permit to scale up the protocol in the range of amounts required for crystallization.

The best approach turned out to be in both cases ion exchange chromatography. Indeed even if the isoelectric points (pI) displayed by Fab fragments depends not only on the type of enzyme used in the hydrolysis but also on the antibody sub-class, it is usually higher than the one of intact immunoglobulins and Fc fragments. So using an anion exchange resin, such as DEAE sephacell, equilibrated with a buffer whose pH is lower than pI of Fab fragments but higher than pI of Fc fragments, the interaction of Fab fragments with positive charged groups of the matrix should be so low to leave them in the flow through. By this approach it was possible to eliminate Fc fragments and a portion of intact immunoglobulins. To increase sample purity eliminating even traces of undigested antibodies size exclusion chromatography was combined.

CHAPTER 6: Fabs crystallization

There are a number of potential bottlenecks in determining a crystal structure, but growing a useful crystal can be the most serious one.

This is particularly true in the case of proteins considering their conformational flexibility and chemical versatility that make this kind of molecules not ideally suited to be stacked into the periodic lattice of a crystal. Indeed protein crystals contain on average 50% solvent, mostly in large channels between the three dimensional array stacked molecules. Moreover the interactions holding the molecules together are hydrogen bonds, salt bridges, and hydrophobic interactions. As a consequence of that protein crystals are quite small (rarely over 10^{-3} mm^3), very fragile, soft and sensitive to all kind of environmental variations.

INTRINSIC PHYSICO-CHEMICAL	BIOCHEMICAL-BIOPHYSICAL
Supersaturation	Conformational sensitivity
Temperature	Binding of ligands
Time	Specific additives
Ionic strength	aging
Diffusion and convection	PURITY
Volume and geometry of set-ups	Conformational hetereogeneities
Solid particles, wall and interface effects	Post-translational hetereogeneities
Pressure, electromagnetic fields	Sequence hetereogeneities
Density and viscosity	Contaminants
Vibration and sound	Batch effects

***Table 1*: Parameters involved in biocrystallization**

Even if biocrystallization involves a much larger number of parameters due to peculiar physical-chemical properties of biological macromolecules (Table 1), chemical precipitants are the most widely used method of achieving supersaturation of macromolecules in order to induce crystallization. In general, the main influence of these compounds is on the solvent (e.g. bulk water) rather than on the solute (protein). Most of them change the chemical potential of the protein in solution; at high concentration they compete for water molecules that hydrate protein surface and so in these conditions protein association becomes energetically more favorable to minimize surface tension effects. For crystallization of proteins, the major classes of precipitants may be divided into four categories (McPherson, 1990):

- salts,
- high molecular weight straight chain polymers (PEG),
- MPD (2-methyl-2,4 pentane diol),
- organic solvents (ethanol, methanol, acetone, isopropanol, DMSO, or tert-butanol)

However the current state of understanding of biocrystallization phenomena is so poor that only a strictly empirical methodology can be employed. Therefore crystallization is a matter of

◊ searching as systematically as possible ranges of individual parameters that impact on crystal formation to find a set or multiple sets of these factors,

◊ optimizing them in order to obtain the best possible crystal for X-rays analysis.

For this reason a vast array of crystallization trials are set up and the obtained information can be used in successive rounds of trials to improve crystal size and quality. This is usually done by initial screening

that is followed by a systematic optimization of conditions and can be based on different approaches:

- **Sparse matrices** for initial crystallization trials are based on conditions, which were most frequently reported in the crystallization literature (Jancarik and Kim, 1991; McPherson, 1992). In this way it is possible to achieve a broad enough sampling of parameter space by random (or near random) combination of conditions to yield initial crystals to be improved upon.
- **Grid screens** are typically based on two dimensional matrices with the two different factors to be tested as the two axes (typically precipitant concentration *vs.* pH). Screening is performed as an iterative process, starting with a course grid over a wide range to find a very fine grid over a narrow range until the two parameters of the grid converges on a best condition. Further optimization can be obtained testing additional factors (temperature, additives).

Four main methods are commonly employed achieve supersaturation in the crystallization of macromolecules:

- **Free interface diffusion**
- **Batch method**
- **Dialysis**
- **Vapor diffusion technique**

In particular the last one, that has been employed, utilizes evaporation and diffusion of water between solutions of different concentration as a means of approaching and achieving supersaturation of macromolecules.

A small volume of high concentrated protein solution is mixed with an about equal amount of reservoir solution containing the precipitants. A

drop of this mixture is put on a glass slide, which covers the reservoir and is sealed with grease (Figure 1).

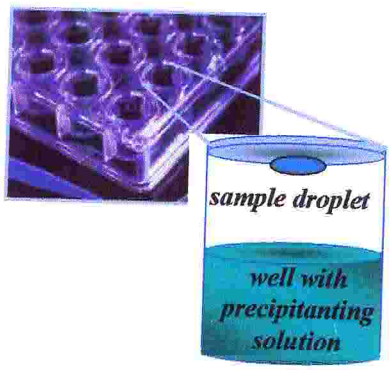


***Figure 1*: Hanging drop**

The difference in precipitant concentration between the drop and the well solution is the driving force, which causes over the time a net evaporation of water from the drop into the reservoir.

A way to insure nucleation and crystal growth consists in seeding (Stura *et al.*, 1991, McPherson, 1988). Usually a protein crystal is crushed to a fine powder in a storage solution to obtain a seed solution, which is then added to the pre-equilibrated supersaturated solution thus ensuring the presence of nuclei for crystal growth. If no crystalline material is available for a particular protein, seed solutions from related proteins (the same protein, but from a different species) may sometimes trigger crystallization of the target protein; this case is generally referred as heterogeneous nucleation.

Figure 2: My cat Minou

By **Microseeding,** seeds are introduced either by adding small volumes of the seed solution directly to the equilibrated protein solution (and thus diluting it) or by touching a crystal by a whiskers (from my cats, Figure 2) and then streaking this across the surface of the drop

(streak seeding). By **macroseeding** small crystals may be enlarged to suitable size for data collection (Thaller *et al.*, 1981). This can be achieved by transferring them from the solution in which they were grown into pre-equilibrated protein/precipitant solutions, after a partial melting in a lower concentration precipitant solution to remove any potential nuclei at its surface. Of course the final pre-equilibrated solution should be in the metastable region of supersaturation to allow crystal growth without nucleation. The entire process can be repeated multiple time in order to increase the size of the crystal.

Methods and results

FabMNAC13 crystallization

The purified FabMNAC13 fragment in 10mM Tris pH 8.0 and 50mM NaCl was concentrated to 5-10 mg/ml using Centricon ultrafiltration unit with a membrane of 30Kda molecular weight cut-off (Amicon). Initial crystallization experiments were based on the sparse-matrix sampling method (Jancarik & Kim, 1991) using Crystal Screen I and II purchased from Hampton Research (Laguna Niguel, CA, USA). All crystallization trials utilized the hanging-drop method and were conducted at 16°C. Drops of 2 µl of the protein stock solution were mixed with an equal volume of the precipitant solution and equilibrated by vapor diffusion with a reservoir solution (0.7 ml) in 24 well Linbro plates. The best results were initially obtained mixing equal volumes containing 2M ammonium sulfate, 5% v/v isopropanol (Crystal Screen II, Reagent #5) and are shown in Figure 3.

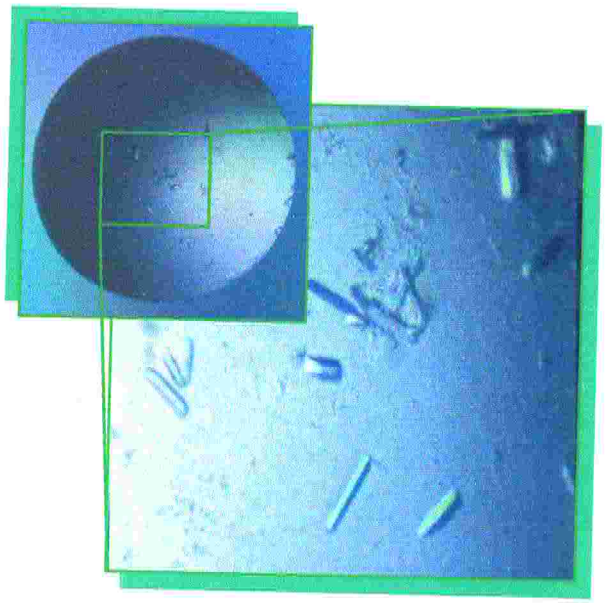


***Figure 3*: Preliminary crystallization condition for FabMNAC13**

By slight variations in sample concentration it was possible to optimize crystallization conditions obtaining "bundles like crystals" as the ones in Figure 4.

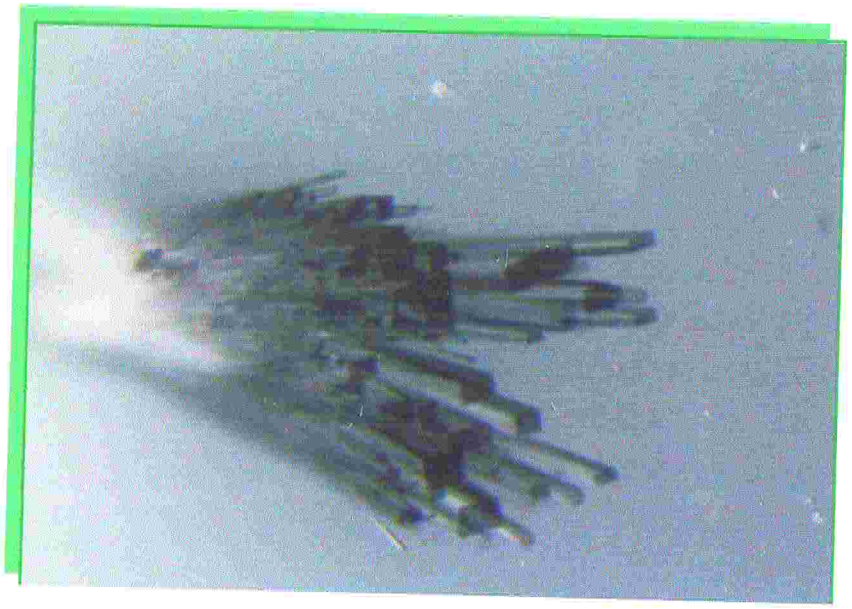


***Figure 4*: Bundles like crystals of FabMNAC13**

In these conditions diffraction-quality crystals grew to their maximum dimensions (*ca.* 0.8 x 0.3 x 0.2 mm) in one week as shown in Figure 5.

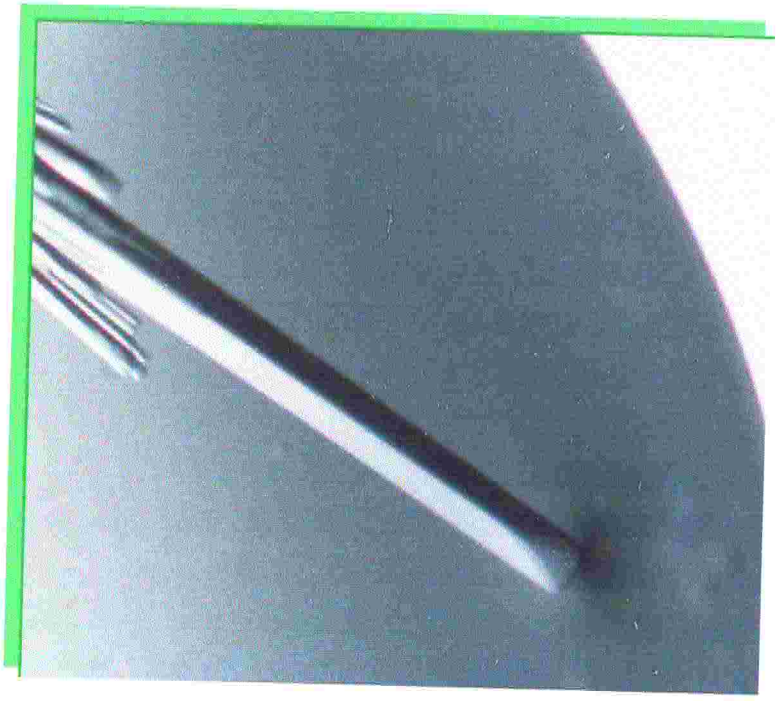


***Figure 5*: Typical crystal of FabMNAC13**

FabαD11 crystallization

In a similar way of what has been previously reported for FabMNAC13 purified FabαD11 fragment in 10mM Na Phosphate pH 7.8 and 50mM NaCl was concentrated to 10 mg/ml using Centricon ultrafiltration unit with a membrane of 30KDa molecular weight cut-off (Amicon). Crystallization experiments have been set up using the hanging drop vapor diffusion method and were conducted at 16°C. Drops of 2 µl of the protein stock solution were mixed with an equal volume of the precipitant solution added to the reservoir (0.7 ml) in 24 well Linbro plates.

Initial screening was based on the sparse-matrix crystallization screening conditions developed by Jancarik & Kim (1991) and commercially available from Hampton Research (Laguna Niguel, CA, USA). It is worth noting that results of the initial screens showed that precipitant concentrations in Hampton Crystal Screen I and II were too high and led to precipitation under most conditions. Anyway reducing precipitant concentrations and/or protein concentrations to half revealed no promising crystallization conditions either.

Besides experiments based on the sparse-matrix sampling method also two different grid screens, both based on PEG3350 as a precipitant, were performed. At first Low Ionic Strength Screen purchased from Hampton Research was used. This screen has been designed for crystallization of soluble proteins in general and in particular of Fab and Fv fragments. Moreover following what reported by Valjakka J *et al.* (2000) a grid screen has been prepared combining increasing PEG3350

concentrations (5%, 10%, 15% and 20%) and increasing pH (Table 2).

Buffers 0. 1M pH	Citric acid 3.0	Na Acetate 4.0	Na Citrate 4.5	Bis tris Propane 5.5	MES 6.0	ADA 6.5	Na cacodylate 7.0	HEPES 7.5	Tris 8.0	Bicine 8.5

Table 2

The results of both sets of crystallization trials were disappointing. Trying to change completely approach a further screening was based on the sparse-matrix crystallization screening conditions purchased from Jena BioSciences. The most promising results were obtained using 20% PEG4000, 0.6M NaCl, 100mM MES pH 6.5 (Kit number 4, solution C2) and could be only partially optimized by slightly lowering the concentration of PEG4000 to a final value of 18%. Typical crystals (Figure 6) grew in one week. It was quite hard to reduce the nucleation

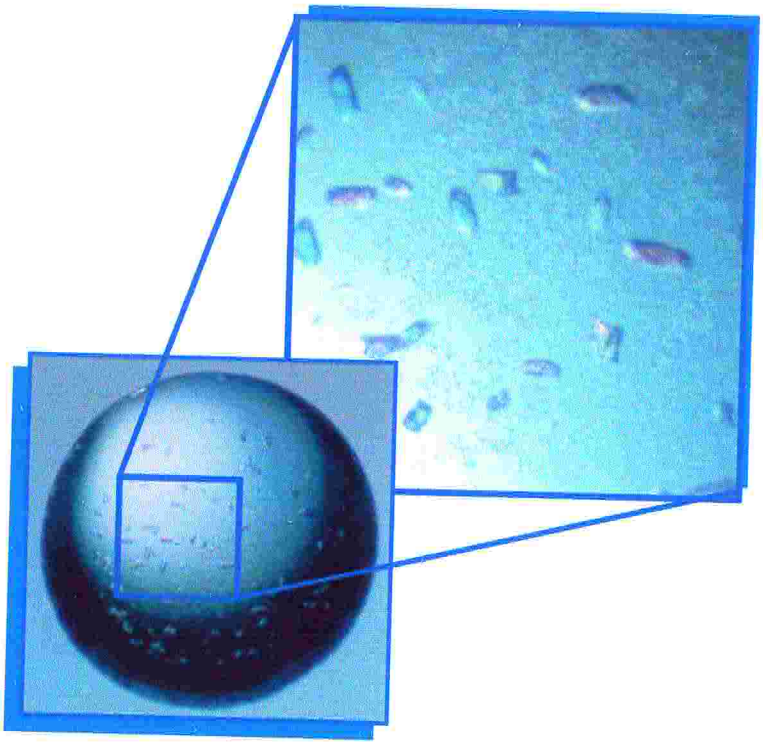


Figure 6: Typical FabαD11 Crystal Form I, obtained in MES buffer

and moreover no great improvements in crystal shape could be achieved even with different ratios protein/precipitant or carring crystallization experiments at 4°C in order to reduce the rate of nucleation and/or growth. From these attempts two different kinds of crystals (denominated Crystal Form I). could be grown:

- ◊ in conditions very similar to the starting ones it was possible to obtain quite high number of crystals per drop, whose shape was not completely well formed and whose size was quite small;
- ◊ with lower concentrations of PEG4000 and adding several types of detergents from Detergent Screen I and II purchased by Hampton Research it was possible to decrease their number slightly increasing their size, even if the vast majority of these crystals was not single and so unsuitable for X-rays data collection.

Considering that the combination of quite high salt (0.6M NaCl) with PEG4000 gave some positive indications, a grid screen analogous to the one reported by Valjakka J *et al.* (2000) and previously set up, was performed in parallel using both PEG3350 and PEG4000 and adding 0.6M NaCl.

Interestingly not only no positive results were obtained using PEG3350 but also using conditions very similar to the ones of crystallization condition found in Jena BioSciences Kit no crystals were obtained using a different buffer (ADA instead of MES) keeping constant pH value, ionic strength, salt and precipitant concentrations.

On the other hand crystals were obtained at a different pH value using 20% PEG4000, 0.6M NaCl, 100mM BTP (bis tris propane) pH 5.5. Also these conditions were hard to optimize in order to reduce the number and to increase the size of crystals. However the crystal growth rate was quite slower in these conditions so that crystals grew over several months and reached slightly larger dimensions. Moreover they appeared to be more

regular and better shaped than the previous ones, as it can be noticed in Figure 7 (and have been denominated Crystal Form II).

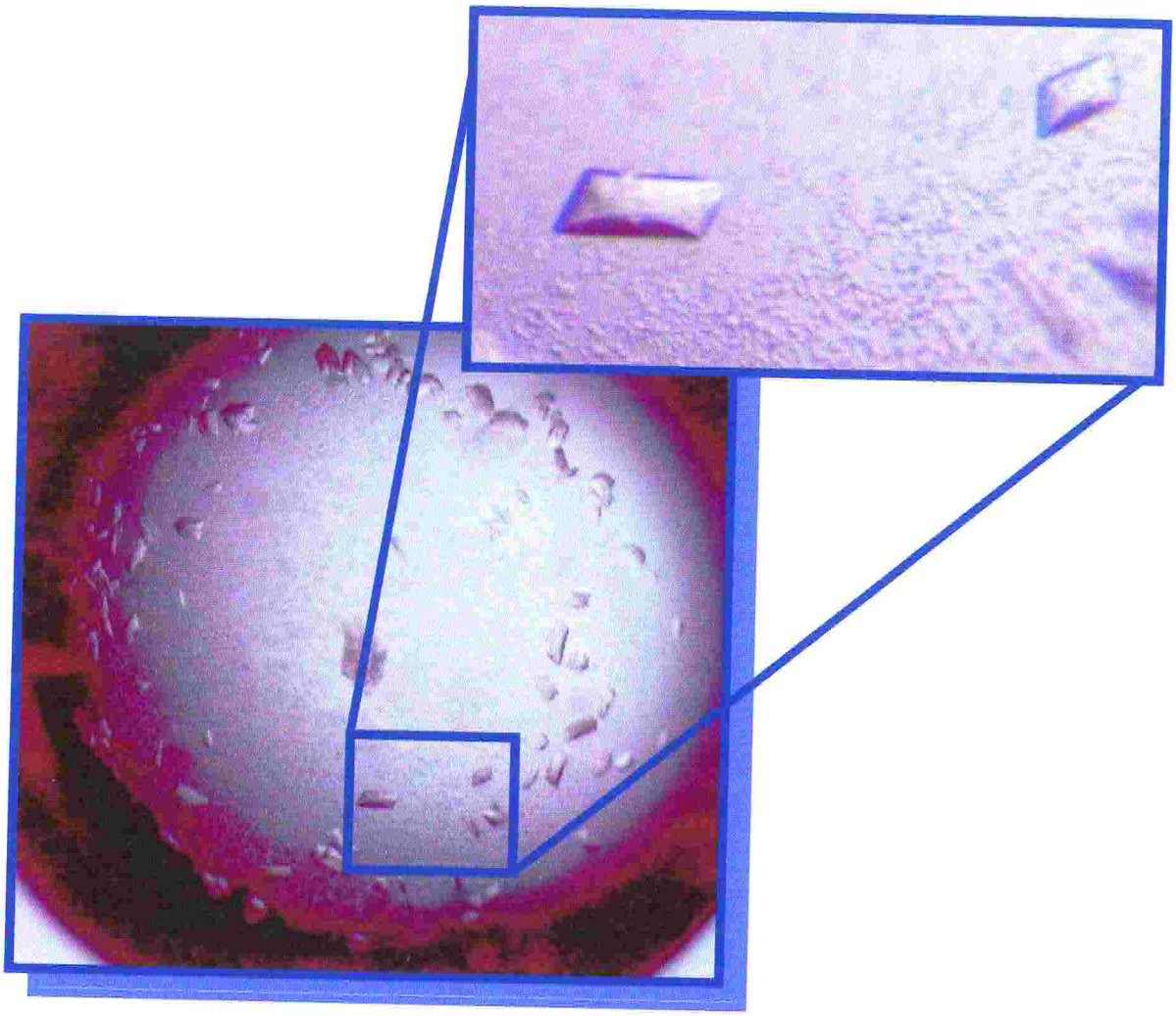


***Figure 7**: Typical Fab αD11 Crystal Form II, obtained in BTP buffer*

Trying to optimize this promising starting condition it was possible to reduce the number of crystals per drop and to further improve crystal shape monitoring different protein/precipitant ratios. The best result is shown in Figure 8 and was obtained using 20% PEG4000, 0.6M NaCl, 100mM BTP pH 5.5 in a protein/precipitant ratio equal to 1.5.

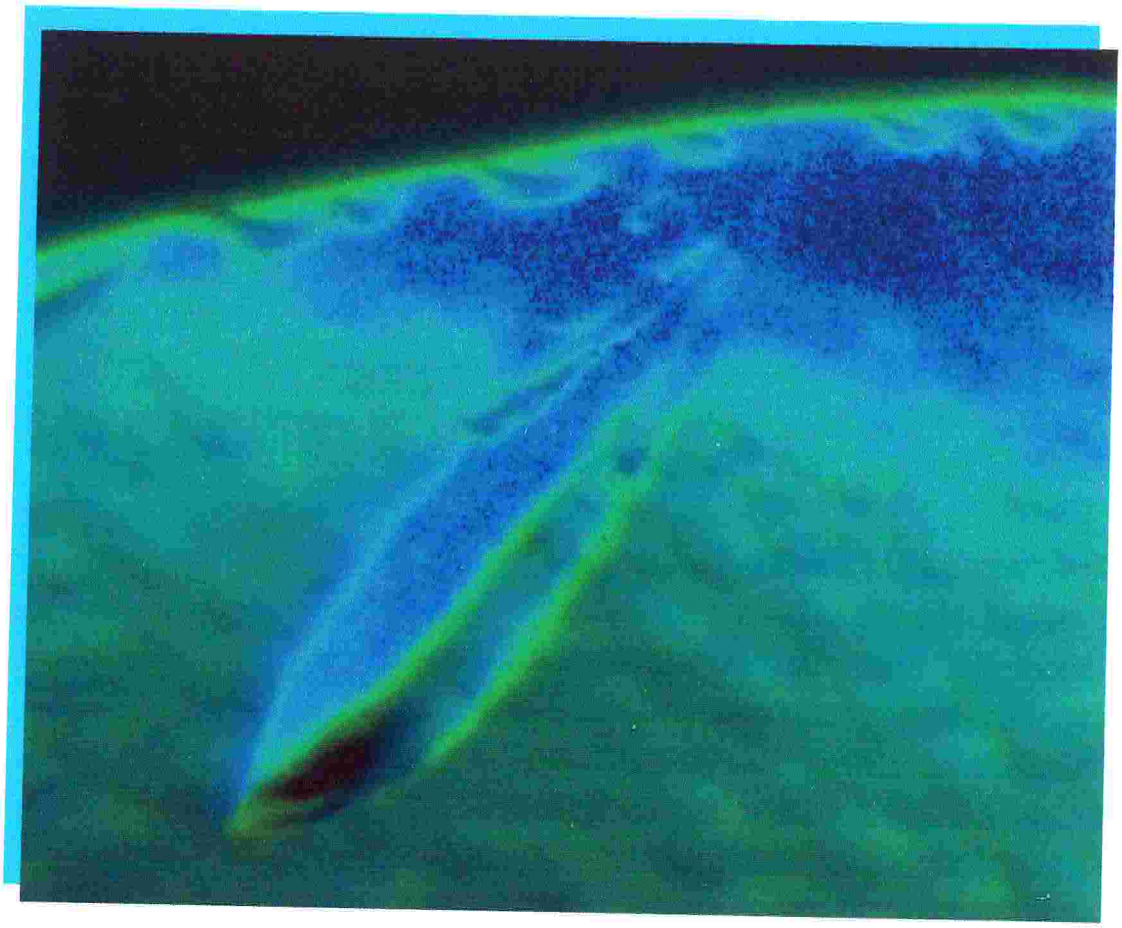


***Figure 8**: Typical Fab αD11 Crystal Form III, obtained in protein:precipitant ratio equal to 1.5.*

It is worth noting that only one crystal per drop grew in this condition and moreover their shapes were slightly different and their dimensions were quite larger than from the previously observed ones, therefore they have been denominated Crystal Form III.

References

Jancarik, J. and Kim, S-H.J. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *Appl. Cryst.*; **24**, 409-411.

McPherson, A. (1992). Two approaches to the rapid screening of crystallization conditions. *J. Cryst. Growth.*; **122**, 161-167.

McPherson, A. (1990). Current approaches to macromolecular crystallization. *Eur. J. Biochem.;* **189**, 1-23.

McPherson, A. and Schlichta, P. (1988). The use of heterogeneous and epitaxial nucleants to promote the growth of protein crystals. *J. Cryst. Growth.;* **90**, 47-50.

Stura, E. A., Nemerow, G. R., and Wilson, I. A. (1991). Strategies in protein crystallization. *J. Cryst. Growth.*; **110**, 1-12.

Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R., and Jansonius, J. N. (1981). Repeated seeding technique for growing large single crystals of proteins. *J. Mol. Biol.;* **147**, 465-469.

Valjakka, J., Hemminki, A., Teerinen, T., Takkinen, K., Rouvinen, J. (2000). X-ray studies of recombinant anti-testosterone Fab fragments: the use of PEG 3350 in crystallization. *Acta Crystallogr D Biol Crystallogr*; **56**, 218-221.

CHAPTER 7: X-rays data collection and processing

To perform an X-ray diffraction experiment besides rotating anode generators producing a X-ray beam of a characteristic wavelength, synchrotron facilities generate an intense, tuneable X-ray radiation. When high energy electrons are deflected by strong magnetic fields, they emit tangentially electromagnetic waves called synchrotron radiation, whose main and more important characteristics are:

- High intensity
- Very broad continuous spectral range
- Narrow angular collimation
- Small source size
- High degree of polarization
- Regularly pulsed time structure

All the X-ray diffraction experiments were performed into two main third generation synchrotron radiation facilities:

Figure 1: Elettra facility

ELETTRA (Figure 1) provides with light of high brightness with photon energies in the range 10 eV to 30 KeV, and with spectral brilliance of up to 10^{19} photons/s/mm^2/mrad^2 /0.1% bw. The working energy reaches 2.4 GeV, with a current of 320 mA.

ESRF (Figure 2) is even more focussed on X-ray diffraction biocrystallography experiments operating in the hard X-ray range: indeed it generates X-ray beams about a trillion times brighter (a factor of 10^{12})

Figure 2: ESRF facility

than those of conventional X-ray sources used in laboratories reaching photon energies between 1 eV to 100 GeV and spectral brilliance in the 10^{20} range. The working energy reaches 6 GeV, with a current of 90 mA..

The primary X-ray beam is monochromated by crystal monochromators and focused by a system of three toroidal mirrors (Figure 3).

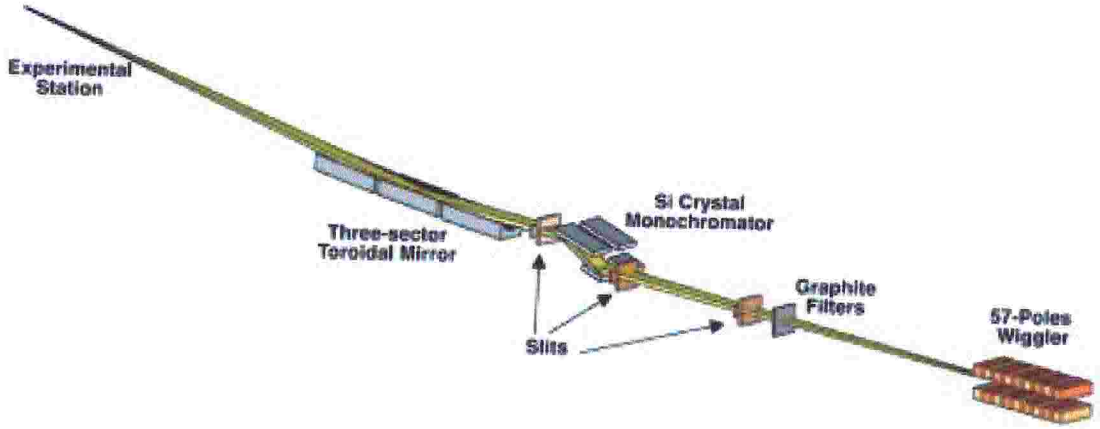


Figure 3: XRD-1 layout

The crystal is mounted on a pin on a goniometer head, which allows positioning the crystal in different orientations. The X-ray diffraction pattern of the crystal is recorded using fast read-out two-dimensional detectors, such as imaging plate (as mar345 IP) or Charge-Coupled Device sensor-based (CCD) detector (as marCCD).

Intrinsic limitations due to diffraction quality, can be dramatically improved by cryo-techniques, in which crystals are flash-cooled to near liquid nitrogen temperatures (~100 K). Many factors contribute to improvements in data quality. In addition to the obvious advantage of reduced thermal vibrations, an important secondary effect is the suppression of radiation damage, permitting a complete data set to be collected from one single crystal. This in turn eliminates errors from merging and scaling of data sub-sets from multiple crystals. To avoid crystalline ice formation that can destroy protein crystal or interfere with data collection giving rise to powder diffraction rings, cryoprotectant agents are added to the mother liquor and flash cooling is performed in order to convert water in amorphous vitreous ice.

Thus large number of these images recorded from different crystal orientations are processed (scaled and merged) into a final list of indexed reflection intensities.

Data collection

Cryoprotection and crystal mounting

For FabMNAC13 crystal mounting 20% glycerol was added to the mother liquor as a cryo-protectant: the cryo-protectant solution (2.2 M ammonium sulfate, 6% v/v isopropanol and 20% v/v glycerol) was first tested in order to verify ice ring formation.

After a brief soaking in this cryo-protectant solution (5-10s) the crystal was fished by a loop (Hampton Research) of suitable dimension and quickly mounted on a goniometer head, allowing a flash-cooling of the crystal under the stream of liquid nitrogen at 100K using an Oxford Cryosystems cooling device (Oxford, UK).

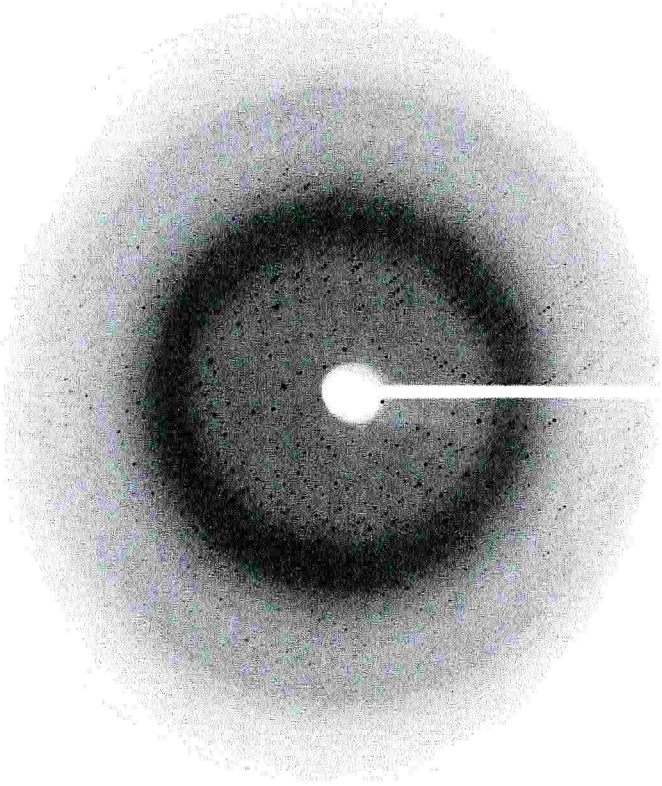
In the case of FabαD11 crystal-mounting no cryoprotectant agent was required, since 20% PEG4000 present in the mother liquor could supply for this function.

Preliminary analysis and data collection

A preliminary analysis of the diffraction pattern of the crystal is necessary to evaluate its quality (mosaicity, diffraction limit, if it is split or not before starting data collection. Important data collection parameters such as oscillation angle, detector to crystal distance and exposure time can be set according this preliminary analysis. Moreover in order to choose the best data collection strategy few images taken from different crystal orientations are used for unit cell, lattice and space group determination. Then an appropriate data collection strategy can be designed.

- ***FabMNAC13 data collection***

An initial X-ray diffraction data set was collected on a FabMNAC13 crystal (0.5 x 0.2 x 0.1 mm) at the XRD1 beam line of the ELETTRA synchrotron light source (Trieste, Italy). In a preliminary analysis of the diffraction pattern it was possible to determine the Bravais lattice using the autoindexing procedure implemented in the program DENZO (Otwinowski & Minor, 1997; Otwinowski, 1993) as primitive orthorombic space group and the resolution limit (2.5Å).



<u>Figure 4</u>: Representative diffraction pattern of FabMNAC13

A higher resolution and more complete X-ray diffraction data set was collected on a FabMNAC13 crystal (0.8 x 0.3 x 0.2 mm) at the ID14-EH1 beam line of ESRF (Grenoble, France). A representative diffraction pattern is shown in Figure 4. Diffraction images were recorded in two

steps at different exposure time and crystal to detector distance to measure accurately both low and high resolution reflections obtaining good I/σ(I) ratio and higher redundancy.

- ***FabαD11 data collection***

In a preliminary analysis many FabαD11 crystals were tested at the XRD1 beam line of the ELETTRA synchrotron light source (Trieste, Italy). Several larger crystals of Form I (grown from 18% PEG4000, 0.6M NaCl, 100mM MES pH 6.5) were mounted to analyse their diffraction patterns. Unfortunately, as it can be noticed analysing the diffraction patterns shown in Figure 5, it was clear that the crystals were not single and therefore they could not be used for data collection even if they diffracted up to 2.0 Å resolution.

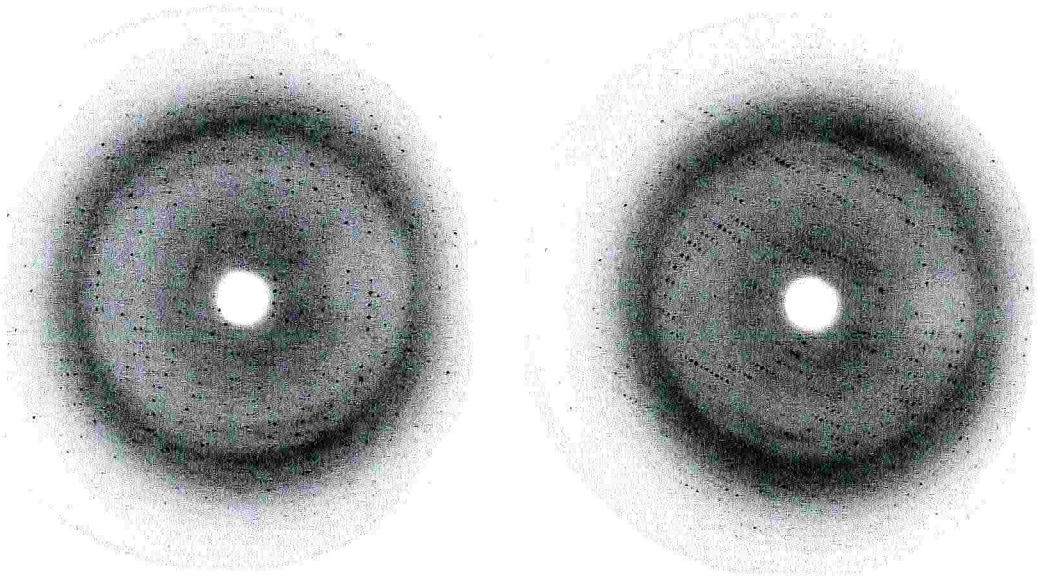


Figure 5: Two diffraction pattern of FabαD11 Crystal Form I, taken at 90° one from the other

All the tested smaller crystal obtained in similar condition showed a very poor diffraction even if they were single. Nevertheless it was possible to

determine the crystal lattice, using the autoindexing procedure, as primitive triclinic P1.

On the other hand testing crystal of Form II (grown in 20% PEG4000, 0.6M NaCl, 100mM BTP pH 5.5), whose appearance was similar, it was possible to obtain a diffraction image of a single crystal, whose resolution limit was around 3.0 Å.

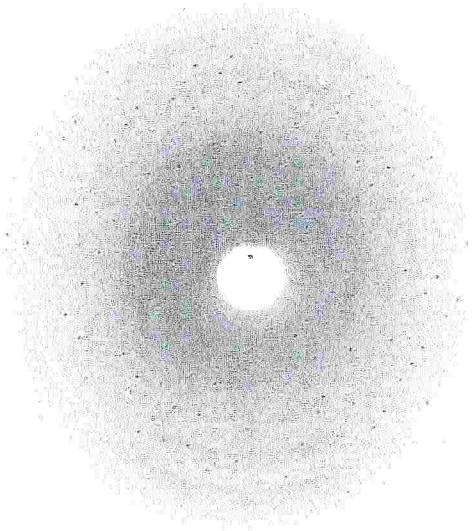


***Figure 6:* Representative diffraction pattern of FabαD11 Crystal Form II**

Thus a 2.7Å resolution data set was collected for FabαD11 at the XRD1 beam line of the ELETTRA synchrotron light source (Trieste, Italy). A representative diffraction pattern is shown in Figure 6.

Also in this case a second data set was collected at 1.7Å resolution on a FabαD11 crystal of Form III (0.7 x 0.4 x 0.2 mm) at the ID14-EH1 beam line of ESRF (Grenoble, France). As shown by the representative diffraction pattern in Figure 7, it was possible to increase both resolution and completeness of the resulting X-ray diffraction data set. Surprisingly during preliminar analysis using the autoindexing procedure it came out

that the crystal lattice was different from the previous one, being monoclinic C centered.

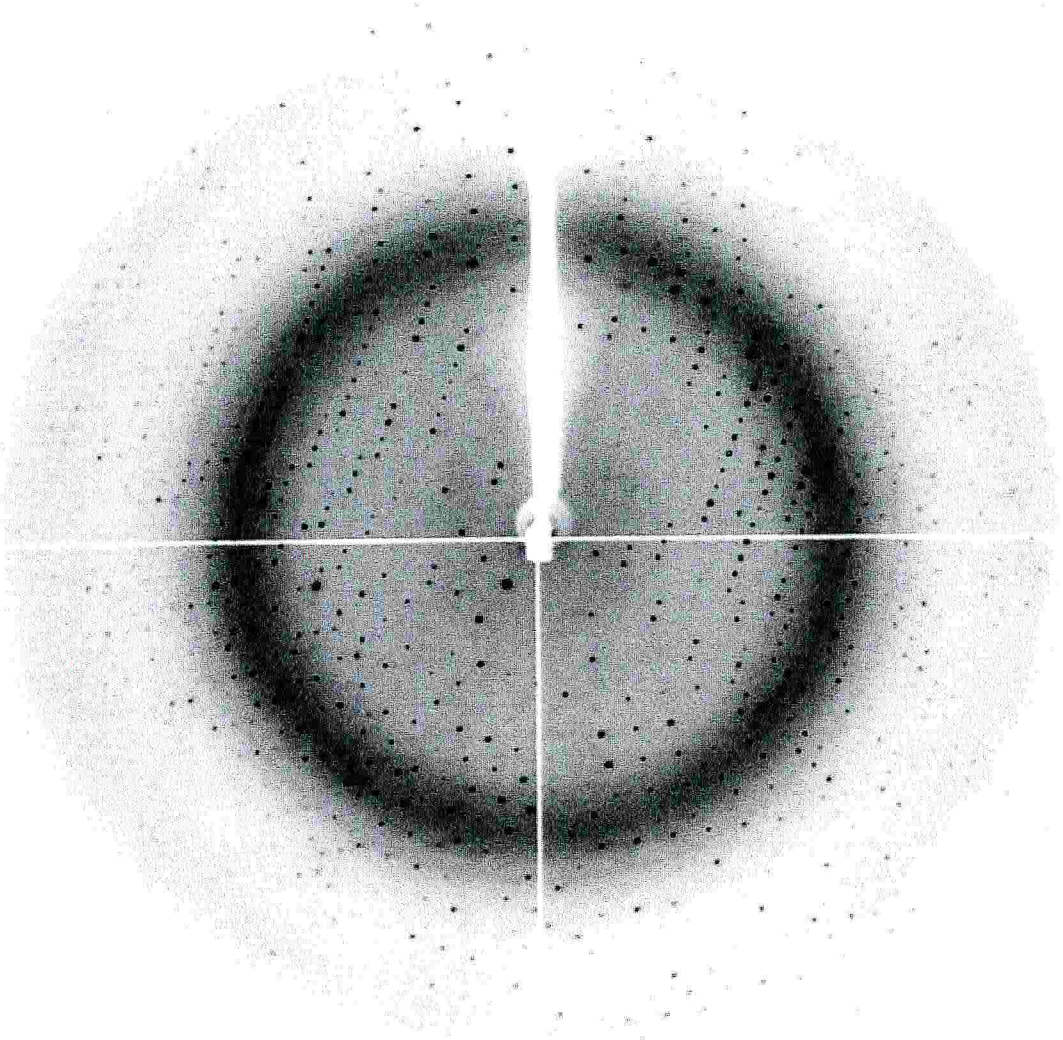


Figure 7: ***Representative diffraction pattern of FabαD11 Crystal Form III***

Data processing

- ***FabMNAC13 data processing***

Therefore both X-ray diffraction data sets, collected for FabMNAC13, were indexed, integrated and subsequently scaled using the programs *DENZO* and *SCALEPACK* respectively, and the *CCP4* package (Collaborative Computational Project, Number 4, 1994) was used for the data reduction. A summary and comparison of the X-ray diffraction data for the two collected crystals is given in Table 1.

Moreover their intrinsic quality is quite good considering the value of R_{sym} that is calculated as

$$R_{sym}(I) = \Sigma_{hkl}\Sigma_i | I_{hkl,i} - \langle I_{hkl} \rangle | / \Sigma_{hkl}\Sigma_i | I_{hkl,i} |$$

with $\langle I_{hkl} \rangle$ mean intensity of the multiple $I_{hkl,i}$ observations from symmetry-related reflections. Indeed this value remains still quite low even for the highest resolution shell. Finally the redundancy in both cases is quite high while the ratio between signal and background ($\langle I/\sigma(I) \rangle$ of measured data) is quite low.

Finally assuming a molecular weight of 46.600 KDa and one molecule in the asymmetric unit it was possible to estimate (Matthews, 1968) the value of the crystal packing parameter V_M (2.1 $Å^3$ Da^{-1}), corresponding to a solvent content of 42 %

Table 1

Crystal parameters, data collection and processing statistics for FabMNAC13

Crystals	I	II
X-ray source	ELETTRA	ESRF
Wavelength (Å)	1.000	0.934
Detector	mar345	marCCD
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	52.78	52.73
b (Å)	67.53	67.55
c (Å)	111.51	111.43
Mosaicity (°)	0.40	0.47
Resolution range (Å)	12.0 - 2.50 (2.59 - 2.50)	17.0 - 1.80 (1.83 – 1.80)
No. of measurements	98688	414115
No. of observed reflection $I \geq 0$	56918	227914
No. of unique reflections $I \gtrsim 0$	14203 (1371)	38392 (1893)
Completeness (%)	99.5 (99.3)	99.5 (99.6)
Redundancy	4.0 (4.0)	5.9 (4.9)
$\langle I/\sigma (I) \rangle$ of measured data	9.4 (4.7)	8.2 (1.1)
R_{sym} (%)	5.7 (15.2)	6.3 (39.8)

- Values in parentheses are for the highest resolution shell.

- ***Fab αD11 data processing***

In a similar way both X-ray diffraction data sets, collected for Fab αD11 Crystal Form II and III, were processed using the programs *DENZO* and *SCALEPACK*.

Table 2 reports and compares refined crystal and detector parameters and the value of statistical indices.

Both data sets are quite complete, even if there is a decrease in the highest resolution shell of the higher resolution data set, its completeness (78.4%) is still acceptable.

Assuming a molecular weight of 48100 Da and two molecules in the asymmetric unit in the case of the triclinic crystal form, the value of the crystal packing parameter V_M is 2.28 $Å^3$ Da^{-1}, corresponding to a solvent content of 45.6 % (Matthews, 1968).

Regarding the centered monoclinic crystal form, assuming a molecular weight of 48100 Da and one molecule in the asymmetric unit the value of the crystal packing parameter V_M is 2.1 $Å^3$ Da^{-1}, corresponding to a solvent content of 47.56% (Matthews, 1968).

Table 2

Crystal parameters, data collection and processing statistics for Fab αD 11

Crystals	**Form II**	**Form III**
X-ray source	ELETTRA	ESRF
Wavelength (Å)	1.000	0.934
Detector	marCCD	marCCD
Space group	P1	C2
Unit-cell parameters		
a (Å)	42.685	114.801
b (Å)	50.626	69.354
c (Å)	102.697	64.104
α (°)	81.977	90
β (°)	89.116	117.02
γ (°)	85.957	90
Mosaicity (°)	0.44	0.40
Resolution range (Å)	47.6 – 2.7 (2.8 – 2.7)	17.0 - 1.70 (1.75 – 1.70)
No. of measurements	124456	492594
No. of observed reflection I ≥ 0	74241	399184
No. of unique reflections I ≥ 0	23413 (2162)	47951 (3198)
Completeness (%)	98.2 (92.4)	97.2 (78.4)
Redundancy	5.7 (5.2)	6.7 (7.5)
$\langle I/\sigma (I) \rangle$ of measured data	29.6 (6.7)	9.5 (2.1)
R_{sym}(%)	11.0 (33.5)	5.8 (27.8)

- Values in parentheses are for the highest resolution shell.

References

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.;* **33**, 491-497.

Helliwell, J. (1998). Synchrotron radiation facilities. *Nat Struct Biol.*; **5 Suppl**, 614-617.

Minor, W., Tomchick, D., Otwinowski, Z. (2000). Strategies for macromolecular synchrotron crystallography. *Structure Fold Des*; **8**, R105-110.

Otwinowski, Z. & Minor, W. (1997). Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*; **276**, 307-326.

Westbrook, E. M., Naday, I. (1997). Charge-coupled device-based area detectors. *Methods Enzymol* ; **276**, 244-268.

CHAPTER 8: Phase solution

When X-rays interact with matter they are scattered in all directions from the electrons they encounter. Thus travelling different distances, they will differ in their relative phases and there will be interference as they add up. This is the reason why crystals amplify the scattering signal to one that can be measured giving rise to diffraction pattern restricted to discrete spots.

Diffraction spots are often called reflections thinking about the crystal as being composed of thousands of mirrors (Bragg planes defined by Miller indices *h,k,l*) that reflect the X-rays with an angle of reflection equal to the angle of incidence. Thus when light rays hit these "mirrors" in phase, they also exit in phase, regardless of where they hit the mirror because they have identical pathlengths. If rays reflected from a plane have identical pathlengths, then rays reflected from different planes must have different pathlengths. Considering Figure 1, simple geometry and trigonometry tells that for the two rays to be diffracted in phase the pathlengths should differ by any multiple of the wavelength

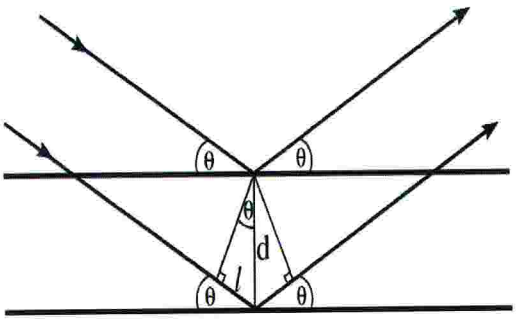


Figure 1

$$n\lambda = 2\,d\,\sin(\theta)$$

also known as the **Bragg's law**.

Therefore while objects on the planes diffract in phase, objects *between* the planes will diffract out of phase with a phase shift proportional to a fraction of the distance to the next plane. In this way a single diffraction event gives information about the positions of objects relative to these sets of planes.

Each reflection of a diffraction pattern is described by the **structure factor,** that is the resultant of N waves scattered into the direction of the reflection HKL by the N atoms in the unit cell. Its expression is:

$$\mathbf{F} = \sum_{j=1}^{N} f_j \exp\left(2\pi i \sigma_j\right)$$

where $\mathbf{f_j}$ is the ***scattering factor*** of the j^{th} atom and σ_j is the ***phase***.

The following equation links the electron density to the structure factor by the Fourier transform operator. Moreover, since the Miller indices implicitly incorporate the cell dimensions, the structure factors can be expressed in terms of Miller indices (h) and fractional coordinates (x):

$$\mathbf{F(h)} = \int_{cell} \rho(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) d\mathbf{x}$$

Since the atoms scattering the X-rays are not fixed in their positions but vibrate around an equilibrium position, their scattering factor is weaken especially at high diffraction angle. This motion is dependent on the temperature and thus called Temperature or Debye-Waller factor, (B-factor). When the vibration is the same in every direction, the isotropic B-factor is defined to by the Debye-Waller equation

$$B = 8\pi^2 \langle u \rangle^2$$

where <u> is the mean displacement of a vibrating atom from its rest position.

Considering that the structure factors are a Fourier transform of the electron density, the latter one is in turn the Fourier transform of structure factors:

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} \mathbf{F(h)} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

In general, an inverse Fourier transform would involve an integral like the forward Fourier transform, but if the object is periodic (like a crystal), it involves just a summation of structure factor contribution, where each amplitude tells the extent to which electrons are concentrated on planes parallel to the Bragg planes, while each phase tells where these concentrations are found. In a diffraction experiment unfortunately only intensities at angle 2θ are observed. The intensities are given by the square of the amplitudes of the structure factors. Thus only structure factor $|F|$ can be measured experimentally, while the phase term is not directly measurable. This is the so called crystallographic phase problem.

These funny pictures give a particularly dramatic illustration of the importance of the phases. In Figure 2 the Cat and his Fourier transform and the Duck and his Fourier transform are respectively represented. Amplitudes are represented by color saturation and brightness, while phases are given by the hue. Figure 3 shows the result of mixing them up, combining the magnitudes from the Cat transform with the phases from the Duck transform. (the brightness from the Cat and the colors from the Duck) and viceversa using the magnitudes from the Duck transform and the phases from the Cat transform. In each case, the image,

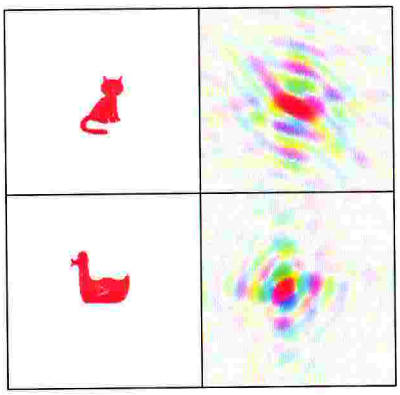


Figure 2

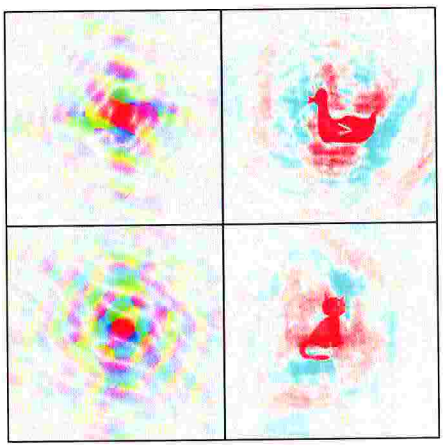


Figure 3

which contributed the phases, is still visible, whereas the image, which contributes the magnitudes, has gone: clearly the phases are dominating.
Among different approaches that can be followed to solve the phase problem, in order to determine both Fab fragments structures molecular replacement was employed.

Molecular replacement

In the early days crystallographers, working on the structures of very simple molecules, were able to make a good guess of the conformation of a molecule and thereafter test it by calculating a diffraction pattern and comparing it to the observed one. If the atoms were positioned in about the right place, then the calculated phases would have resulted approximately correct and a useful electron density map would have computed by combining the observed amplitudes with the calculated phases. When the model became reasonably accurate, such a map would have shown features, that were missing from the model so that the model could have been improved. For proteins, such a guess on the structure is possible only if a closely-related protein structure is already known: this approach that makes use of prior structural information is called molecular replacement.
Molecular replacement can be used with a good model for a reasonably large fraction of the molecule. This means that the model must be fairly complete and must share at least 40% sequence identity with the unknown structure considering that in general the level of resemblance of two protein structures correlates quite well with the level of sequence identity.
To carry out molecular replacement, the model structure has to be placed in the correct orientation (specifying three rotation angles) and position

(specifying three translational parameters) in the unknown unit cell. So if there is one molecule in the asymmetric unit of the crystal, the molecular replacement problem is a 6-dimensional problem that can usually be separated into two 3D problems. A rotation function can be computed to find the three rotation angles, and then the oriented model can be placed in the unit cell by a 3D translation function.

T he rotation and translation functions are based on the Patterson function ($P_{(u)}$). It is defined as the Fourier transform of the structure factor amplitudes:

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{u}),$$

where h extends over all the available part of the reciprocal lattice and u, the coordinate of the Patterson cell.

It is basically an interatomic vector map. Each peak in the map corresponds to a vector between atoms in the crystal and the intensity of the peak is the product of the electron densities of each atom. This is shown schematically in Figure 4.

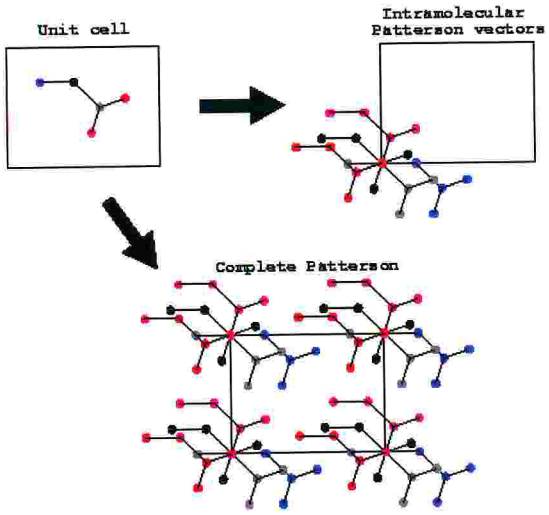


Figure 4

There are two categories of vectors in the Patterson map:

◊ Intramolecular vectors (linking atoms in the same molecule) that
◊ depend only on the orientation of the molecule;
◊ Intermolecular vectors that depend both on the orientation and position of the molecule.

Therefore the Rotation function exploits the fact that the intramolecular vectors depend only on the orientation of the molecule, not on its position in the unit cell.

Indeed in order to quantify the overlap between the observed from the collected diffraction (P_{nat}) and calculated from the model coordinates (P_{mod}) Pattersons, the following product function is used:

$$\mathbf{R}(\kappa,\phi,\psi) = \int_{r_{min}}^{r_{max}} \mathbf{P}_{nat}(\mathbf{u})\mathbf{P}_{\mathbf{mod}}(\kappa,\phi,\psi,\mathbf{u})d\mathbf{u}$$

where the rotation is expressed in terms of kappa/phi/psi. In this procedure, the corresponding positions in the observed and calculated Patterson maps are multiplied, obtaining a maximum when the two Pattersons overlap. Conceptually this consists in physically rotating search model, calculating the Patterson and comparing it with the native Patterson, as shown in Figure 5.The top half shows that if the search model has a random orientation, its corresponding Patterson vectors do not correspond to the observed Patterson. Otherwise with the correct orientation of the search model the calculated Patterson matches the observed Patterson as shown in the bottom half.

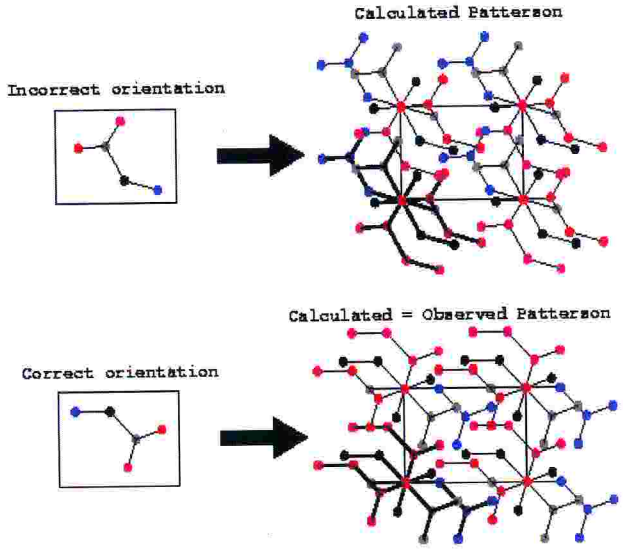


Figure 5

The product of the two Pattersons at all points in a spherical shell is summed, and that is the rotation function score.

Considering that on average the intramolecular vectors will be shorter than the intermolecular vectors, they are enriched near the origin. Therefore the "contamination" by intermolecular vectors can be reduced by restricting the computing of the rotation function to a sphere surrounding the origin. In addition, since there is a huge peak at the origin, this area is not used in the analysis either. In the integration over a spherical shell, a region near the origin is typically omitted to exclude the large origin peak of the Patterson, which would add a large constant term. Then to place the correctly oriented search model at the correct position in the unit cell, a **Translation function** again compares calculated and observed Pattersons. The information about the position is present in the intermolecular vectors using the full Patterson (minus the origin) by a product function, evaluated by a Fourier transform using the correlation theorem. This is illustrated by the formulas below.

$$T(\mathbf{t}) = \int_{cell} P_{2\to 1}(\mathbf{u}-\mathbf{t}) P_{nat}(\mathbf{u}) d\mathbf{u}$$

$$= \frac{1}{V}\sum_{\mathbf{h}} \left\{ F_1(\mathbf{h}) F_2^*(\mathbf{h}) \right\}^* \left| F_O(\mathbf{h}) \right|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{t})$$

$$= \frac{1}{V}\sum_{\mathbf{h}} F_1^*(\mathbf{h}) F_2(\mathbf{h}) \left| F_O(\mathbf{h}) \right|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{t})$$

Phasing by molecular replacement

Given that a number of crystal structures of Fab fragments are available, the most appropriate method to solve the crystallographic phase problem was molecular replacement in both cases. So a parallel close scrutiny of the structures in the Protein Data Bank (Berman *et al.*, 2000) was performed to find out suitable models for the determination of both FabMNAC13 and FabαD11. Then preliminary models for the crystal

structure of both Fab molecules was obtained by molecular replacement using the program AMoRe (Navaza, 1994).

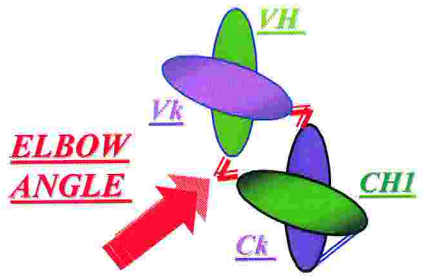


Figure 6: Elbow angle

It is well known that the so-called elbow angle of Fab molecules (schematized in Figure 6) is the most variable parameter among different crystal forms. Therefore to take this into account the search model was divided into variable and constant domains. Moreover to monitor the importance of a difference at the level of the elbow angle, molecular replacement was performed in a parallel way using as models both the whole Fab and the constant and variable domains separately. Molecular replacement protocols and results were quite similar in the solution of the structures of both Fab fragments. Indeed in both cases the rotation function gave clear indications for the orientation for the variable and constant domains. These orientations were used in turn for the translation function, in which the correct solution proved to be the highest peak in each case. A two body translation function was then calculated to confirm the consistency of the independent solutions for the variable and constant domains and to place these two components on a common crystallographic origin. Rigid-body refinement was subsequently performed to optimize the orientation and position of the two domains. The final models had structurally sensible pairing of the heavy and light variable domains with no steric obstruction to the packing of the Fab molecules in the unit cell, thus confirming the correctness of the solution.

- ***FabMNAC13 phasing***

In particular the search for Fab MNAC13 homologues revealed that the most suitable model appeared to be 1BM3: the structure of an immunoglobulin Opg2 Fab-peptide complex (Kodandapani *et al.*, 1999). This Fab was chosen on the basis of sequence identity (70 and 88% for the heavy and light variable domains, respectively) and resolution (2.0Å). The length of the CDRs matches the loop length in the FabMNAC13 with the following exceptions: MNAC13_CDRL1 is shorter by one residue and MNAC13_CDRH3 is shorter by two residues, as shown in Figure 7.

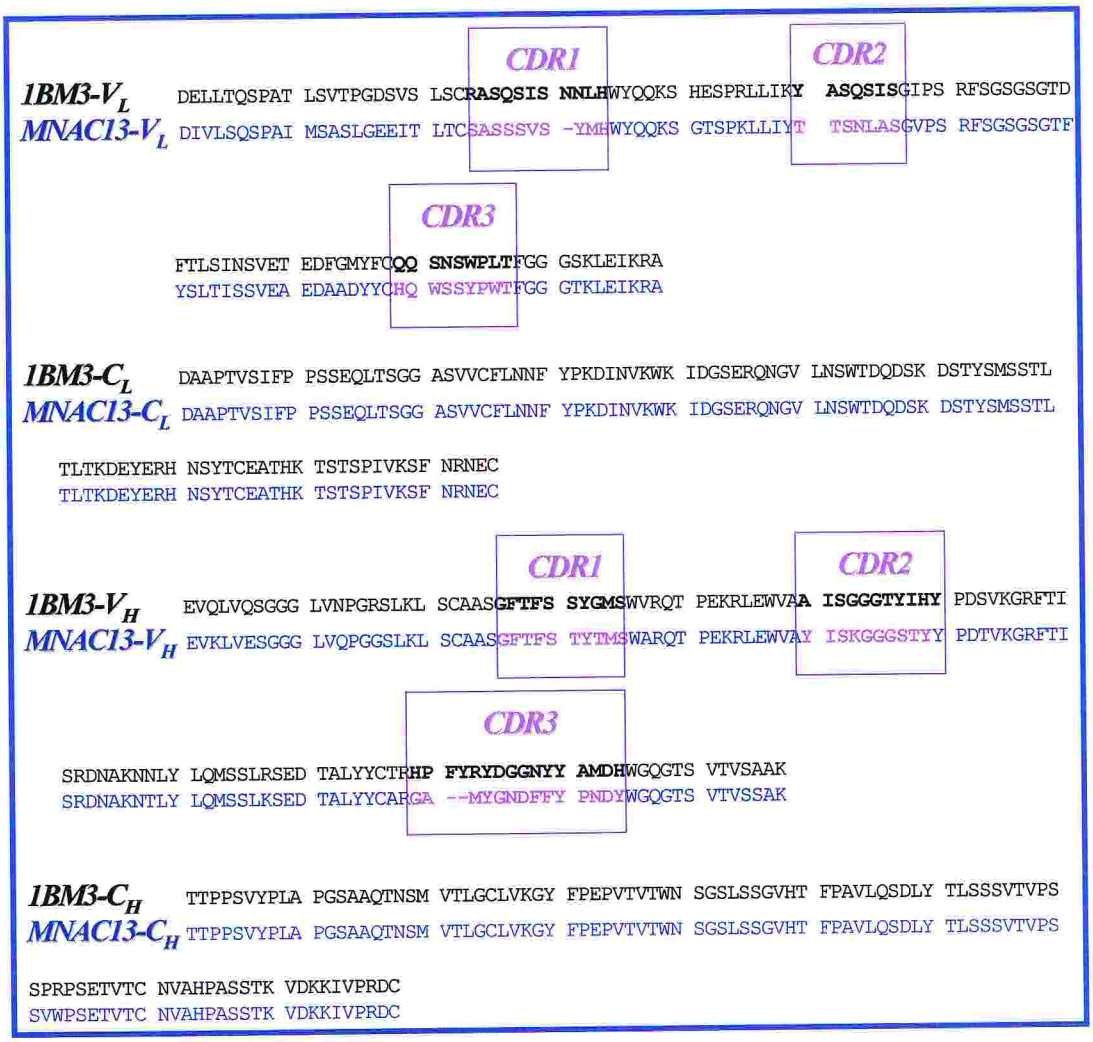


Figure 7: MNAC13-1BM3 alignment

Molecular replacement calculations, whose details are given in Table 1, used all data between 15.0Å and 3.5Å resolution obtained from low resolution data set collected at ESRF.

In particular it is worth noting the crucial importance of the elbow angle as it appears if one compares the results obtained in the two parallel searches. Indeed the search using the whole Fab as a model showed a quite poor correlation based on amplitudes of only 18.0% and after rigid body refinement an high value of the crystallographic R-factor of 51.4%. On the other hand the value of these parameters improved quite a lot using as a search model the constant and variable domains separately, resulting in 53.7% for the correlation and 39.8% for the crystallographic R-factor.

Figure 8 shows a quite evident difference between the two solutions positioned inside the unit cell:the whole Fab (light blue) is superimposed on the separate constant and variable domains (where the heavy and the light chain are drawn in green and in yellow respectively).

It is worth noting that such a disagreement is the result of a quite poor variation in the amplitude of the elbow angle between the model and the target.

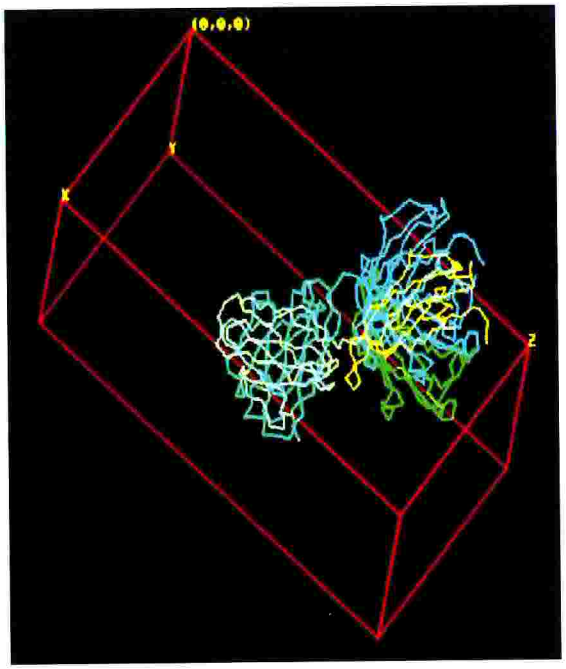


***Figure 8*: AmoRe solutions of Fab MNAC13**

Table 1: AMoRe Structure Solution

Cross-rotation and translation function found in the solution of Fab-MNAC13

A. *Fast rotation function*

Peak	α	β	γ	C_f	R_f	C_i	T_f	R_i
Whole Fab								
1	135.0	28.2	196.8	9.9	53.2	14.8	5.9	40.30
2†	112.7	81.6	248.4	9.8	53.6	14.0	3.4	
Variable domains								
1	110.7	20.4	141.1	12.3	52.2	20.9	11.6	35.05
2†	174.1	60.8	349.7	9.0	52.9	14.0	5.8	
Constant domains								
1	94.7	13.3	170.7	11.3	53.3	17.2	13.9	27.96
2†	12.2	87.4	317.3	9.2	54.0	13.7	7.1	

B. *Translation function*

Peak	x	y	z	C_f	R_f	C_i	T_f
Whole Fab							
	.3298	.1595	.3984	17.1	51.7	17.7	100
Variable domains							
1	.0974	.0775	.4674	28.9	48.3	30.6	100
2†	.0987	.2736	.4694	21.8	50.8	23.6	77.5
Constant domains							
1	.1675	.3102	.2357	23.1	50.6	21.6	100
2†	.1694	.2478	.2345	17.8	52.1	18.6	86
	.1783	.3130	.7349	(origin shifted)			
Variable and constant domains (individual orientation and position determined as above)							
	.0000	.0000	.0000	42.9	43.8	44.9	100

C. Rigid Body Refinement

Peak	α	β	γ	x	y	z	C_f	R_f	C_i	D_m
Whole Fab										
	134.4	28.7	197.0	.3330	.1616	.3986	18.0	51.4	18.3	36.9
Variable and constant domains										
	106.5	20.7	143.9	.1004	.0757	.4680				
	94.5	13.9	173.3	.1684	.3073	.7355	53.7	39.8	54.8	32.4

α,β,γ = Euler angles (°).
x,y,z = Translations (fractionary).
C_f = Correlation of amplitudes (x100).
R_f = Crystallographic R-factor (x100).
C_i = Correlation of intensities (x100).
C_p = Truncated Patterson correlation (x100).
R_i = Radius of integration (Å).
T_f = Fast-translation-function value (% of maximum value).
†The highest noise peak is shown for comparison.
D_m = Minimal CoM-CoM distance (Å) with current symmetry related and preceding positions.

◊ ***FabαD11 phasing***

A similar procedure was followed for the solution of the structure of FabαD11. Among its homologues obtained by a database search 1CIC, the structure of an idiotope-anti-idiotope complex FabD1.3-FabE225 (Bentley *et al.*, 1990), was chosen to perform molecular replacement. FabD1.3 appeared to be the most suitable model on the basis of resolution (2.5Å) and sequence identity (81.95 and 82.65% for the heavy and light variable domains, respectively). The length of the CDRs matches the loop length in the FabαD11 with the unique exception of FabαD11_CDRH3, that is six residues longer, as shown in Figure 9.

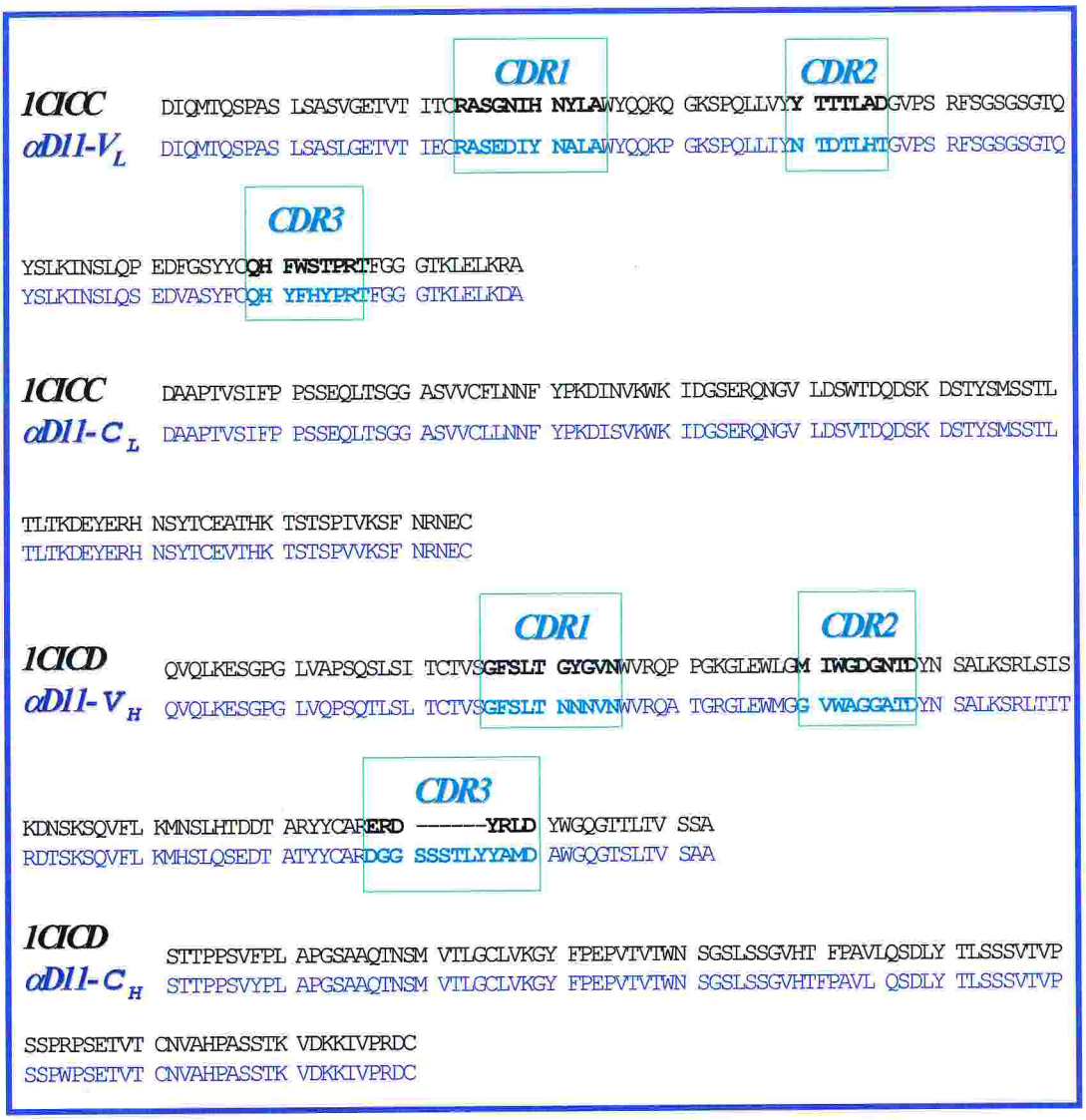


Figure 9: αD11-1CICC/1CICD alignment

Also in this case molecular replacement was performed by the program AmoRe, using both the whole Fab and the constant and variable domains separately as a probe models. In particular the variable domains comprised residues 1 to 109 of the light chain and 1 to 114 of the heavy chain, while the constant domains comprised the rest of the molecule (residues 110 to 214 of the light chain and 115 to 218 of the heavy chain). Molecular replacement calculations, whose details are shown in Table 2, used all data between 15.0Å and 3.5Å resolution obtained from a low

Table 2 - AMoRe Structure Solution

Cross-rotation and translation function found in the solution of Fab-αD11

A. Fast rotation function

Peak	α	β	γ	C_f	R_f	C_i	C_p	R_i
Whole Fab								
1	20.2	105.7	225.5	18.7	51.4	17.5	8.5	39.98
2	56.0	90.3	49.3	17.0	51.9	15.7	8.2	
3†	45.2	118.6	236.1	13.4	52.7	12.8	5.1	
Variable domains								
1	210.2	58.0	49.1	21.0	49.7	21.3	18.1	34.66
2	222.7	104.8	230.0	20.0	50.2	19.2	13.6	
3†	24.5	112.3	148.1	11.3	52.5	13.3	6.4	
Constant domains								
1	95.1	136.0	148.1	19.2	53.8	17.9	12.4	30.61
2	337.1	48.5	323.0	16.3	54.6	14.7	11.2	
3†	58.6	72.8	348.8	13.1	55.5	11.8	4.7	

B. Translation function

Peak	α	β	γ	x	y	z	C_f	R_f	C_i	T_f
Whole Fab										
1-Body Search										
1	20.2	105.7	225.5	.0000	.0000	.0000	18.7	51.4	17.5	8.5
2	56.0	90.3	49.3	.0000	.0000	.0000	17.0	51.9	15.7	8.2
2-Body Search										
1	20.2	105.7	225.5	.0000	.0000	.0000	18.7	51.4	17.5	42.7
2	56.0	90.3	49.3	.8220	.2486	.5047	26.0	49.0	24.7	100
1-Body Search										
Variable domains										
1	210.2	58.0	49.1	.0000	.0000	.0000	21.0	49.7	21.3	18.1
2	222.7	104.8	230.0	.0000	.0000	.0000	20.0	50.2	19.2	13.6

Constant domains

1	95.1	136.0	148.1	.0000	.0000	.0000	19.2	53.8	17.9	12.4
2	337.1	48.5	323.0	.0000	.0000	.0000	16.3	54.6	14.7	11.2

2-Body Search

Variable domains

1	210.2	58.0	49.1	.0000	.0000	.0000	21.0	49.7	21.3	42.7
2	222.7	104.8	230.0	.3892	.6275	.2787	33.0	46.5	33.8	100

Constant domains

1	95.1	136.0	148.1	.4607	.4119	.7952	28.8	48.1	28.5	100
2	337.1	48.5	323.0	.7676	.1041	.5095	27.9	48.8	28.1	100

3-Body Search

Variable domains

1	210.2	58.0	49.1	.0000	.0000	.0000	21.0	49.7	21.3	42.7
2	222.7	104.8	230.0	.3892	.6275	.2787	33.0	46.5	33.8	35.7

Constant domains

1	95.1	136.0	148.1	.4674	.4103	.7915	41.7	44.1	42.5	100
2	337.1	48.5	323.0	.7650	.1023	.5081	39.7	44.8	39.9	100
3†	337.1	48.5	323.0	.7051	.9285	.1123	31.1	47.1	31.1	30.9

4-Body Search

Variable domains

1	210.2	58.0	49.1	.0000	.0000	.0000	21.0	49.7	21.3	42.7
2	222.7	104.8	230.0	.3892	.6275	.2787	33.0	46.5	33.8	35.7

Constant domains

1	95.1	136.0	148.1	.4674	.4103	.7915	41.7	44.1	42.5	34.8
2	337.1	48.5	323.0	.7624	.1061	.5145	46.6	42.6	46.5	100
3†	337.1	48.5	323.0	.9083	.1344	.1050	38.8	45.4	39.5	30.9

α,β,γ = Euler angles (°).
x,y,z = Translations (fractionary).
C_f = Correlation of amplitudes (x100).
R_f = Crystallographic R-factor (x100).
C_i = Correlation of intensities (x100).
C_p = Truncated Patterson correlation (x100).
R_i = Radius of integration (Å).
T_f = Fast-translation-function value (% of maximum value).
†The highest noise peak is shown for comparison.

C. *Rigid Body Refinement*

Peak	α	β	γ	x	y	z	C_f	R_f	C_i	D_m
Whole Fab										
1	20.4	105.5	225.8	.0006	.0002	.0004	18.7	51.4	17.5	42.7
2	55.6	90.5	48.9	.8137	.2509	.5056	26.6	48.8	25.1	42.7
Variable domains										
1	211.6	59.2	49.9	-.0005	-.0028	.0003	21.0	49.7	21.3	42.7
2	223.5	105.3	229.6	.3932	.6321	.2799	33.0	46.5	33.8	35.7
Constant domains										
1	95.0	136.5	148.2	.4682	.4047	.7878	41.7	44.1	42.5	34.8
2	338.2	49.2	321.4	.7686	.0998	.5071	53.5	39.7	52.3	35.9

α,β,γ = Euler angles (°).
x,y,z = Translations (fractionary).
C_f = Correlation of amplitudes (x100).
R_f = Crystallographic R-factor (x100).
C_i = Correlation of intensities (x100).
D_m = Minimal CoM-CoM distance (Å) with current symmetry related and preceding positions.

resolution data set collected at ELETTRA on the Crystal Form II. Translation function searches were performed for the orientations identified from the rotation function. From these searches, a full Fab model was assembled and adjusted with rigid body refinement. The only difference respect to the protocol followed in the solution of the structure of FabMNAC13 is due to the presence of two molecules in the asymmetric unit in the case of FabαD11 instead of the only one in the case of FabMNAC13. As a consequence two solutions, related by a two-fold non crystallographic symmetry (NCS), were found in the triclinic space group for FabαD11.

It is interesting to note that in this second case the difference in elbow angle is not so large as between FabMNAC13 and the respective employed searching model. Indeed the results obtained using the whole Fab as a search model were quite similar in terms of correlation based on

amplitudes and the crystallographic R-factor to the ones obtained using separately the constant and variable domains.

The molecular replacement method was performed again for phase calculation and structure solution of the FabαD11Crystal Form III, using all data between 15.0Å and 3.5Å resolution from an high resolution data set collected at ESRF. In this second case the probe model was the solution obtained by the previous run with AmoRe after a preliminary structure refinement. Indeed as exposed in details in the next chapter, low resolution structure of Crystal form II was partially refined: non homologous residues were mutated, while the less conserved (CDRH3) or poorly defined loop regions were completely deleted to minimize model bias.

Table 3 shows main details of molecular replacement calculations. Taking in consideration that the elbow angle shows a great variability not only among different Fab fragments but among the different crystal forms of the same Fab fragment, molecular replacement was performed, using the constant and variable domains separately as a search models.

Table 3: AMoRe Structure Solution

Cross-rotation and translation function found in the solution of Fab αD11

A. Fast rotation function

Peak	*α*	*β*	*γ*	C_f	R_f	C_i	C_p	R_i
Variable domains								
1	150.7	155.3	41.5	17.0	51.1	19.9	17.1	31.59
2†	45.8	85.1	272.5	6.5	53.9	7.9	5.9	
Constant domains								
1	19.3	64.4	73.4	11.8	54.5	12.3	15.9	27.16
2†	3.5	69.6	118.6	5.9	56.0	6.4	5.2	

B. Translation function

Peak	*x*	*y*	*z*	C_f	R_f	C_i	C_p
Variable domains							
1	.1433	.000	.4471	30.6	47.8	29.1	100
2†	.4943	.000	.4524	20.7	50.5	18.9	64.9
Constant domains							
1	.3617	.000	.1957	21.8	51.6	18.4	100
2†	.4965	.000	.4928	15.0	53.4	13.2	71.2

C. Rigid body refinement

Peak	*α*	*β*	*γ*	*x*	*y*	*z*	C_f	R_f	C_i	C_p
Variable and constant domains										
	151	155.4	43	.1424	.0005	.449				
	17.8	63.7	73.2	.3625	.9532	.1991	55.0	38.9	49.7	35.9

α,β,γ = Euler angles (°).
x,y,z = Translations (fractionary).
C_f = Correlation of amplitudes (x100).
R_f = Crystallographic R-factor (x100).
C_i = Correlation of intensities (x100).
D_m = Minimal CoM-CoM distance (Å) with current symmetry related and preceding positions.
C_p = Truncated Patterson correlation (x100).
R_i = Radius of integration (Å).
T_f = Fast-translation-function value (% of maximum value).
†The highest noise peak is shown for comparison.

References

Beauchamp, J. C., Isaacs, N. W. (1999). Methods for X-ray diffraction analysis of macromolecular structures. *Curr Opin Chem Biol*; **3**, 525-529.

Bentley, G. A., Boulot G, Riottot MM, Poljak RJ. (1990). Three-dimensional structure of an idiotope-anti-idiotope complex. *Nature*; **348**, 254-257.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*. **28,** 235-242.

Kodandapani, R., Veerapandian, L., Ni, C. Z., Chiou, C.-K., Whittal, R. M., Kunicki, T. J. & Ely, K. R. (1998). Conformational change in an anti-integrin antibody: structure of OPG2 Fab bound to a beta 3 peptide. *Biochem. Biophys., Res. Comm.*; **251**, 61-66.

Navaza, J. (1994). AMoRe: an automated package for Molecular Replacemet. *Acta Cryst.*; **A50**, 157-163.

CHAPTER 9: Fitting, refinement and validation

Starting from the initial phases, derived by molecular replacement solution, electron density Fourier maps with coefficients $2F_o$-F_c and F_o-F_c respectively have been calculated.

Thereafter an iterative procedure, consisting in two alternate steps, has been employed as follows:

- Fitting the model to the maps by manual model building, using interactive computer graphics software "O" (Kleywegt and Jones, 1994);
- Refinement of model using protocols as implemented in the CNS suite, Crystallography and NMR System (Brünger *et al.*, 1998). Hence the schematic layout of the procedure comprised:

1. ***RIGID BODY MINIMIZATION***

That included three steps: first the whole Fab was treated as an independent rigid body, then the heavy and light chains have been individually refined. Finally the variable and constant domains in either of the heavy and light chains were allowed to adjust their positions.

2. ***SIMULATED ANNEALING + POSITIONAL REFINEMENT + B-FACTOR REFINEMENT***

Then the model was subjected to simulated annealing molecular dynamics (Brünger *et al.*, 1997) according to the slow cooling protocol described in the CNS manual.

As a result the calculated electron density Fourier maps allowed to identify the most prominent features (such as side chain mutations, insertions and deletions), that have been applied to the initial model used for the molecular replacement search. Indeed the step of model building was guided by the alignment between the amino acidic sequences of the model and the target in order to mutate the residues, that did not match, and find out regions, whose lengths were not conserved. In particular among them the CDRH3 loop of the heavy chain, being the most variable, required major interventions.

Several alternated iterative cycles of refinement (positional and B-factor) and manual building allowed the model to be completed.

At this stage an automated water picking protocol was able to localize not only water molecules, but also ions and ligands that were fitted into electron density Fourier maps densities and included in the very last cycles of the refinement.
Finally keeping the model in terms of stereochemical constraints as close as possible to ideal, an optimisation of the positional weight *wa* and the B-factor weight *r-weight* was performed, followed by cycles of alternating B-factor and positional refinement. This step was necessary to better take into account the X-ray crystallographic term.

In general it is worth noting that fitting a model into an electron density Fourier map, refining it to agree with the diffraction amplitudes, and

validating the result are closely intertwined processes whose crucial limit is under-determination. However during the refinement the phase information computed from the atomic model gradually becomes more and more accurate. In order to monitor this progress during refinement, the agreement between the observed and calculated structure factors can be expressed by the conventional disagreement factor:

$$R = \Sigma(|F_o - F_c|)/\Sigma(F_o)$$

The target function used by CNS for minimization is:

$$E_{tot} = w_a E_{effective} + E_{empirical}$$

where the $E_{effective}$ comprises the X-ray term and the $E_{empirical}$ contains the stereochemical restraints, such as bond lengths, bond angles and van der Waals contacts, torsional, packing, impropers (chirality) (Engh and Huber, 1991). During minimization of the energy function, the problem of data overfitting can be circumvented by the cross-validation procedure (Brünger, 1992). Most of the data (working set) are used for the refinement of the atomic model and to compute the conventional R-factor, while the remaining data (test set) are flagged and used to compute the R_{free}-factor in the same way as the conventional R-factor but using only a subset of data (usually 10%) that instead is not used during the refinement. Therefore both R and R-free factor should smoothly decrease at every stage until convergence has been reached.

Fab MNAC13

Fitting and refinement

As mentioned before in order to follow the progress of the refinement and to allow cross-validation 10% of the reflections were flagged, both the R and R_{free} were monitored (Figure 1).

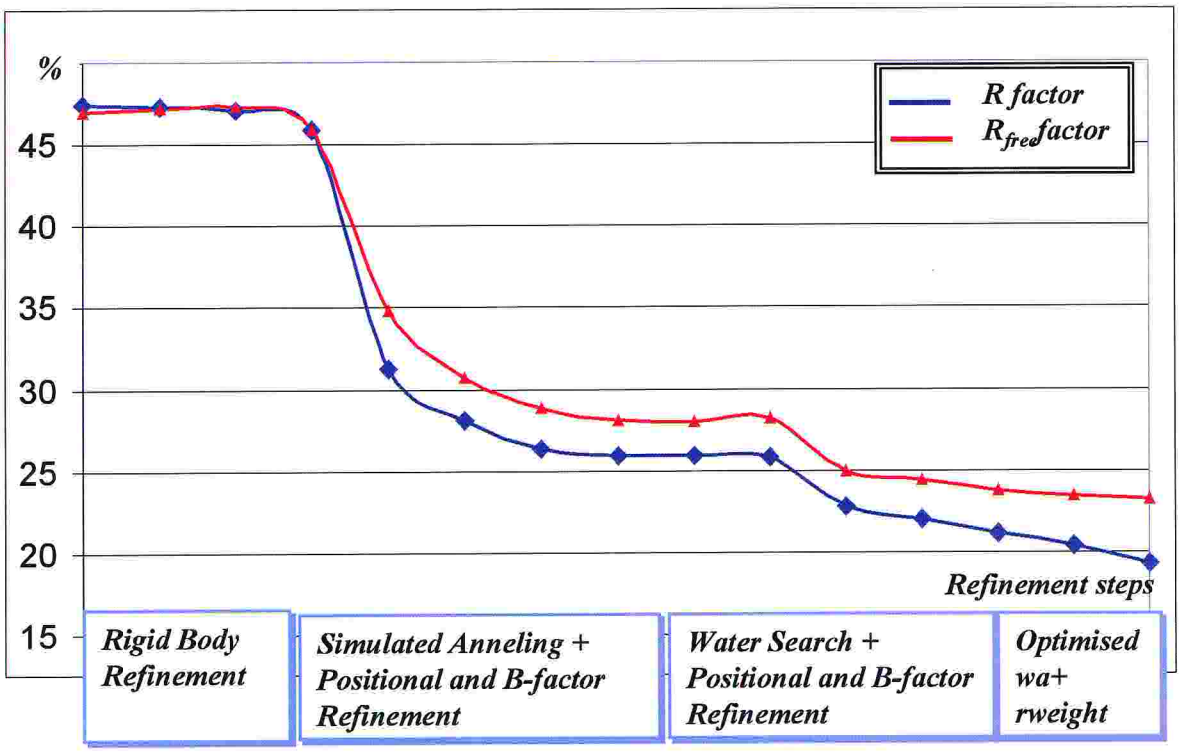


***Figure 1*: R and R_{free} factor plots of Fab MNAC13 structure solution**

In Figure 2 the sequence alignment between Fab MNAC13 and the model, used in molecular replacement, is reported. Residue numbering and definitions of the CDRs follow the conventions of Kabat *et al.* (1991). Residue numbers are preceded by L or H for the light and heavy chains respectively.

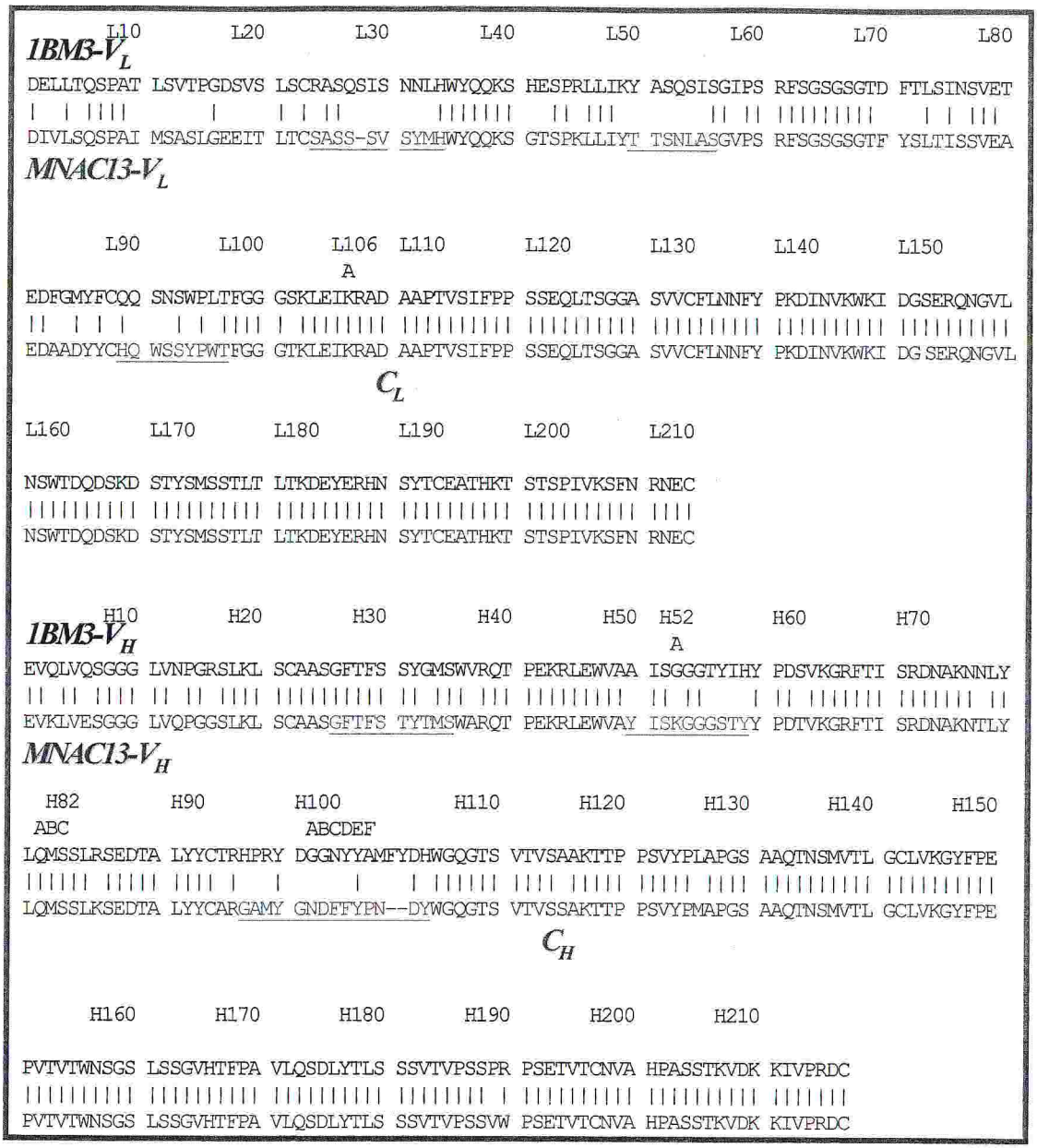


Figure 2: Target/model alignment for FabMNAC13

Regarding the first step of model building it is interesting to note that unfortunately in the initial model, used in molecular replacement, the structure of the all CDR3 loop of the heavy chain was absent. In addition in two other loops the experimental density showed a quite bad agreement with the model and was also quite discontinuous and broken.

The in initial electron density Fourier maps also revealed that the loop regions encompassing residues L196-L204 of the light chain and residue H129-133 of the heavy chain were ill defined and thus were omitted in the following refinement steps. As the refinement proceeded and phases improved, the missing regions, mentioned above, could be eventually traced.

Validation

The entire molecular structure of the MNAC13 Fab fragment was refined at high resolution (1.8Å), which is notable since many Fab crystals diffract on average up to 2.7-2.8-Å resolution.

The final model consists of 3244 protein non hydrogen atoms (comprising 208 amino acids in the light and 221 amino acids in the heavy chain, with four intramolecular disulphide bonds: L22-L88 and L134 L194 for the light chain and H22 H93 and H142 and H197 for the heavy chain) and 351 solvent atoms, four sulphate ions, one Tris molecule and one isopropanol molecule.

The final refinement statistics and parameters describing the quality of the model are presented in Table 1.

The average B-factor for all the protein atoms is 23.55Å^2. It is worth noting that all the CDRs but CDRH3 showed average temperature factor below this value. Beside its higher mobility nevertheless the electon density of the polypeptide backbone corresponding to the CDRH3 was clearly defined as well as the side chains, except for Asn H101.

Table 1: Quality of refined structure of Fab MNAC13:

Refinement statistics

Resolution range	(Å)	39-1.778
Final R factor		19.35%
Final R_{free} factor	**(10% of data)**	23.22%
Rms deviations		
bond lenghts (Å)		0.008
bond angles (°)		1.456
dihedral angles (°)		27.29
improper angles (°)		0.928
Average isotropic B factors (Å^2)		
All protein		23.55
Light chain		24.14
Heavy chain		22.99
L-CDRs	L1	18.79
	L2	22.40
	L3	17.93
H-CDRs	H1	23.05
	H2	23.33
	H3	34.27
Water molecules		31.95
Ions (sulphate)		55.94
Tris		46.06
Isopropanol		32.60
% of main-chain dihedral angles in Ramachandran plot are in:		
fully allowed regions		91.2%
additonal allowed regions		8.2%
generously allowed regions		0.3%
disallowed regions		0.3%

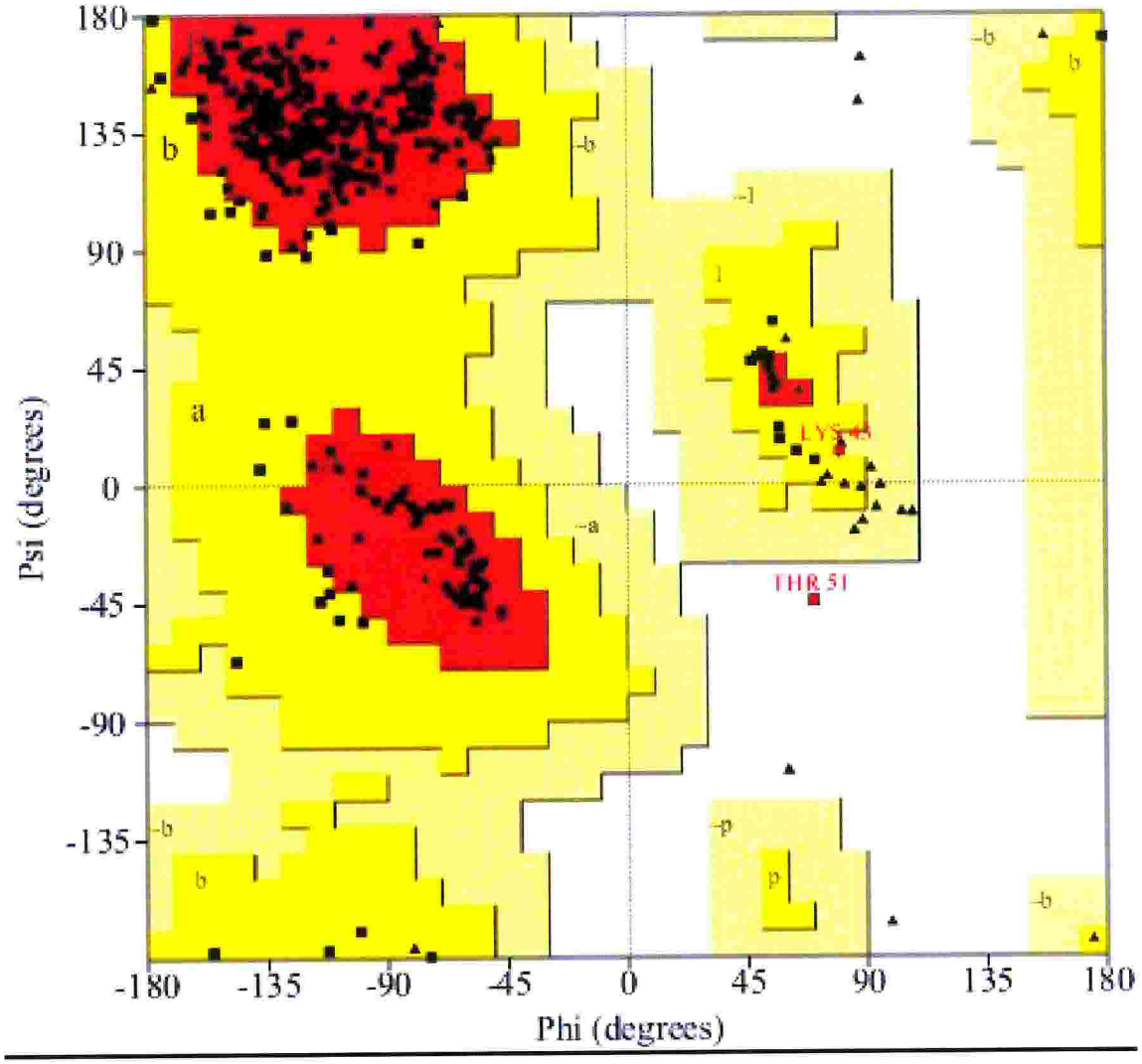


***Figure 4**: Ramachandran plot of main chain torsion angle(ϕ/ψ) For light and heavy chains of Fab MNAC13*

The final model geometry was analyzed with PROCHECK (Laskowski *et al.*, 1993). The only residue, that was located in disallowed region of the Ramachandran plot (Figure 4), was Thr L51. It is worth noting that these unfavorable ϕ/ψ values (70.1; −44.2) can be explained by the fact that this residue takes a part in a classic γ-turn. This seems to be a common feature in the majority of the V_L domains in the PDB database (Honegger, unpublished data).

The final electron density maps were of high enough quality in most regions such that we could identify previous DNA sequencing errors within the heavy and light chains. Indeed in the following six positions (three in the light and three in the heavy chain) the maps clearly were not consistent with the residues that had been assumed on the basis of the cDNA sequence. Therefore the following appropriate changes were made according to the electron density map:

L5	**Ser**	⇒	**Thr**	**H99**	**Tyr**	⇒	**Phe**
L19	**Ile**	⇒	**Val**	**H101**	**Asn**	⇒	**Met**
L45	**Lys**	⇒	**Val**	**H104**	**Tyr**	⇒	**Arg**

Two carboxyl terminal residues of both chains (L211-L212 and H216-217), few residues (L198-L200 in the constant part of the light chain and H130-H133 in the constant part of the heavy chain) and the side chain od Asn100A) were omitted from the final model because of their weak electron density.

Moreover electron densities for disulphide bonds were ill defined and these residues have been modeled to be in their most probable conformations. Besides that in many cases the side chains of exposed acidic residues were poorly defined (respectively Glu L18, L81, L105, L122 and Asp L142 in the light chain and Glu H1 in the heavy chain).

Both observations could be explained on the of radiation damage (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000). Indeed the principal highly specific effects of X-ray irradiation result in breakage of disulphide bonds and loss of the carboxyl groups of acidic residues.

Fab αD11

Fitting and refinement

As mentioned before two crystal forms for Fab αD11 were obtained and two data sets were collected (a low resolution one for Crystal form II -triclinic P1- and an high resolution one for Crystal form III -monoclinic C2-), processed to solve the phase problem and sequentially refined.

- ***Crystal form II preliminary fitting and refinement***

Regarding molecular replacement performed on the low resolution data set, both the rotational searches (using the entire Fab fragment and the segmented Fab respectively) resulted in a pair of clear solutions with correlation coefficients significantly higher than those of the next highest solutions (see Chapter 8). These two solutions in the triclinic space group were related by a two-fold non-crystallographic symmetry (NCS).

As mentioned before, after structure solution, several steps of rigid body refinement were performed. In particular before the segmentation of the model, an additional step was included to refine first of all both Fab fragments together by rigid body minimization.

Afterwards a $2F_o$ -F_c map was calculated and the molecular packing was checked showing no conflicts between molecules related by translational symmetry. The elbow regions connecting the variable and constant domains of both the light and heavy chains could be easily discerned in the electron density map, yielding an intact molecule for structure refinement.

After rigid body refinement, that reduced R factor value to 37.39%, one round of torsional angle molecular dynamics using strict CNS were

applied, followed by two steps of simulated annealing, positional and B-factor refinements.

Extensive manual model rebuilding was performed on *2Fo-Fc* maps at this stage using O. Most of the residues, that did not match between the model and the target (Figure 5), were mutated and poorly defined regions

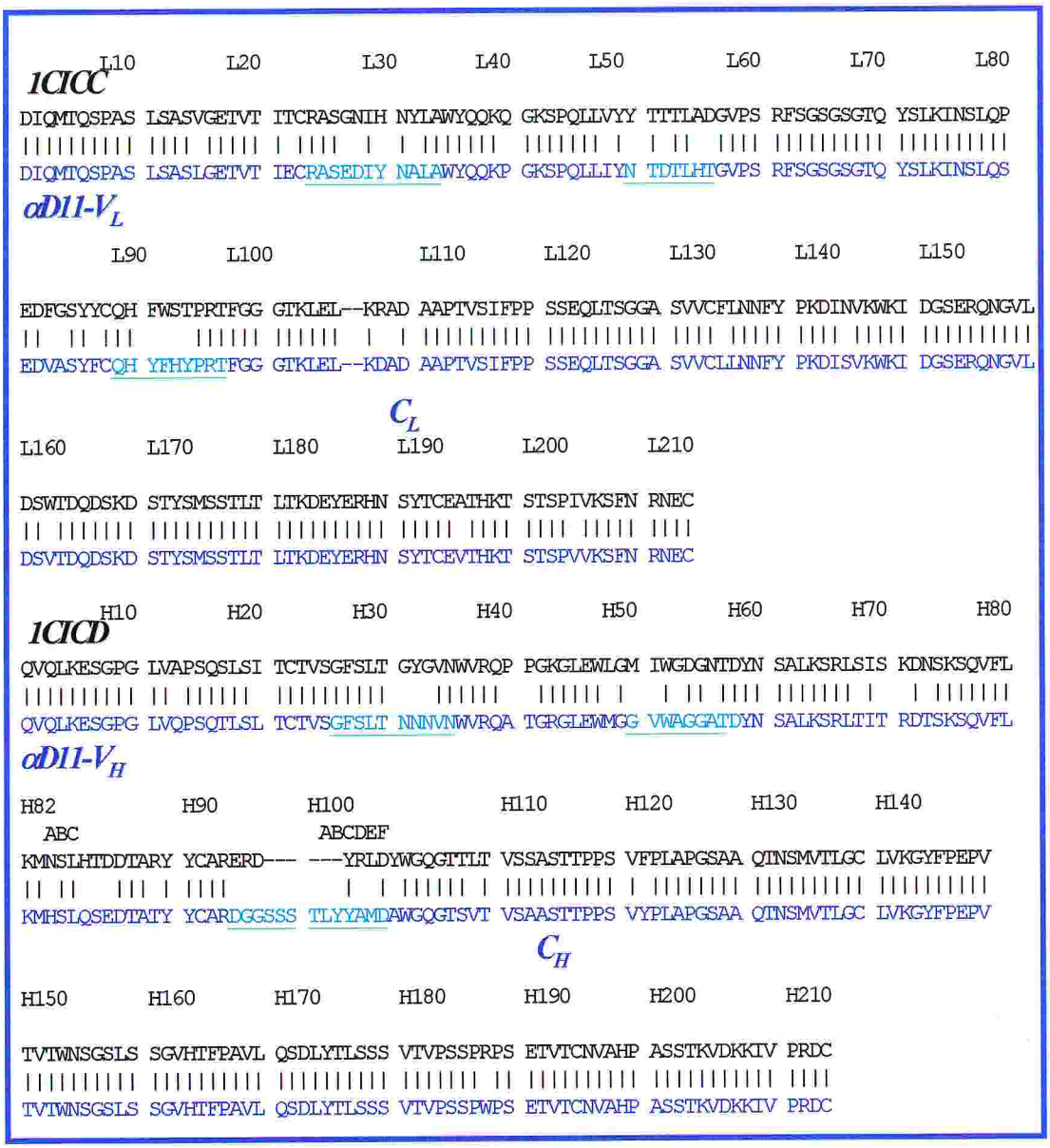


Figure 5: Target/model alignment for Fab αD11

(H126-H132) and CDRH3 loop (that was completely different both in length and in sequence) were deleted in both molecules.

In the next cycle of refinement tight NCS restraints were introduced and B-factors were refined individually, leading to a decrease in the R factor value to 34.74%.

The overall electron density was essentially unambiguous for the trace of the main chains (except for the regions mentioned above and for the carboxyl termini of both domains) and the general positions of the side chains. As it can be expected for data at this resolution, the details of the side chain conformations and the torsion angle of some of peptide oxygen atoms were not well determined.

Fitting and refinement of this model was stopped at this early stage because in the meanwhile an higher resolution data set was collected at ESRF on a different crystal form.

Nevertheless as mentioned before this partially refined model was employed as a searching model for the molecular replacement with the higher resolution data set.

- ***Crystal form III fitting and refinement***

As mentioned above for Fab MNAC13 refinement, also the progress of Fab αD11 refinement was followed monitoring the behavior of the R and R_{free} (shown in Figure 5). After molecular replacement ant three steps of rigid body refinement (analogously to the protocol performed for Fab MNAC13), a *2Fo-Fc* map was computed. There was clear density for most of side chains, that had not previously mutated; and especially for the excised CDRH3 segments.

Instead of a manually building of the excised regions into the structure on the basis of the map, automatic refinement and model building were performed by the program *ARP-wARP* (Lamzin and Wilson, 1993, 1997),

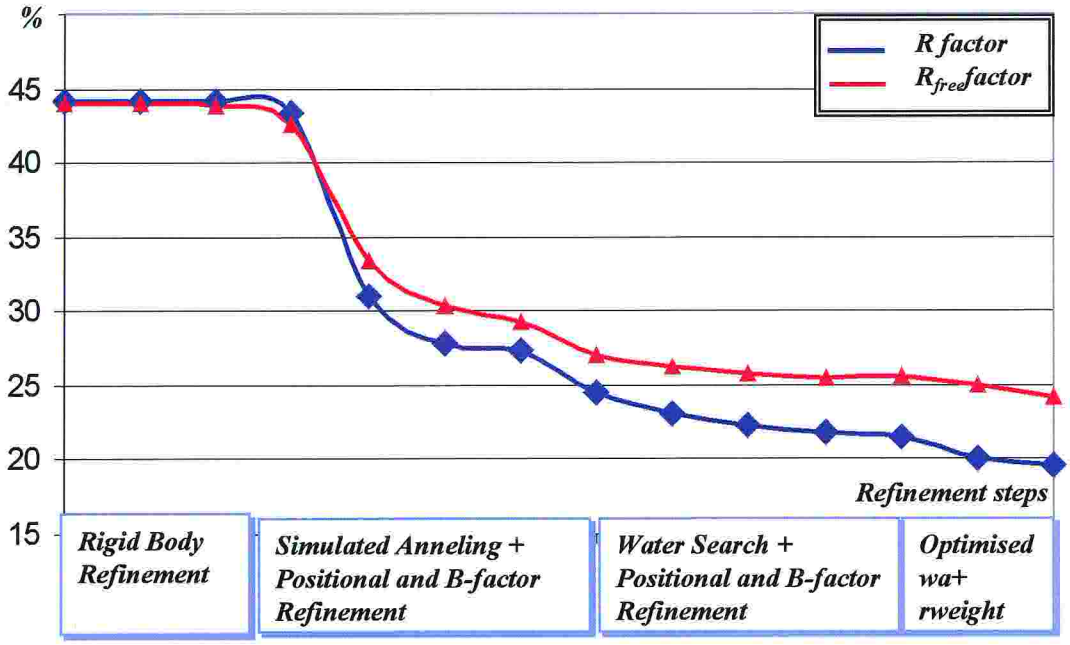


Figure 5: R and R_{free} factor plots of Fab αD11 structure solution

Considering the high resolution and the quality of this second data set, it was possible to use successfully this program, allowing a complete rebuilding of the CDRH3 loop with a perfect fit of model to map. On the other hand the electron density for the second missing regions (H126-H132) was too week to obtain any result.

Crystal form III validation

The final model of Fab αD11 was composed of 3229 protein atoms, (comprising 213 amino acids in the light and 216 amino acids in the heavy chain, with four intramolecular disulphide bonds: L23-L88 and L133 L193 for the light chain and H22 H92 and H139 and H194 for the heavy chain) plus 403 water molecules and one chloride anion.

Overall refinement statistics and final model parameters are given in Table 2.

Table 2: Quality of refined structure of Fab αD11:

Refinement statistics

Resolution range	(Å)	30-1.7
Final R factor		19.54%
Final R_{free} factor	**(10% of data)**	24.22%
Rms deviations		
bond lengths (Å)		0.0096
bond angles (°)		1.6571
dihedral angles (°)		27.40
improper angles (°)		1.048
Average isotropic B factors (Å^2)		
All protein		25.58
Light chain		24.14
Heavy chain		22.99
L-CDRs	L1	23.87
	L2	18.33
	L3	20.16
H-CDRs	H1	22.32
	H2	18.96
	H3	22.15
Water molecules		38.80
Ions (chloride)		20.58
% of mainchain dihedral angles in Ramachandran plot are in:		
fully allowed regions		91.6%
additonal allowed regions		7.6%
generously allowed regions		0.5%
disallowed regions		0.3%

The average B-factor was 25.58Å^2 for all atoms in the protein. As for FabMNAC13, also in this second Fab structure all the residues that had B-factors greater than this value were concentrated at the level of the carboxyl and amino termini of both chains and at the level of the segment H126-H132 in the heavy chain, where the electron density was very weak. Moreover considering that in general there is a good correlation between the B-factor and the secondary structure of the chain, it is worth noting that all six CDRs showed temperature factors considerably below the average for the all protein (the lowest for CDRL2 and CDRH2). In particular B-factor for CDRH3 loop was inside the range of those for the other hypervariable regions; indeed both the polypeptide backbone and the position of the side chains were quite well defined and could be traced with high confidence. This finding could be an index that it is very likely that this loop was quite ordered, less exposed and less mobile, especially compared with the homologous region in Fab MNAC13. As a consequence of that it might undergo a less significant conformational change upon antigen binding.

The final model geometry was analyzed with PROCHECK (Laskowski *et al.*, 1993). As for Fab MNAC13 regarding the Ramachandran plot (Figure 6) only Thr L51 (0.3% of total) had ϕ/ψ angle combinations (72.8; –47.3) that fell just outside of allowed regions, due to its position at the level of a classic γ-turn.

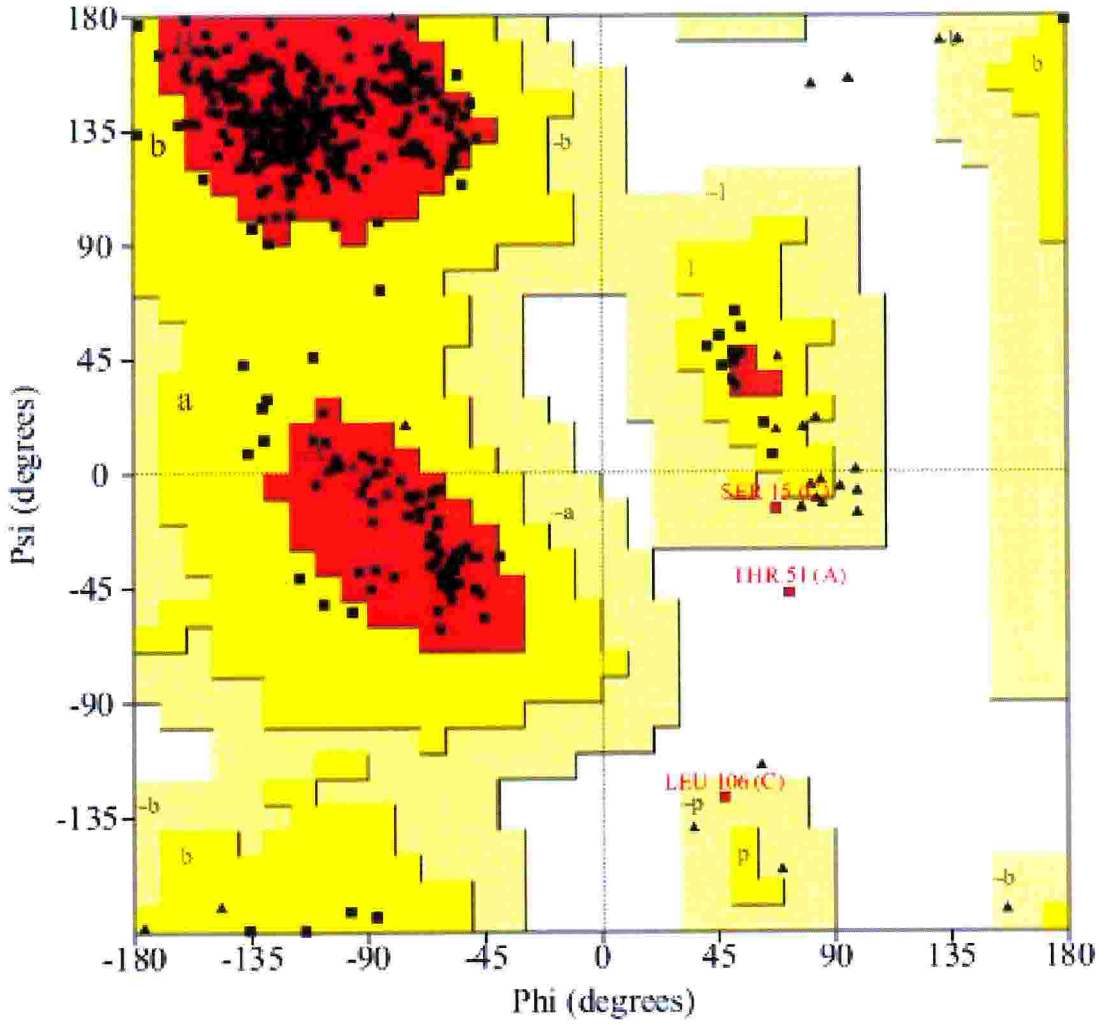


***Figure 6**: Ramachandran plot of main chain torsion angle(ϕ/ψ) For light and heavy chains of Fab αD11*

The final electron density maps were of high quality in most regions such that also in for this second Fab fragment it was possible to identify previous cDNA sequencing errors. In particular in the following four positions (three in the light and one in the heavy chain), appropriate changes, reported below, were performed according the electron density (that was not consistent with the residues that had been assumed on the basis of the cDNA sequence):

L17	**Glu**	**⇒**	**Lys**	**L93**	**His**	**⇒**	**Gly**
L76	**Ans**	**⇒**	**Gln**	**H16**	**Gln**	**⇒**	**Asn**

On the other hand as reported above the electron density of certain regions within the constant domains was too weak to reliably trace the polypeptide chain and thus these regions were not included in the final model. They consist in the two carboxi terminal residues of both chains (L212-L213 and H212-H213) and an extended disordered loop (H126-H132), connecting strand A and B in the constant part of the heavy chain (C_{H1}). It is worth noting that these disordered regions are located in corresponding positions of the missing residues in the final model of FabMNAC13, even if in this second structure a whole loop is missing. Moreover these situations are not uncommon in Fab models and in particular the same loop in of the C_{H1} domain is poorly defined in many Fab structures and several author have reported the complete absence of electron density for the corresponding region of their structures (Whitlow *et al.,* 1995, Barbas *et al.,*1997). Poor tracing of this part of the C_{H1} domain probably result from dynamic or static disorder in this particular loop (Banfield *et al.,*1996) or possible rearrangements of disulphide bridges within the C_{H1} domain (Whitlow *et al.,* 1995). In Fabα D11 structure the region of the molecule containing the loop borders a solvent channel. Any intermolecular contacts that could stabilize a defined conformation are thus absent.

As for FabMNAC13, electron densities for disulphide bonds were often poorly defined as well as the carboxylic group of exposed acidic residues (respectively Glu L22, L81, L105 and Asp L142 in the light chain and Asp H72 and H 94 and Glu H85 in the heavy chain) probably due to radiation damage.

References

Banfield, M., King, D., Mountain, A, Brady, R. L. (1996). Structure of the Fab Fragment of a Monoclonal Antibody Specific for Carcinoembryonic Antigen. *Acta Crystallogr.;* **D52**, 1107-1113

Barbas, C. F. 3rd, Heine, A., Zhong, G., Hoffmann, T., Gramatikova, S., Bjornestedt, R,. List, B., Anderson J, Stura EA, Wilson IA, Lerner, R. A. (1997). Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science*; **278**, 2085-2092

Brünger, A. T., (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Cryst.;* **D54,** 905-921

Brünger, A. T., Adams, P. D. and Rice, L. M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*; **5**, 325-336.

Brünger, A. T. (1992). Free R-value – a novel statistical quantity for assessing the accuracy of crystal structures. *Nature;* **355**, 472-475.

Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr.;* **D56**, 328-341.

Engh, R. A. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.;* **A47**, 392-400.

Kleywegt, G. J. and Jones, A. T. (1994). Checking your imagination: application of the free R-value. *Structure*; **2**, 1241-1258.

Lamzin, V. S. and Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr.***D49**, 129-147

Lamzin, V. S. and Wilson, K. S. (1997). Automated refinement for protein crystallography. *Methods in Enzymol.;* **277,** 269-305.

Laskowski, L. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.;* **26,** 283-291.

Ravelli, R. B. G. & McSweeney, S. M. (2000). The 'fingerprint' that X-rays can leave on structures . *Structure*; **8**, 315-328.

Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J., Sussman, J. L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc Natl Acad Sci U S A.;* **97**, 623-628.

Whitlow, M., Howard, A. J., Wood, J. F., Voss, E. W. Jr, Hardman, K. D. (1995). 1.85 A structure of anti-fluorescein 4-4-20 Fab. *Protein Eng*; 8, 749-761

CHAPTER 10: FabMNAC13 structure

Overall folding

The MNAC13 Fab fragment assumes the overall immunoglobulin folding as observed in the other solved Fab fragments structures. A schematic ribbon diagram pattern, created by the program *VMD*, is shown in Figure 1. The light chain and the heavy chain are colored in cyan and in yellow respectively. Variable and constant domains are located at the top and at the bottom of the diagram respectively.

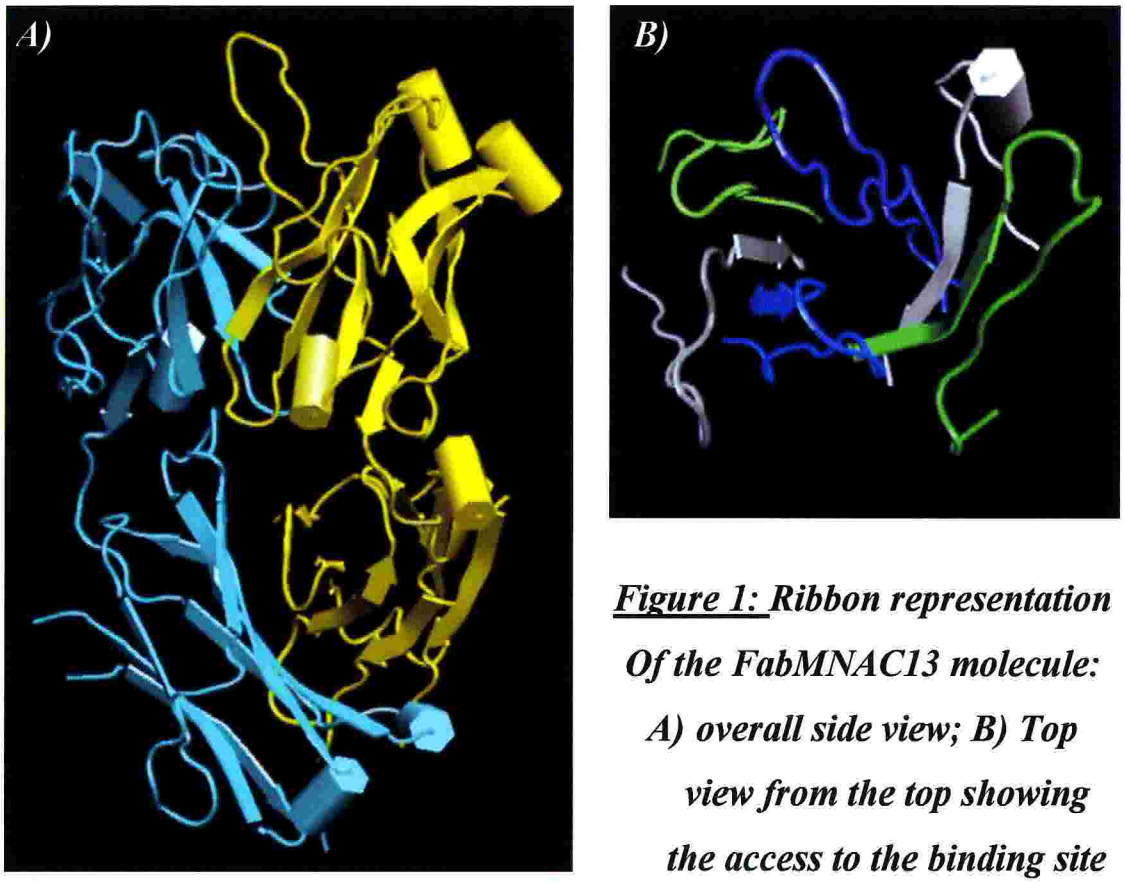


***Figure 1:* Ribbon representation Of the FabMNAC13 molecule: A) overall side view; B) Top view from the top showing the access to the binding site**

Overall the Fab fragment comprises four typical immunoglobulin domains(Figure 1A), two each from the light (V_L-V_H) and heavy chains

(C_L-C_{H1}). Each domain is characterized by a β-barrel structure consisting of β-sheets connected by an internal disulfide bridge.

The determined orientation of the two interchain domain pairs with respect to each other is described by the elbow angle. It is defined as the angle formed by the pseudo two fold axis relating the V_L-V_H and the C_L-C_{H1} domains. For FabMNAC13 its value is of 118°, that is quite out of the range observed for other Fab fragments (Padlan, 1994; Wilson and Stanfield, 1994).

Crystal packing

The two MNAC13 Fab molecules in the unit cell are packed as shown in Figure 2 where the molecules are represented by a Cα trace (light chains in blue and heavy chains in white).

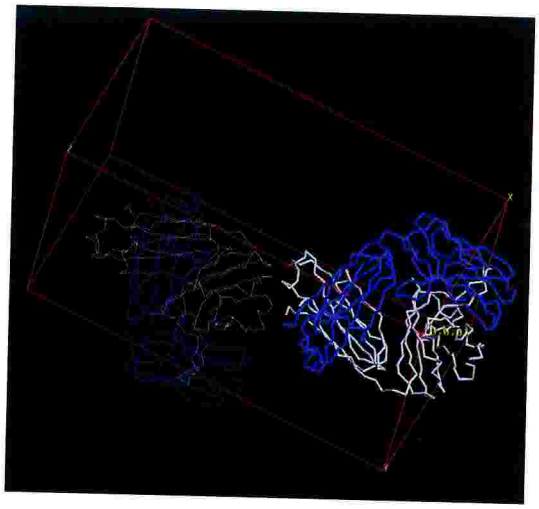
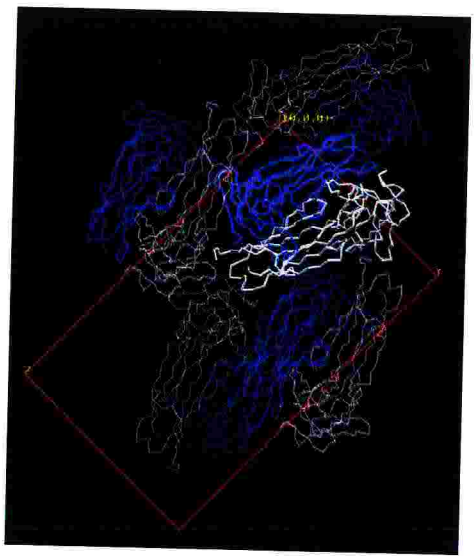


***Figure 2** FabMNAC13 packing inside the unit cell*

Packing views, along the *b* and *c* axis respectively, are shown in Figure 3.



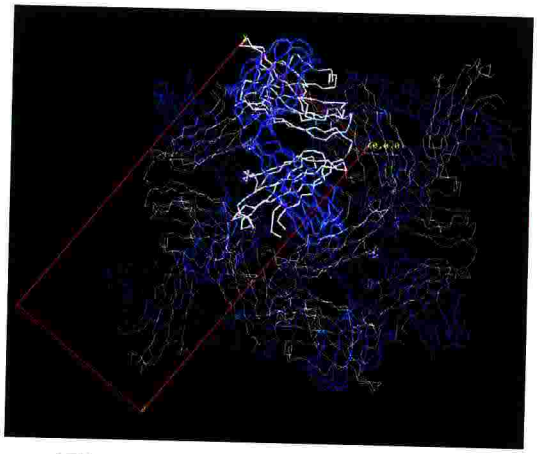


***Figure 3** Perspective views of packing seen from b and c axis*

FabMNAC13 Antigen Binding site

The six surface loops (shown in Figure 1B) extend from the N-terminal variable domains of both the heavy and the light chains and, due to their high variability in sequence, are called CDRs (Complementarity Determing Regions). Results from the crystallographic analysis of many antibody-ligand complexes strongly suggest that the binding site of an antibody is primarily composed by CDR residues, although on rare occasions neighboring framework residues have been found to be involved also. Thus the CDR surface of MNAC13 most probably portrays the topography of its combining site.

The conformations of these hypervariable loops can be classified according to the observations of Chothia and Lesk (1987). Four out of six CDRs of MNAC13 are canonical:

◊ CDRL1 (Figure 4) belongs to the canonical group 1, and extends across the top of the β-sheet framework with the Val L29 packed inside a deep cavity formed by residues Ile L2, Met L33 and Tyr L71. These residues are the major determinants of this conformation.

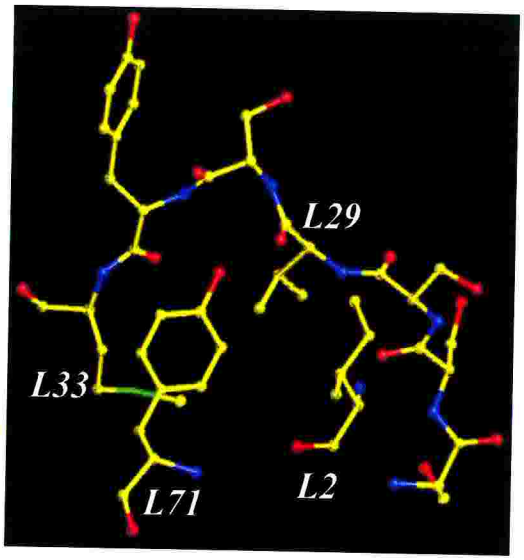


***Figure 4*: Canonical conformation of CDRL1**

◊ CDRL2 (Figure 5A) belongs to group 1 (the only group identified so far), joining two adjacent framework strands. Due to the presence of a hydrogen bond between Tyr L49 and Asn L53 (2.8Å), CDRL2 results in a three residues hairpin with all the side chains pointing towards the surface.

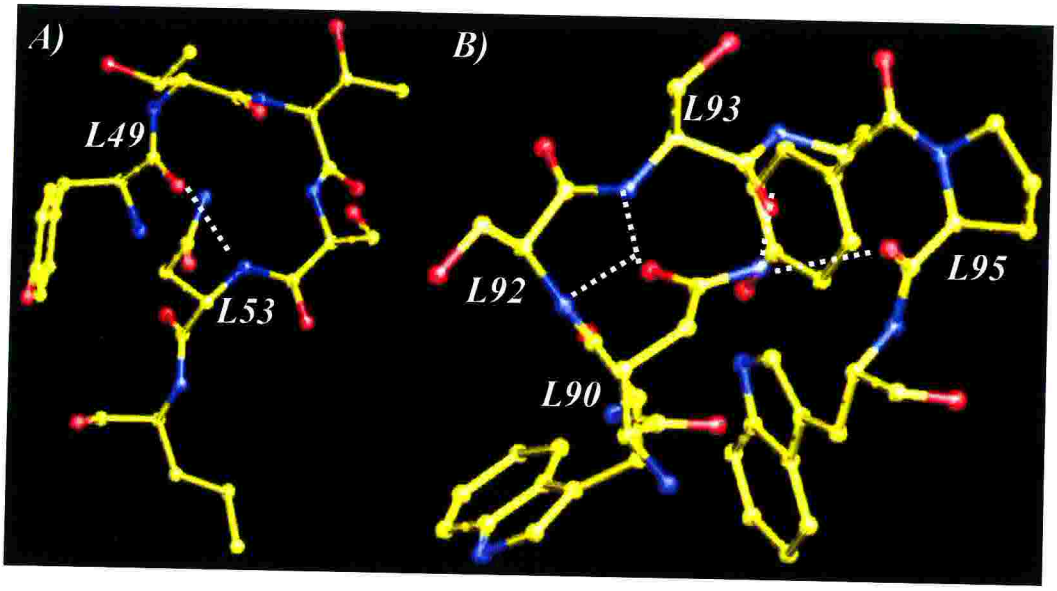


***Figure 5**: Canonical conformation of CDRL2 (A) and CDRL3 (B)*

◊ CDRL3 (Figure 5B) belongs to group 1, whose extended conformation is due to Pro L95 with a *cis* peptide and the α_L conformation of residue L92. Furthermore the NH_2 side chain moiety of the framework residue Asn L90 is involved in two hydrogen bonds with the main chain carbonyls of residues L93 (3.10Å) and L95 (3.17Å). the carbonyl moiety L90 is instead involved in hydrogen bonding with the amide residues of L93 (3.01Å) and of L92 (2.84Å).

◊ CDRH1 (Figure 6) follows the conformation of the canonical group 1, characterized by a sharp turn due to Gly H26. The residue Phe H29 deeply buried within the framework structure and is packed against the side-chains of Met H34 and the main chain of residues H72 and H77.

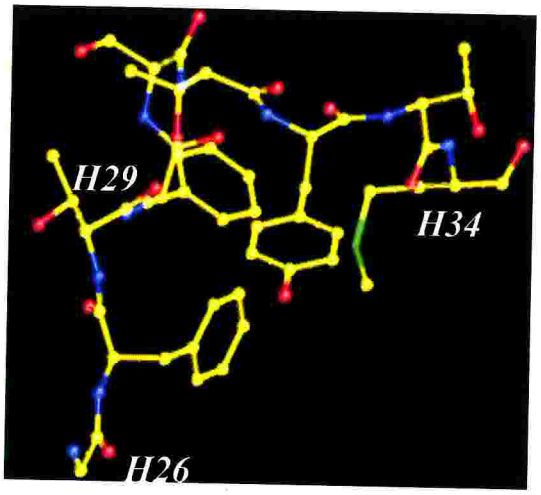


***Figure 6**: Canonical conformation of CDRH1*

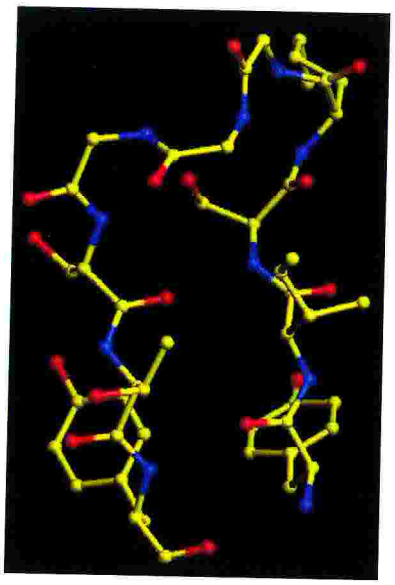


Figure 7:Uncanonical conformation of CDRH2

The CDRH2 (Figure 7) does not belong to any of the known canonical groups.

The third hypervariable region of the heavy chain CDRH3 has not been included in the canonical-structure description of the conformational repertoire of the three hypervariable regions because of its far higher variability in length, sequence and structure.

In the case of FabMNAC13 (Figure 7) the CDRH3 is composed by 13 residues. The length of this CDRH3 is significantly longer than the observed mean length of 8.7 for the equivalent hypervariable loops in murine antibodies (Wu *et al.*, 1993).

Recent studies however (Oliva *et al.*, 1998; Morea *et al.*, 1998; Shirai *et al.*, 1996) have identified some regularities in the conformation of CDRH3, allowing for some group classification, although to a less extent than for the other five antigen-binding loops.

Indeed, defining the CDRH3 region as encompassing the residues between Cys H91 and Gly H104, it can be divided into

- ◊ a stem, that includes residues proximal to the framework (four residues from the N terminus and six residues from the C terminus),
- ◊ a head, that includes residues in between.

Even if the heads or apices of the loops have a very wide variety of conformations, the stem region appears to have a limited repertoire of conformations, as in the canonical structure model of other antigen-binding loops. Indeed among the CDRH3 structures with more than ten

residues, two major classes can be identified according to the presence of either

◊ a β-bulge at residue H101

◊ a regular hydrogen-bonding pattern of the β-sheet hairpin, with no bulge at residue H101.

A bulged versus non-bulged stem conformation is primarily dictated by the sequence, through the formation of a salt bridge (3.5Å) between the side-chains of an Arg at position H94 and an Asp at position H101, as it is shown for CDRH3 of FabMNAC13 by a red circle (Figure 8).

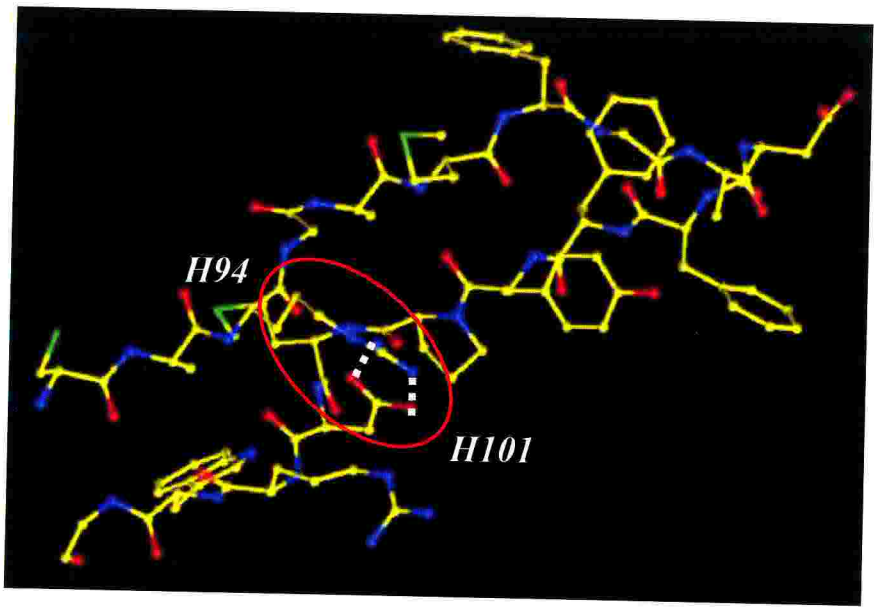


Figure 8: ***Conformation of CDRH3***

However CDRH3, unlike the other five antigen-binding loops, exibits a conformation that strongly depends on the environment. Indeed in particular those with bulged stems present many very different conformations of the head (Morea, 1998).

A molecular surface representation that covers the CDRs of FabMNAC13 is shown in Figure 9. The light and the heavy domains are colored in cyan and in yellow respectively. The CDRs1, CDRs2 and CDRs3 in both domains are white, green and blue respectively. The same "color code" has been used in the ribbon representation (see Figure 1).

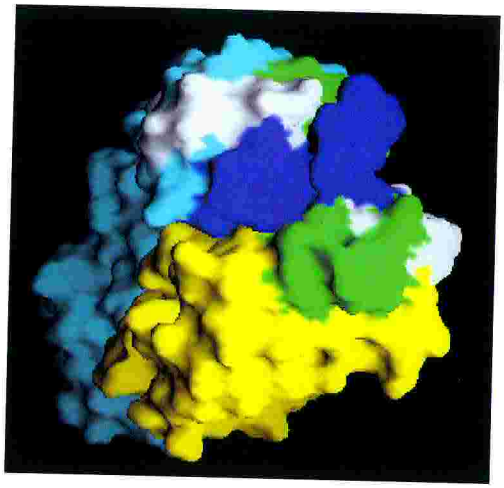


Figure 9: ***FabMNAC13 CDRs surface***

Strikingly the Fab MNAC13 CDRs surface resulted to be rather rough and not flat as observed for several other antibodies that have been raised against intact protein antigens. Instead antibodies raised against haptens or other smaller ligands usually display pronounced grooves or pockets on their CDR surfaces (Padlan, 1994).

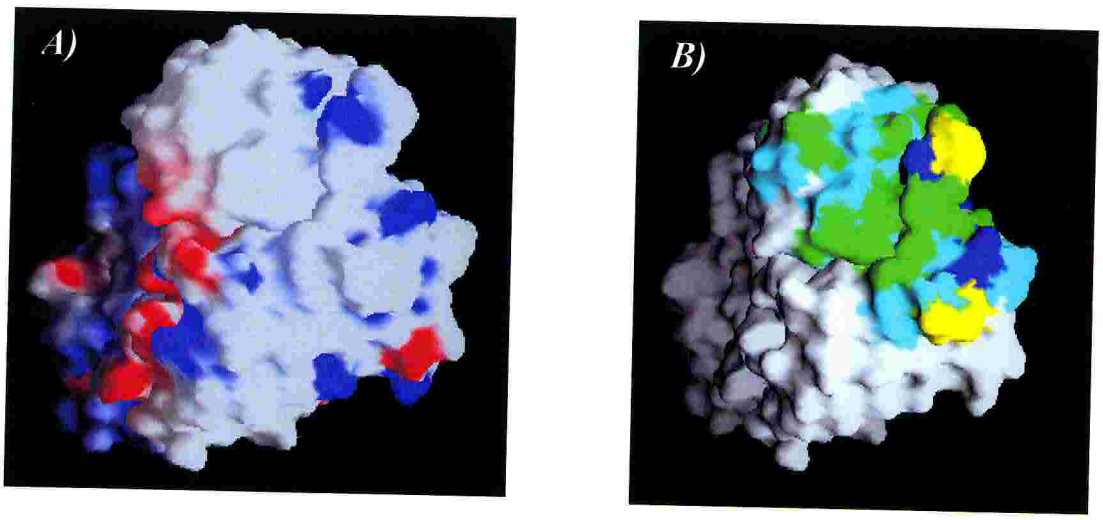


Figure 10: Physico-chemical properties of FabMNAC13 surface:

A) electrostatic potential;

B) hydrophobic and hydrophilic characteristics

Figure 10 shows the physico-chemical properties of the FabMNAC13 surface, calculated using the program GRASP. In particular Figure 10A shows the electrostatic potential of FabMNAC13 surface (with the acidic residues colored in red and the basic ones in blue). These calculations included only the protein formal charges using a protein dielectric constant of 2.0, a solvent dielectric constant of 80.0 and an ionic strength of 0.145M. Solvent accessible surfaces were calculated using the algorithm of Lee and Richards (1971) and a probe radius of 1.4Å.

Figure 10B shows the hydrophobic and hydrophilic patches of the antigen binding site surface of FabMNAC13 (with hydrophilic residues colored in cyan, hydrophobic in green, positively charged residues in blue and Gly in yellow).

Overall the surface encompassed by the CDR loops contains mostly neutral or uncharged residues (Figure 10B). However, two strongly charged patches of only a few residues each are localized in the heavy chain loops. The side chains of Lys H52A (CDRH2) and of Arg H94 (CDRH3) are partially directed away from the main antigen surface (Figure 10).

A close analysis of the combining site revealed the existence of an additional electron density in the 2Fo-Fc map. An isopropanol molecule used as additive (5% v/v) in the crystallization medium, could be readily fitted.

It is worth noting that in several reported structures ordered water molecules and alcohols have been found to replace the ligand in uncomplexed structures (Mattos and Ringe, 1996). The isopropanol molecule is hydrogen bonded via a water molecule to the hydroxyl group of Ser L92. This is the only hydrophilic spot in an otherwise hydrophobic crevice, in which the apolar side of the alcohol molecule is deeply

embedded making contacts with the aromatic rings of Tyr L32, Trp L91, Phe H99 and Phe H100D (Figure 11).

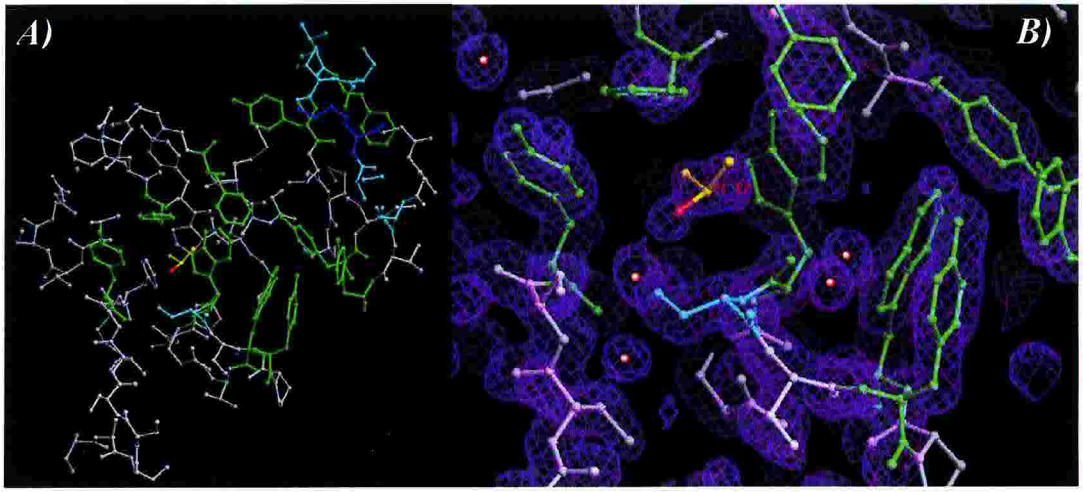


***Figure 11**: Isopropanol molecule inside the FabMNAC13 binding site: A) Top overview; B) closer view of the crevice together with the final 2Fo-Fc map contoured at 1.7 σ.*

Moreover deeply in the crevice and below the isopropanol molecule, there is a water molecule blocked by an hydrogen bonding network with the two carbonyl oxygen atoms of Tyr L32 and Ser L92, the amide nitrogen of Trp L91 and the NE2 of His L34 (Figure 12).

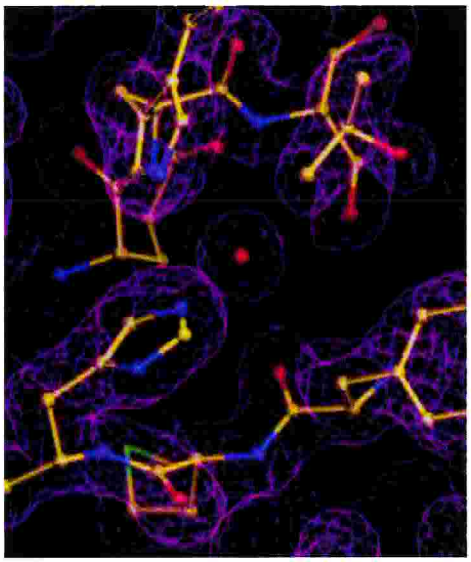


***Figure12:** Water molecule in the bottom of the crevice.*

Discussion

In the absence of a three-dimensional structure for the FabMNAC13 - TrkA extracellular domain complex, it is only possible to guess the

structural basis for the high affinity of the binding. The 7-8 kcal/mol required to increase the affinity from weak ($K_d \approx 10^{-4}$-10^{-5} M) to strong ($K_d \approx 10^{-10}$ M) binding could derive from the formation of salt bridges, of hydrogen bonds, especially when involving charged groups (Fersht *et al.*, 1985) or by hydrophobic interactions.

However, as mentioned before, structural studies of many antibody-ligand complexes have strongly suggested that the combining site of an antibody is primarily portrayed by the topology of CDRs surface. The FabMNAC13 CDRs structure shows some main interesting and unique features that may lead to hypothesize not only on its binding mode to the antigen but also indirectly on the TrkA-NGF interaction.

A first striking feature is the topology of its antigen binding site, that is characterized by significant and deep pockets at the level of the CDR surface. As mentioned above, this peculiarity makes FabMNAC13 closer to antibodies raised against small ligands than to antibodies raised against intact proteins.

Moreover it was quite interesting and unexpected to identify the presence of a bound small ligand. From this point of view it is worth noting that what has been by chance found in the case of FabMNAC13, has been exploited in the Multiple Crystal Structure Method (MCSM). This approach uses X-ray crystallography of a target protein in several organic solvents in order to find binding sites on the protein surface and to characterize potential ligand interactions within these sites (Mattos and Ringe, 1996). The rational behind is that the aqueous buffer, in which the protein crystals are kept during data collection, can be considered as a probe of the protein surface. Indeed in high-resolution protein crystal structures a number of ordered water molecules are clearly observed in sites, where the potential of mean force has a well-defined local minimum and that are highly complementary to the shape and chemical

composition of H_2O. However binding affinity on the surface is not uniform, as demonstrated by NMR experiments, due to the existence of rapidly exchanging water molecules with rates that vary considerably. Therefore MCSM extends this concept to use, instead of water, any liquid small organic molecule (if the protein crystal retains its integrity) as probe to locate other types of binding sites on proteins. Therefore repeatedly solving the crystal structure in several organic solvents (each chosen to mimic a particular functional group) it is possible to map out the specificity

The main and striking difference among this technique and the result obtained in FabMNAC13 structure concerns the concentration of the organic solvent. Indeed isopropanol was not used as a solvent nor as a precipitating agent to induce crystallization, but on the contrary as an additive (only 5% or even less since its volatility). Therefore its well defined location at the level of the FabMNAC13 antigen binding site is an indication of a quite high affinity for this organic molecule. Moreover the isopropanol molecule binds in a unique geometry and interacts within the pocket reflecting complementarity in shape and electrostatic properties. In particular even if the MCSM would alike consider isopropanol to mimic the functional group of threonine, the binding mode, observed in the FabMNAC13 structure, involves its hydrophobic part to be buried, while its hydrophilic region is exposed on the edge of the crevice. This finding may support the general idea that binding sites on the surface of proteins contain a small number of partially exposed hydrophobic residues complementary to an equally small portions of the ligand.

Figure 12 compares the FabMNAC13 binding site in presence and in absence of the isopropanol molecule: analyzing the antigen binding site depressions resulting in an exposed hydrophobic patch. In general similar

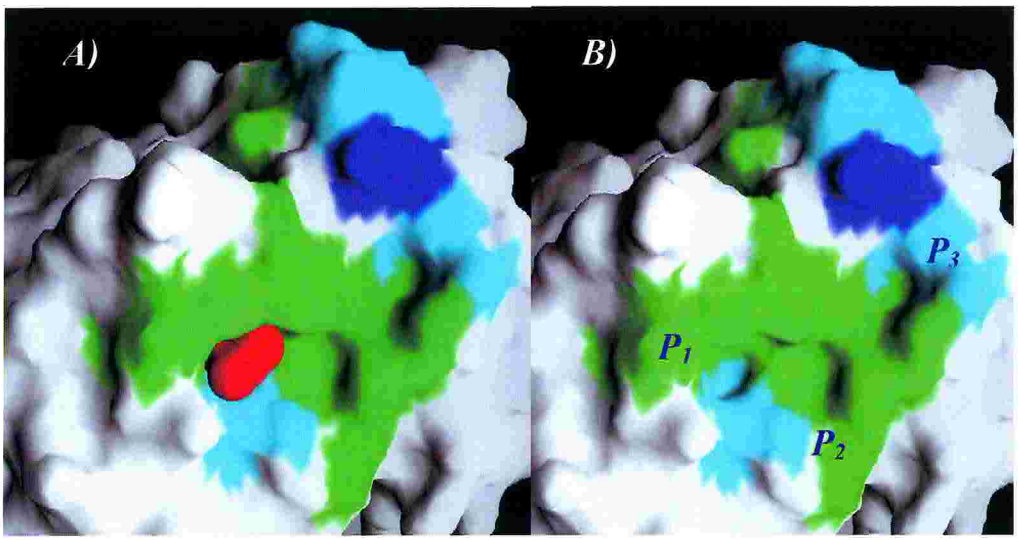


***Figure 12**: FabMNAC13 binding site surface A) with the isopropanol molecule inside; B) Empty crevice.*

small hydrophobic "hot spots" appear to be the principal contributors to binding affinity between ligands and their target proteins. Chothia *et al.* (1985) has shown that aromatic residues play a dominant role in antigen binding.

Hence the FabMNAC13 binding site may be portrayed by a channel, surrounded by hydrophobic side chains on each side, dipping deep in three hydrophobic pockets:

- ◊ the first one (***P1***) is the hydrophobic crevice (Tyr L32, Trp L91, Phe H99 and Phe H100D), in which the apolar side of the isopropanol molecule seats;
- ◊ an even larger hydrophobic pocket (***P2***) in the middle (Tyr L94, Trp L96, Tyr H50 and Tyr H58);
- ◊ a smaller apolar depression (***P3***) (Tyr H32, Phe H27 and Phe H29).

Moreover considering that the CDRH3 loop is quite extended and flexible (as underlined by dotted red line in Figure 13), it would be expected to

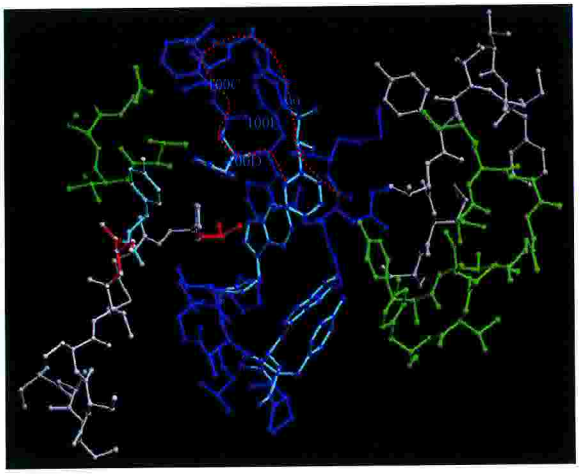


***Figure 13**: CDRH3 extended conformation in FabMNAC13 binding site*

affect the surface presented for antigen binding. Although the corresponding electron density for this loop is rather weak, not allowing for the side chain of Ans (H 100A) to be traced, this flexibility suggests a possible direct interaction with the antigen. Then upon binding this loop might assume a stable conformation.

Indeed if both the P_1 and P_2 hydrophobic clefts directly accomodate the epitope, CDRH3 might contribute to these interactions by the aromatic rings of Phe H100C and Tyr H100E significantly enlarging P_1 pocket.

Besides the main aromatic features of this crescent-shaped valley, it is worth noting that both sides are lined by polar residues. In addition at the right edge of the crevice a positively charged residue Lys H52A is present. The crucial localization of this basic residue suggests that its potential ionic interaction (salt bridge) may account, as part or as whole, for the high antigen affinity.

Moreover the structural features of FabMNAC13 binding site may be combined with biochemical data (epitope mapping), presented in the first part of this thesis and summarized in Figure 14.

AGWILTELEQSATVMKSGGLPSLGLTLA	*TrkA Del2*
GLPVNDEIPDPAK	*Phage epitope 1*
YYWLTLFGGDPAK	*Phage epitope 2*

***Figure 14:** Result of FabMNAC13 epitope mapping.*

Although the consensus is not so evident, these results seem to confirm the structural portrait of FabMNAC13 binding site. Indeed both the two peptides, selected by phage display, and the epitope mapped on TrkA extracellular domain by serial deletions, are characterized by the presence of a stretch of apolar residues, followed by one or two acidic residues. Such a region is very likely to match fairly well to the hydrophobic patch described above.

To verify that this epitope is really likely to be exposed on the surface of TrkA d4, it was difficult to model it on the basis of TrkA d5 structure (Wiesmann *et al.*, 1999), due to very poor sequence homology and to a quite different length of these subtype I immunoglobulin like domains (Schneider *et al.*, 1991).

Therefore TrkA d4 folding has been evaluated using two servers for biomolecular modeling to crosscheck their results (Figure 15).

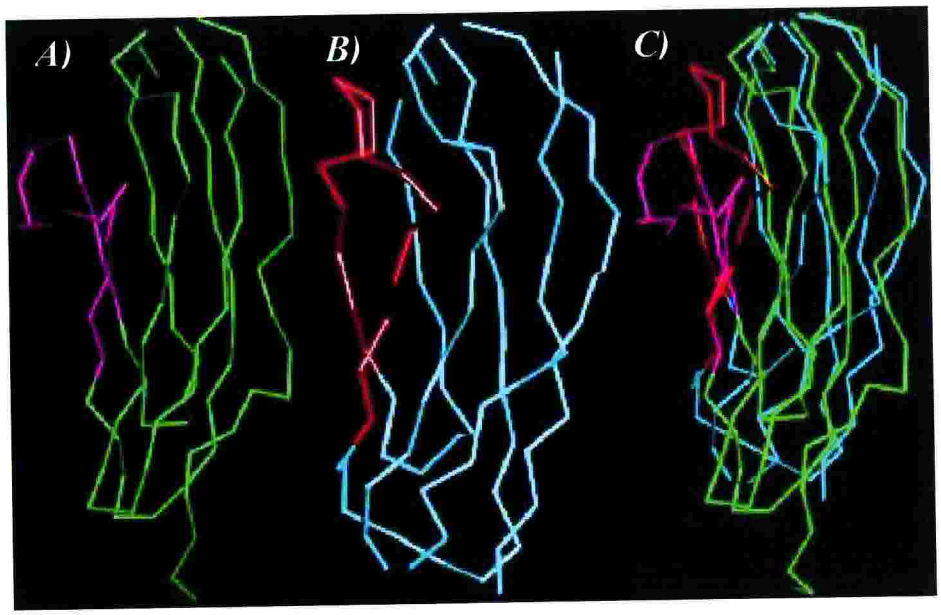


***Figure 15:* TrkA-d4 models:**

A) based on telokin template; B) based on anti estrogen Fab light chain template; C) superimposition of the two models

3D-PSSm (Kelley *et al.*, 2000), that performs protein fold recognition using 1D and 3D sequence profiles coupled with secondary structure and solvation potential information, used telokin (Holden *et al.*, 1992), the prototype of subclass I of the immunoglobulin folding superfamily (Figure 15A) as a template to model TrkA-d4. Instead the homology modelling server 3D-JIGSAW (Bates *et al.*, 2001; Bates & Sternberg, 2001) based its prediction on the structure of the light chain of an anti 17β estradiol Fab fragment (Lamminmaki & Kankare, 2001), which belongs to the of subclass V of the immunoglobulin folding superfamily (Figure 15B). The two models are consistent as shown by the superimposition of their Cα backbones in Figure 15C. Indeed the overall folding is quite similar, the main difference being located at the level of MNAC13 epitope (shown in red and in magenta in Figure 15A and 15B respectively).

However it is worth nothing that in both models this region is exposed on the surface of the domain and thus accessible for MNAC13 antibody recognition and binding.

In this respect Figure 16 shows a schematic ribbon diagram pattern, created by the program *Setor*, of the second model of Trka-d4 domain. The side chains of the MNAC13 epitope, depicted in magenta, result to be clearly solvent exposed.

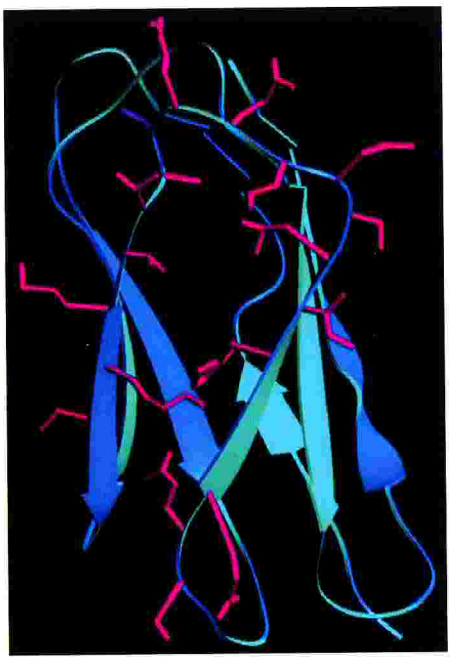


Figure 16: Ribbon representation of the model of TrkA-d4 domain.

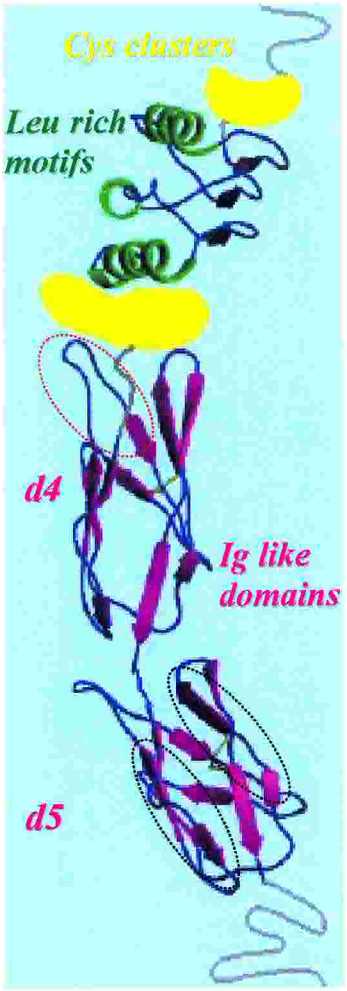


Figure 17: ***TrkA extracellular domain***

As a final consideration it is worth noting, that taken together, these data seems to suggest a more complex scenario for the NGF-TrkA interaction. Indeed these results underline that together with the second Ig-like domain of TrkA extracellular domain (d5, whose predicted secondary structure is shown in Figure 17), also the first one (d4) may play a crucial role in the observed high affinity of NGF for TrkA. Therefore beside NGF binding site on the second Ig-like domain (indicated by dotted black line), a cluster of hydrophobic residues near a negatively charged spot of the first Ig-like domain (indicated by dotted red line), may very likely be involved in the first steps of NGF docking, recognition and/or binding.

References

Bates, P. A., Kelley, L. A., MacCallum, R. M., Sternberg, M. J. (2001). Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*; **45** Suppl 5, 39-46.

Bates, P. A. and Sternberg, M. J. (2001). Model building by comparison at CASP3: Using expert knowledge and computer automation. *Proteins*; **45** Suppl 5, 47-54.

Chothia, C., Novotny, J., Btuccoleri, R. E. and Karplus, M. (1985). Domain association in immunoglobulin molecules: the packing of variable domains. *J. Mol. Biol;* **186**, 651-663.

Chothia, C., and Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol;* **196**, 901-917

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Wayne, M. M., and Winter, G. (1985). *Nature;* **314,** 235-238

Holden, H. M., Ito, M., Hartshorne, D. J., Rayment, I. (1992). X-ray structure determination of telokin, the C-terminal domain of myosin light chain kinase, at 2.8 A resolution. *J Mol Biol*; **227**, 840-51.

Lamminmaki, U., Kankare, J. A. (2001). Crystal Structure of a Recombinant Anti-Estradiol Fab Fragment in Complex with 17 Beta-Estradiol. *J.Biol.Chem.* **276,** 36687

Lee, B. and Richards, F. M. (1971). *J. Mol. Biol;* **55**, 379

Mattos, C. and Ringe, D. (1996). Locating and characterizing binding sites on proteins. *Nature Biotechnol.;* **14,** 595-599.

Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A. M. (1998). Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.* **275**, 269-294

Oliva, B., Bates, P. A., Querol, E., Aviles, F. X., and Sternberg, M. J. (1998). Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction. *J. Mol. Biol.;* **279**, 1193-1210

Padlan, E. A. (1994) Antibody-Antigen Complexes, R. G. Landes, Austin

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Mol. Immunol;* **31**, 169-217

Schneider, R. & Schweiger, M. (1991). A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogenic trk and trkB tyrosine kinase receptors. *Oncogene;* **6**, 1807-1811

Shirai H, Kidera A, Nakamura H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Lett.;* **399**, 1-8

Wiesmann, C., Ultsch, M. H., Bass, S. H. & de Vos, A. M. (1999). Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor . *Nature;* **401**, 184-188.

Wilson, I. A., and Stanfield, R. L. (1994). *Curr. Opin. Struct. Biol.* 4, 857-867

Wu, T. T., Johnson, G., Kabat, E. A. (1993). Length distribution of CDRH3 in antibodies. *Proteins*; **16**, 1-7.

CHAPTER 11: FabαD11 structure

Overall folding

The FabαD11 exhibits the typical immunoglobulin fold, as shown in Figure 1 with the same orientation and "color code" used in the previous chapter for FabMNAC13.

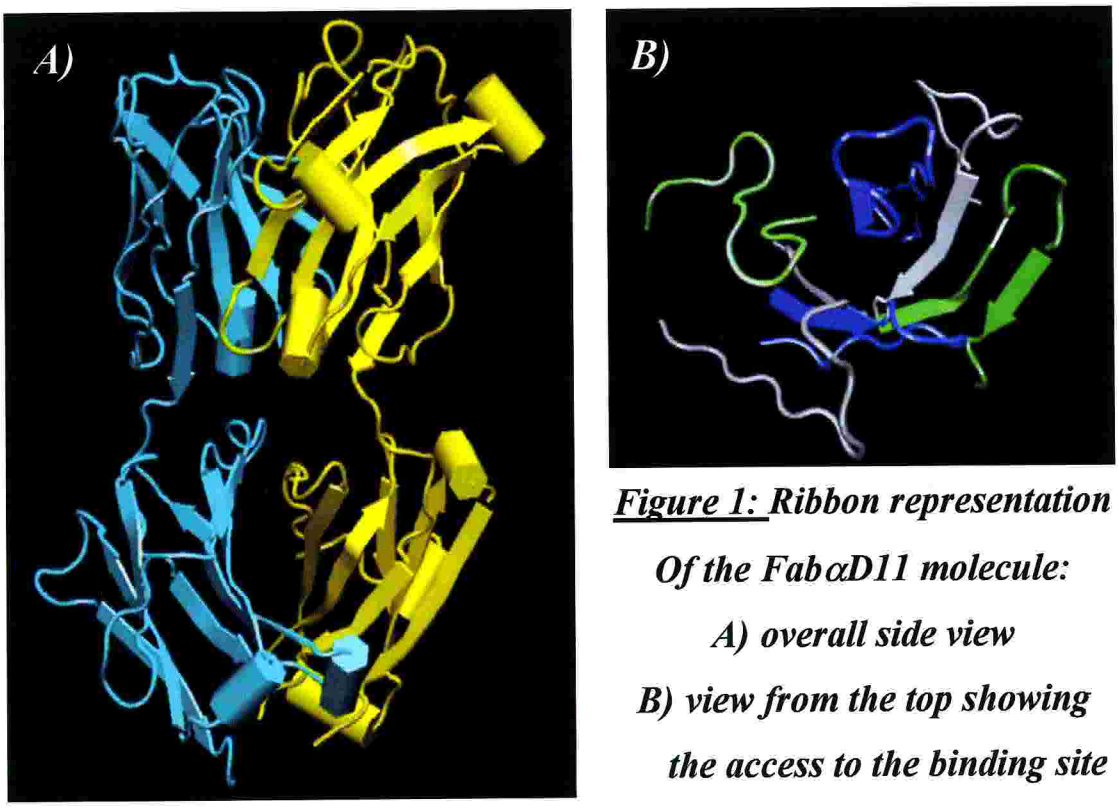


***Figure 1:** Ribbon representation Of the FabαD11 molecule: A) overall side view B) view from the top showing the access to the binding site*

The variable and constant domains (Figure 1A) of both chains contains two large β-pleated sheets, that pack face to face and are linked by the conserved disulphide bond, formed by Cys L214 and Cys H215. Each of the CDRs is a loop connecting two β-strands of the framework of the variable regions (Figure 1B). The elbow angle between the variable and constant domains is 165°, that is in the range observed for other Fab fragments (132°-172°), and there are no contacts between them.

Crystal packing

Since FabαD11 crystallized in two different space groups it is quite interesting to compare the two packing schemes. They are shown in Figure 2 and in Figure 3 respectively for the Crystal form II (triclinic space group) for the Crystal form III (monoclinic space group).

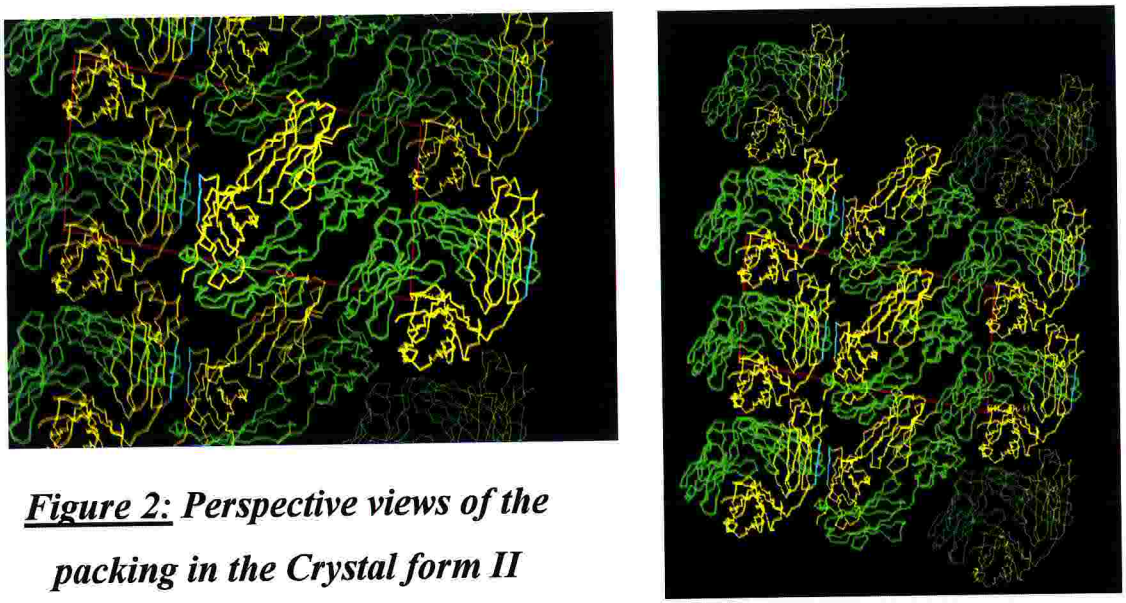


Figure 2: ***Perspective views of the packing in the Crystal form II***

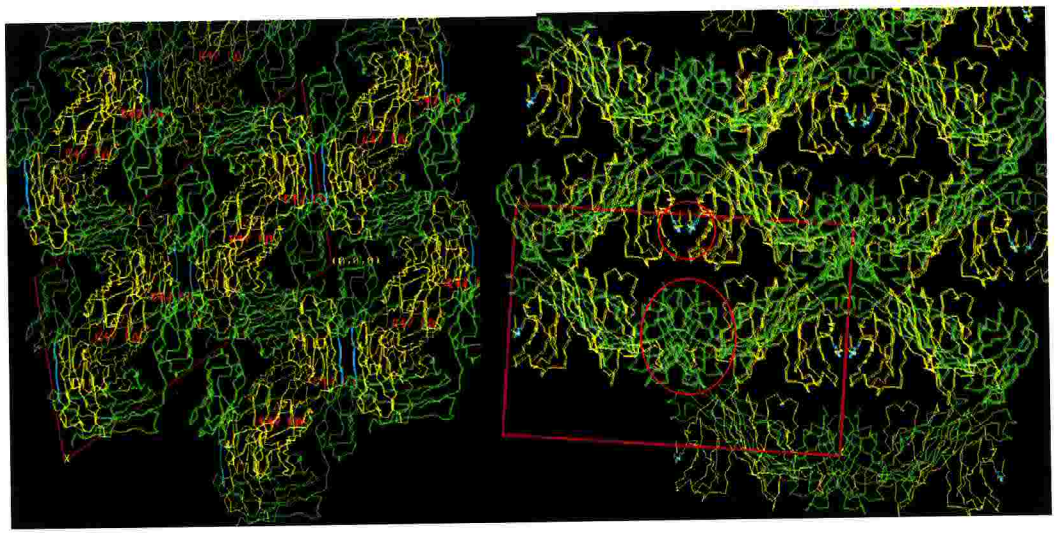


Figure 3: ***Perspective views of packing in the Crystal form III seen along b and c axis***

Both monoclinic and triclinic crystal forms of the FabαD11 display similar crystal packing features. This finding suggest that the non-

crystallographic quasi-perfect two-fold (177°), that relates the two molecules in the triclinic cell, corresponds very likely to the crystallographic two-fold axis in the monoclinic C2 space group.

It is worth noting that this mode of packing into both unit cells is rather unusual and complex. It has been previously observed (Cygler *et al.*, 1987) that Fab fragments often pack in a head-to-tail fashion (defining the variable domain as head and the constant domain as tail), with the variable domains packing against the constant domains of symmetry-related Fab fragments.

In the monoclinic as well as in the triclinic space groups, FabαD11 main crystal contacts involve instead tail-to-tail interactions through their C_H domains, (Figure 4). Namelly a short intramolecular antiparallel β-sheet is observed between the β-strand stretch from Ser H203 to Lys H209 of each molecule (Figure 4B).

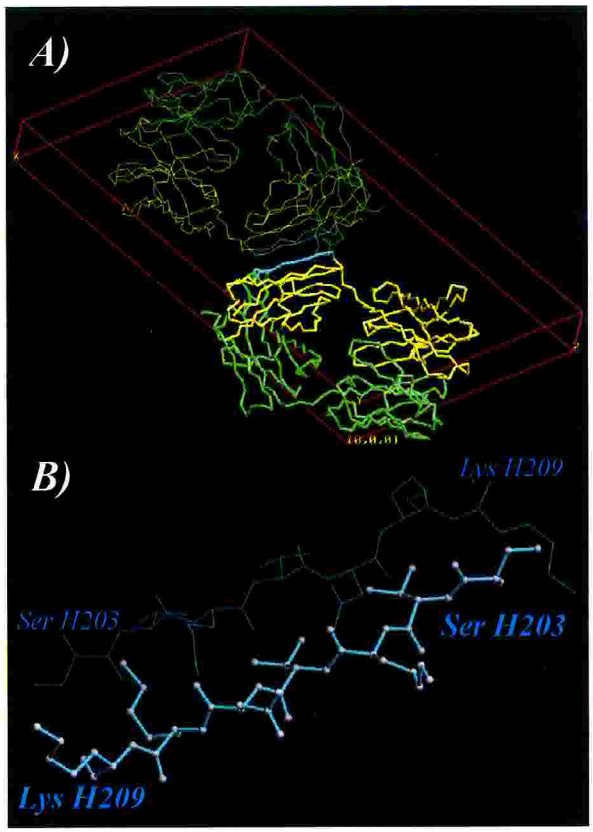


Figure 4: Main packing contacts: A) in the unit cell; B) closer view.

Further additional contacts, involve different domains (Figure 5B) of three symmetry related molecules. It is interesting to underline that the antigen binding sites of two of the three molecules are partially involved these secondary interactions, as shown in Figure 5A. Namely head-to-head contacts occur among the apices of all three light hypervariable loops. Two symmetry-related molecules show an head-to-head antiparallel hydrophobic

interaction via the aromatic rings of Tyr L30 in CDRL1, Tyr L91, Phe L92 and Tyr L94 in CDRL3 and hydrogen bonds involving Tyr L30 in CDRL1 and Asp L52 in CDRL2 (as shown in Figure 5A).

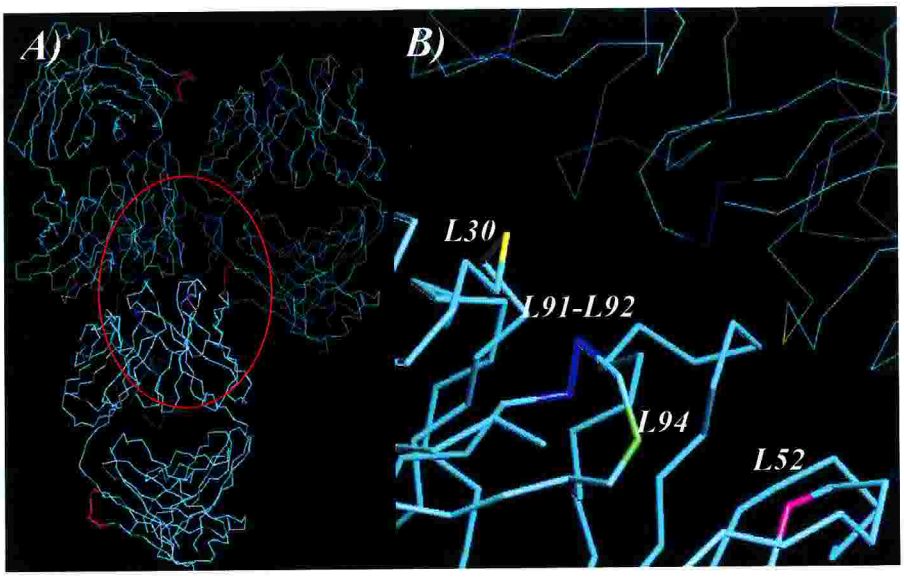


***Figure 5:** Additional packing contacts: A) overall view; B) closer view of the interaction involving binding sites of symmetry-related molecules.*

Moreover the region between residues Ser L65 and Ser L72 in one of them is involved into a loose head-to-tail van der Waals interactions with the stretch from Lys L198 and Ser L203 at the C terminus of the light chain of a third symmetry-related molecule. Figure 6 summarizes all the regions of FabαD11 that are involved both in main and secondary crystal packing contacts (the color code is the same for the Figures 4 and 5). It is clear that the main and secondary contacts involve opposite sides of the molecule.

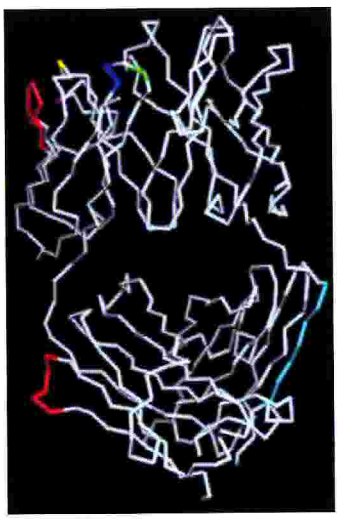


***Figure 6:** Overview of FabαD11 crystal contacts*

Reconsidering on the light of these data the disapponiting results obtained in the crystallization studies for the recombinant form of this antibody (Chapter 4), it is quite evident that linkers spanning from 15 to 20 aminoacids would be of a suitable lenght for a proper orientation of the two corresponding variable domains in the single chain Fv fragment.

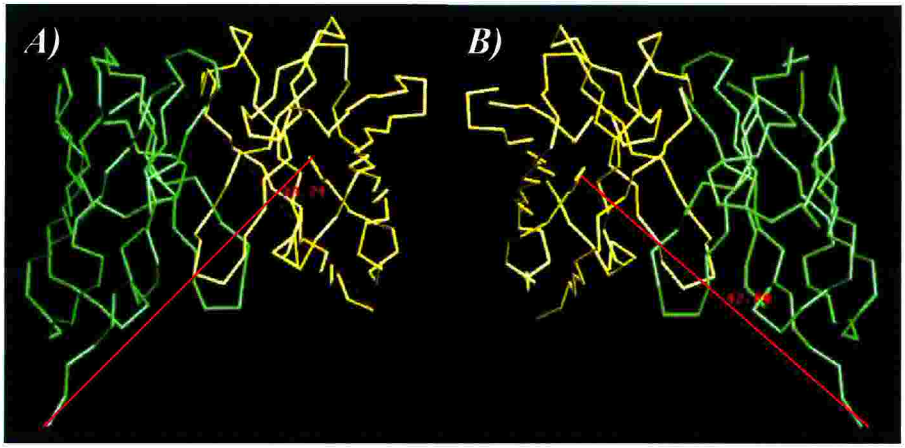


Figure 7: Fv fragment in FabαD11 crystal structure: distance between A) the N terminus of V_L and the C terminus of V_H B) the N terminus of V_H and the C terminus of V_L

Indeed in Figure 7 the distances between the N terminus of V_L and the C terminus of V_H (33.7Å as shown in Figure 7A) and between the N terminus of V_H and the C terminus of V_L (42.7Å as shown Figure 7B) are depicted, since both arrangements have been employed in the two constructs. Supposing that the distance between two neighbouring C_α atoms is around 3.8Å, the shorter of the two linkers should account for a 57Å spacing while the longer one for a 76Å spacing: both of them would allow the interaction between the two domains observed in this crystal structure.

Moreover supposing that the scFv might crystallize with the same packing mode described above, it is interesting to note either the N and C

termini of the variable domains in the two possible arrangements are buried and therefore not accessible to allow a linker to connect them in an intramolecular fashion (Figure 9).

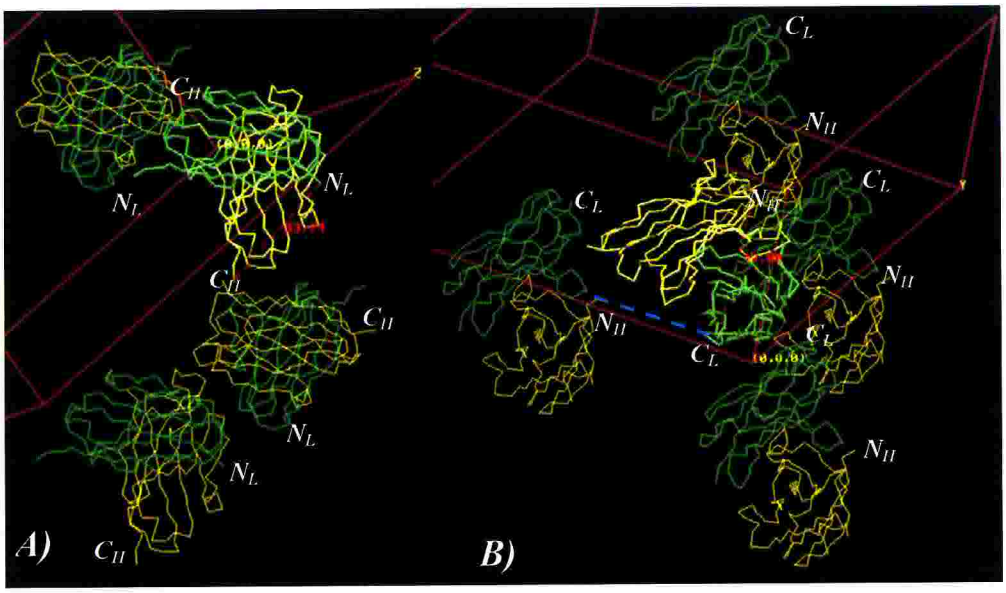


Figure 9: Fv fragment in Fab αD11 crystal packing: relative 3D arrangements between A) the N_L (N terminus of V_L) and the C_H (C terminus of V_H); B) the N_H (N terminus of V_H) and the C_L (C terminus of V_L)

Therefore αD11 scFv fragments could not interact in a similar way, if they would have crystallized as monomers because the steric obstruction of both amino and carboxi termini. On the other hand the alternative possibility is that they would have crystallized as dimers (and in that case the linker should have been intermolecular). Again it seems difficult to build a linker connecting N_L to C_H (Figure 9A) in an intermolecular way due to the steric hindrance or because they result to be very far apart from each other. This seems to be a direct consequence of the head to head packing in the present crystal form III. In contrast in the N_H to C_L configuration an intermolecular linker seems to be easily built in terms of distance and accessibility (dotted blu line in Figure 9B).

FabαD11 Antigen Binding site

Analysis of the main-chain conformation of the CDRL1, CDRL2, CDRL3, CDRH1, and CDRH2 hypervariable regions by sequence and three-dimensional structure comparison with other immunoglobulins confirms that standard conformations based on hypervariable loop length and sequence (canonical structure) are maintained.

In particular according to the definitions of Chothia and Lesk (1987), five out of the six CDRs of FabαD11 result to be canonical:

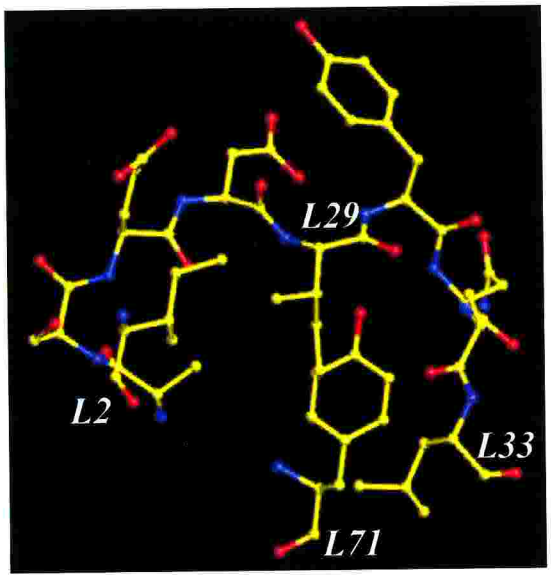


Figure 9:
Canonical conformation of CDRL1

◊ CDRL1 (Figure 9) belongs to the canonical group 2. It extends across the top of the β-sheet framework with the side chain of Ile L29 deeply buried in a cavity between residues Ile L2, Leu L33 and Tyr L71.

◊ CDRL2 and CDRL3 loops are both of class 1 (Figure 10 A and B respectively). In particular CDRL2 links two adjacent strands in the frame work β-sheet. This region is a three residues hairpin, since residues Tyr L49 and Thr L53 are hydrogen bonded to each other (2.8Å).

CDRL3 links two adjacent strands in the frame work β-sheet. The

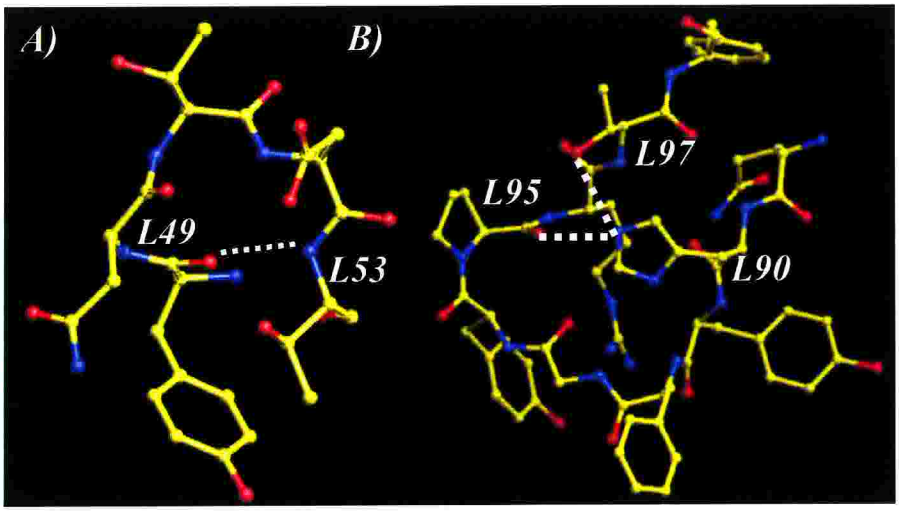


***Figure 10*: Canonical conformation of CDRL2 (A), CDRL3 (B)**

presence of Pro L95 with a cis peptide and the α_L conformation of residue Phe L92 put residues from L93 to L96 in an extended conformation. Important determinants of this conformation is the hydrogen bond formed by the side chain of framework residue His L90 (Nε2) to the main chain carbonyl of L95 (3.34Å). The Nε2 of His L90 also interacts with OH group of Thr L97 (3.14Å). The carbonyl of L93 instead is not involved in any hydrogen bonding as expected for this canonical conformation.

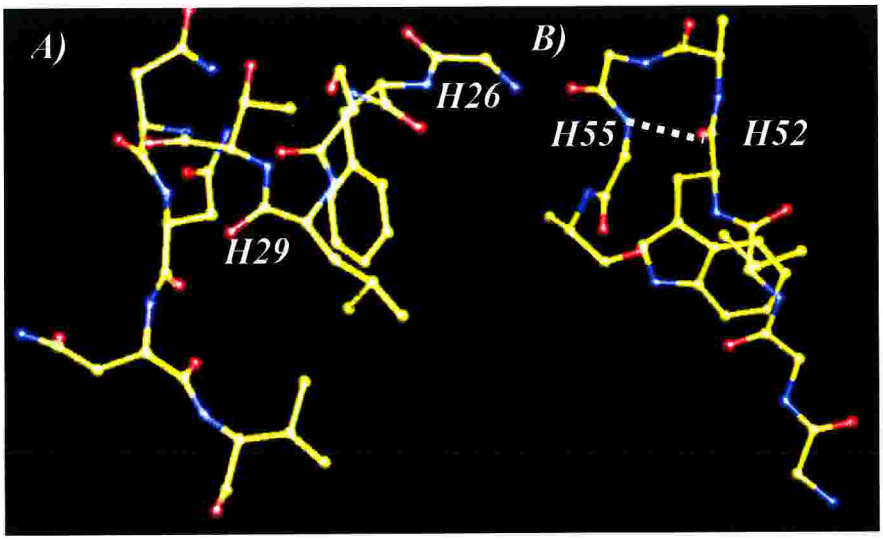


***Figure 11*: Canonical conformation of CDRH1 (A) and CDRH2 (B)**

◊ CDRH1 (Figure 11A) and CDRH2 (Figure 11B) are judged to belong to the canonical class 1.

CDRH1 region, packing across the top of the V domain, is characterized by a sharp turn due to Gly H26 with Leu H29 deeply buried within the framework structure.

CDRH2 region, linking two adjacent strands in the framework β-sheet, is a seven residues turn, with a three residues apex, formed by a hydrogen bond between residues H52 and H55 (2.73Å), which is a conserved Gly. The other four residues in the turn are part of the framework structure.

As discussed in the previous chapter, due to the high diversity of the apex of CDRH3, a comprehensive canonical classification scheme has only been introduced for the stem of the loop.

Also in FabαD11, the stem of CDRH3 adopts a kinked conformation due to the presence at the two sides of the loop of Arg H94 and Asp H101, which form a conserved salt bridge (2.68Å) in the kinked-type loop (Figure 12).

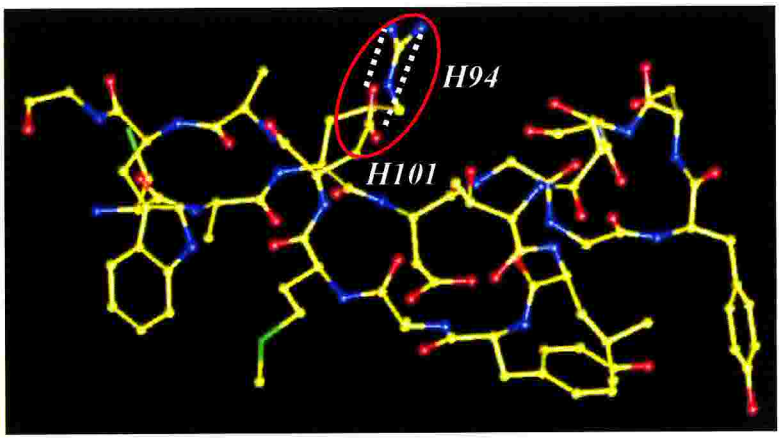


Figure 12: Conformation of CDRH3

Thus, molecular recognition is accomplished by the structural motifs used by all immunoglobulins to recognize antigen or hapten-protein conjugate.

The molecular surface that covers the CDRs of FabαD11 is shown in Figure 13 (again with the same color code that had been used above and in the previous chapter). It is interesting to note that also in this second structure the CDRs surface display pronounced grooves and pockets,

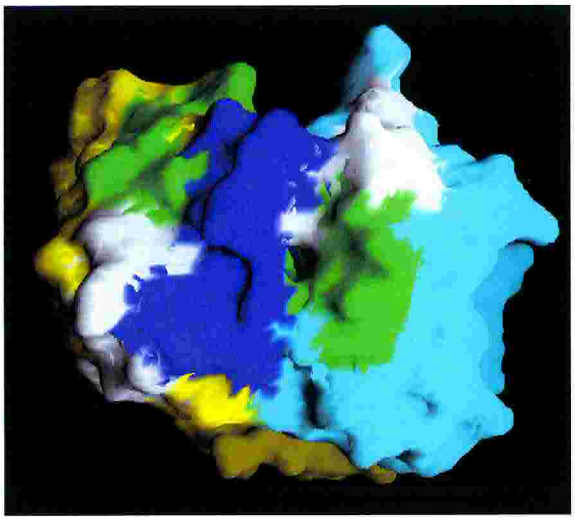


***Figure 13**: FabαD11 CDRs*

in contrast with the relatively flat CDRs surfaces typical of several other antibodies raised against intact protein antigens.

The physical-chemical properties of the FabαD11 surface are analyzed in Figure 14, showing the molecule in the same orientation of Figure 13.

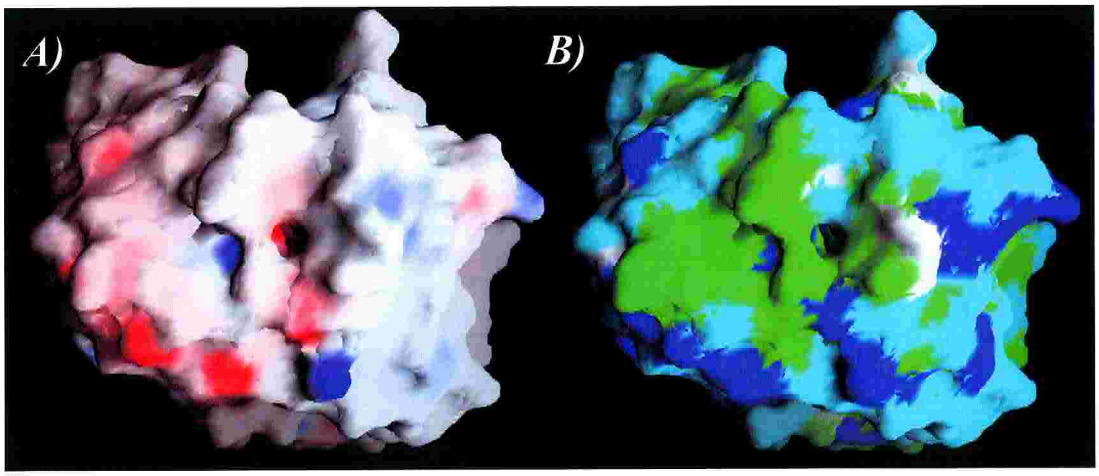


***Figure 14**: Physico-chemical properties of FabαD11 surface:*

A) elettrostatic potential;

B) hydrophobic and hydrophilic characteristics

In particular Figure 14A shows the electrostatic surface potentials of FabαD11 (with acidic residues in red while the basic ones are in blue),

calculated using the program GRASP. The hydrophobic and hydrophilic feature of FabαD11 surface (with hydrophilic residues in cyan, hydrophobic in green, positively charged residues in blue and Gly in yellow) are represented in Figure 14B.

Overall the surface of the antigen binding site is mostly composed by neutral or uncharged residues, with the exception of three strongly charged zones. Indeed there are two acidic patches, one is composed by residues Glu L27 and Asp L28 in the CDRL1 and the other formed by residues Asp H94 and Asp H100F in the CDRH3. In both cases the negatively charged surface seems to be exposed on the surface of the antigen binding area (Figure 14). Moreover deeply buried at the bottom of two otherwise hydrophobic clefts there is the basic residue Arg L96 localized in the light chain loop CDRL1.

Discussion

As mentioned in the previous chapter, also here in the absence of a crystal structure for FabαD11 NGF complex, it is only possible to speculate on the structural basis for the high affinity of the binding.

A common feature between these two Fab structures, although in contrast to what generally reported regarding antibodies raised against intact proteins, consists in the overall topology of the antigen binding site. Indeed also in the structure of FabαD11 the analogous region is slightly convex with a shallow groove or channel running the width of the molecule resulting in a bipartite binding surface, as shown in Figure 15.

In particular both of these two clefts are surrounded by apolar side chains nearly on each side:

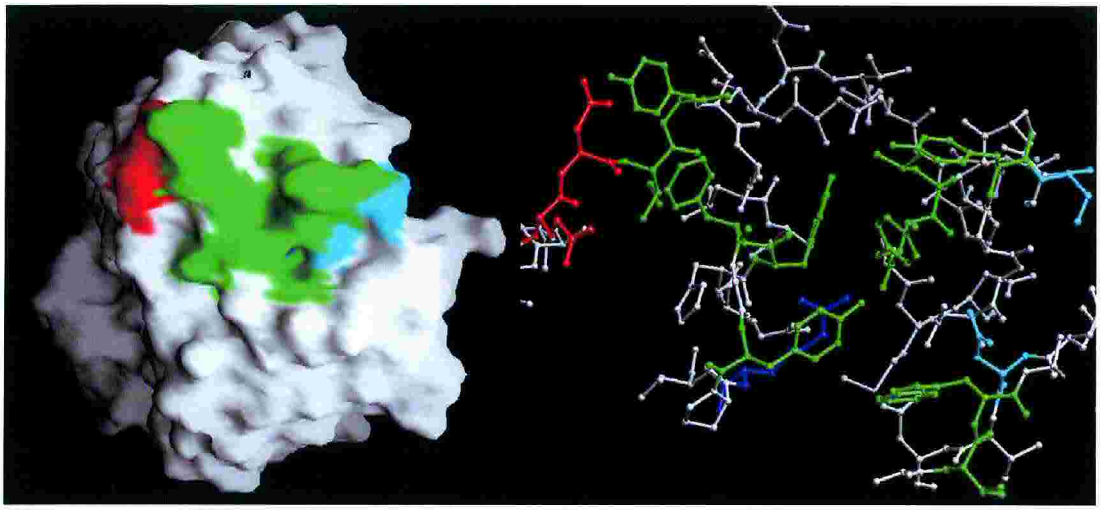


***Figure 15*: Fab αD11 binding site**

- ◊ a first mainly hydrophobic pocket is lined by Ile L29, Tyr L30, Tyr L91, Phe L92 and Tyr L94;
- ◊ on one edge of the second cleft, otherwise surrounded by the aromatic side chains of Tyr H51, Trp H52, Tyr H100B and Tyr H100C, there is an hydrophilic spot due to the presence of Asn H32 and Asn H33.

The bottom of these two pockets is shared and mainly hydrophobic, except for a central polar bulge containing Arg L96. Moreover near one side of this hydrophobic "horseshoe" shaped crevice, there is a negatively charged patch (Glu L27 and Asp L28), that protrudes from the binding surface.

Some other features may contribute to the high affinity binding, such as flexibility in order to improve the binding of FabαD11 to NGF. Although Asn in CDRs have been frequently found to form hydrogen bonds with main chain atoms, thus stabilizing the conformation of the local structure (Padlan, 1990), in FabαD11 the central two of the stretch of four Asn (from H31 and H35) in CDRH1 are solvent exposed. Thus the fact, that they are not engaged in hydrogen bonds, should result in a reduced stability and higher flexibility of CDRH1 loop, that constitute one edge of the bipartite hydrophobic binding surface (Figure 15). Indeed if both

hydrophobic clefts directly accommodate the epitope, CDRH1 might contribute significantly enlarging the smaller second pocket due to the presence of the two Asn.

Other two regions, that are almost certainly flexible, are the CDRH2 and CDRH3 loops due to the presence of Gly residues: Gly H50, H54 and Gly H55 in CDRH2 and Gly H95 and H96 in CDRH3 respectively.

Another interesting feature regarding Fab αD11 binding site is related to its packing described above. Indeed among the four symmetry-related molecules, that closely interact, two in particular are engaged into a head-to-head contact, leading the respective binding sites in close vicinity (Figure 16).

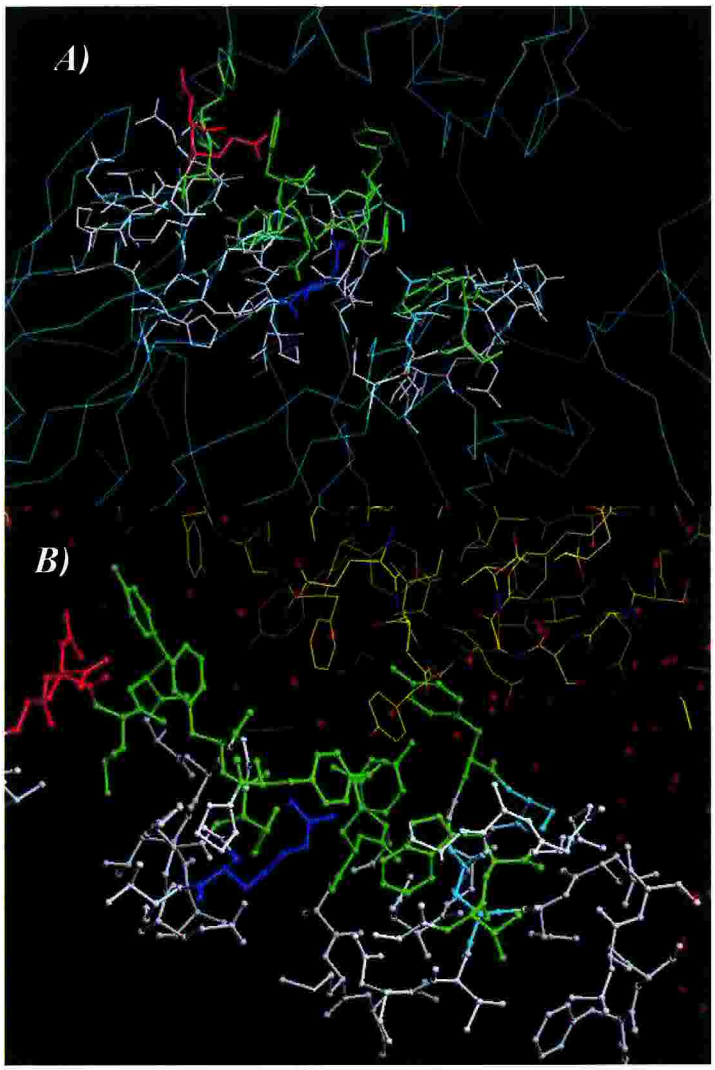


Figure 16: ***Interactions at the level of Fab αD11 binding sites between symmetry-related molecules: A) overall view of the three symmetry-related molecules; B) closer view of two interacting binding sites.***

In particular the aromatic residues of CDRL1 (Tyr L30) and CDRL3 (Tyr L91, Phe L92 and Tyr L94) seem to mediate this contact by hydrophobic interaction in an antiparallel fashion against corresponding aromatic residues of a symmetry-related molecule.

The stacking of the aromatic rings of CDRL1 and CDRL3 belonging to different symmetry-related molecules is shown in Figure 17, together with the final 2Fo-Fc map. Ii is worth noting the presence of several ordered water molecules within the binding site.

The involvement of aromatic residues to the crystal packing may suggest its importance in NGF recognition. Indeed the overall bipartite depression on the surface, characterized by a greater than average degree of exposure

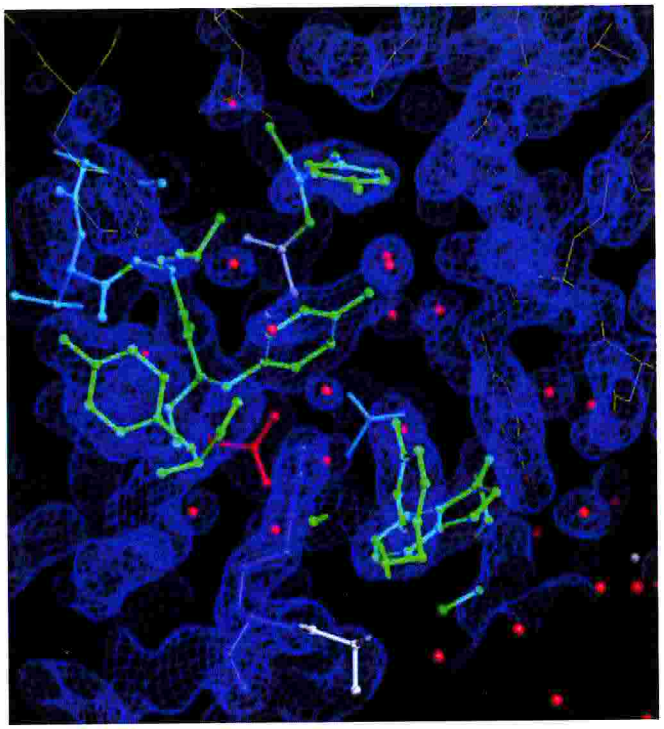


Figure 17: Fab αD11 binding sites together with the final 2Fo-Fc map contoured at 1.7 σ.

of hydrophobic groups, can be considered as hydrophobic "hot spots", that are thought to be the principal contributors to the interaction between ligands and their target protein. In addition these sites can be occupied by

ordered water molecules that are displaced with a gain in entropy upon complex formation, allowing for more favorable electrostatic interactions between the protein and the ligand.

Such an overall portrait of FabαD11 binding site appears to be consistent with previous epitope mapping data. In particular these studies identified Lys 95 and the L2 loop (whose aminoacidic sequence is shown in Figure 18) as a necessary and sufficient condition for NGF binding. Indeed it seems to be rather likely that the bipartite aromatic cleft may interact with one or more apolar residues in the L2 loop, that is localized in an highly hydrophobic region of NGF surface (Figure 18). Moreover the specificity and strength of interaction could be ascribable to the formation of salt bridges, involving buried charged groups such as Arg L96 on CDRL3 and Asp 41 on L2 of NGF. Beside this, the importance of Lys 95 on L5 loop of NGF can be clearly explained in this scenario. Indeed if it forms a salt bridge with one of the negatively charged residues Glu L27 and Asp L28, this positively charged residue may be responsible, in part or in whole, for the high affinity.

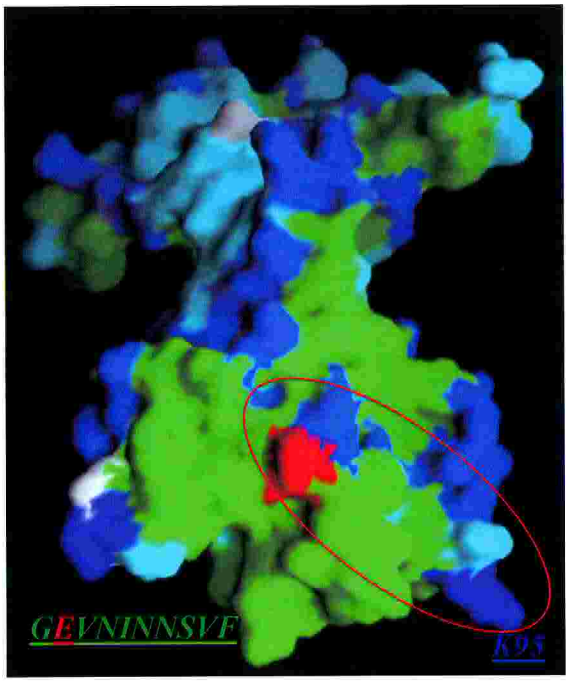


Figure 18: FabαD11 epitope on NGF surface

References

Padlan, E. A. (1990) *Proteins Struct. Funct. Genet.* **7,** 112-124

CHAPTER 12: Conclusions and perspectives

As mentioned before the progress in the area of monoclonal antibodies, following the discovery by Kohler and Milstein (1975), showed their wide range of biological activities and specificities, and indicated a tremendous potential in the treatment and diagnosis of human diseases (Bach *et al.*, 1993).

As a consequence the detailed structural information on the antibody binding site structure paves the basis for the study of complex protein-protein interactions. In addition these structural information are a pre-requisite for the rational design of diagnostic reagents and drugs, based on the activities of the antibodies (in particular as summarized in Figure 1 antibody's humanization, development of CDRs peptides or peptide mimetics).

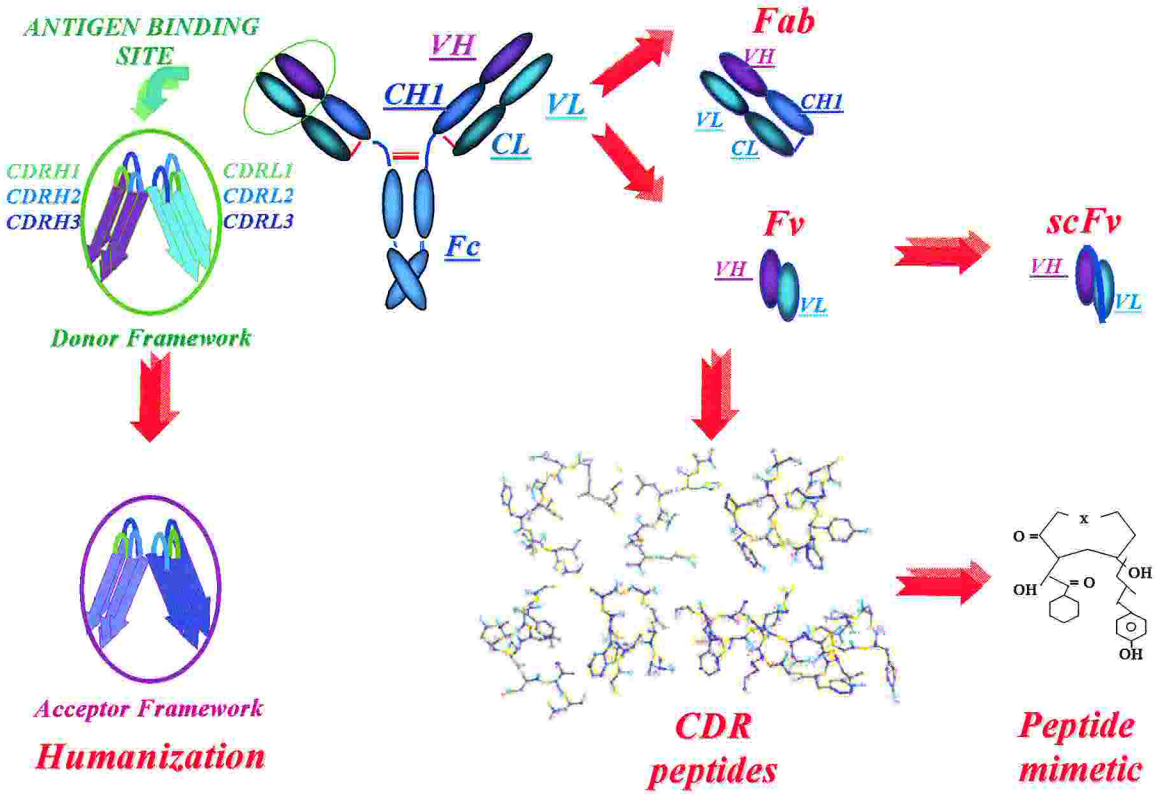


***Figure 1*: Antibody derived molecules**

Insight on TrkA-NGF interaction

The structural characterization of an antibody binding site is essential in molecular recognition studes and to explore structure-function relashionships of receptor and ligands. From this point of view they allow to reduce complex macromolecular interactions to better defined models. In particular structural analysis of antibody-anti-idiotype pairs has been extensively employed as a model to study complex protein-protein interactions (Williams *et al.*, 1990). The concept of ligand mimicry has been also extended to blocking anti-receptor antibody. Indeed the possibility that structures within the CDRs of these antibodies may display similarity with the original receptor ligand has been documented (in the reovirus system, in the fibrinogen-receptor system and in the TSH-receptor system), thereby validating this immunologic approach to explore ligand-receptor interactions (Taub and Greene, 1992). Moreover the same immunologic approach may be applied for anti-ligand neutralizing antibodies, that would mimic the original receptor binding site at the level of primary or secondary structure and thus form an internal image of the receptor binding site.

Therefore combining biochemical and structural data presented in this thesis several considerations may give interesting insights on NGF-TrkA mode of recognition and binding.

It is worth of noting that FabαD11 binding site on NGF does not overlap with the two NGF patches, that have been previously shown to interact with TrkA-d5 (black circles in Figure 2).

Therefore the binding sites of these two blocking antibodies might portray the two complementary surfaces, on the ligand and the receptor, involved in preliminary events of NGF binding to TrkA.

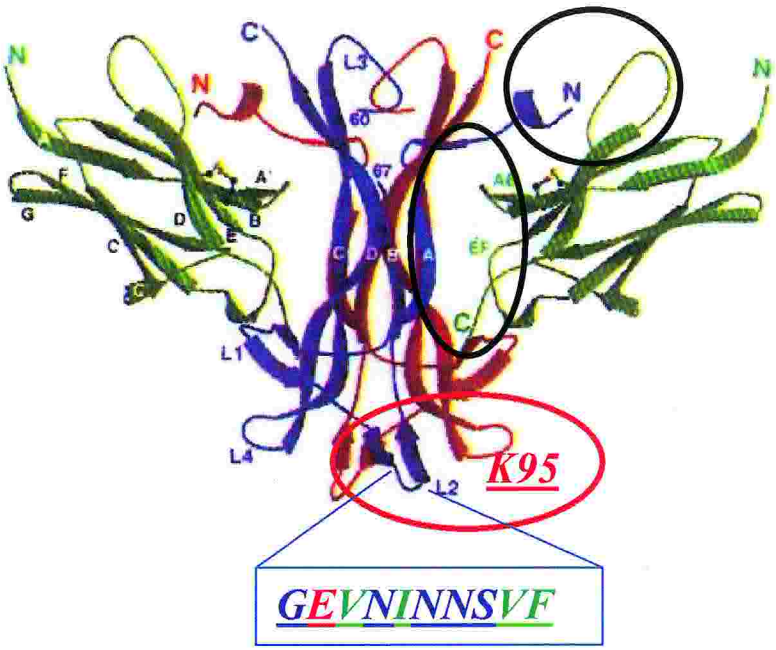


Figure 2: FabαD11 binding site respect to the complex NGF-TrkAd5

This speculation may be supported by similarities of physico-chemical properties of respective surfaces. Indeed it is possible to put in evidence some common features of the FabMNAC13 binding site and the NGF epitope respectively. Comparing these two surfaces (Figure 3), they share the presence of an exposed Lys (colored in blu) near a hydrophobic cleft.

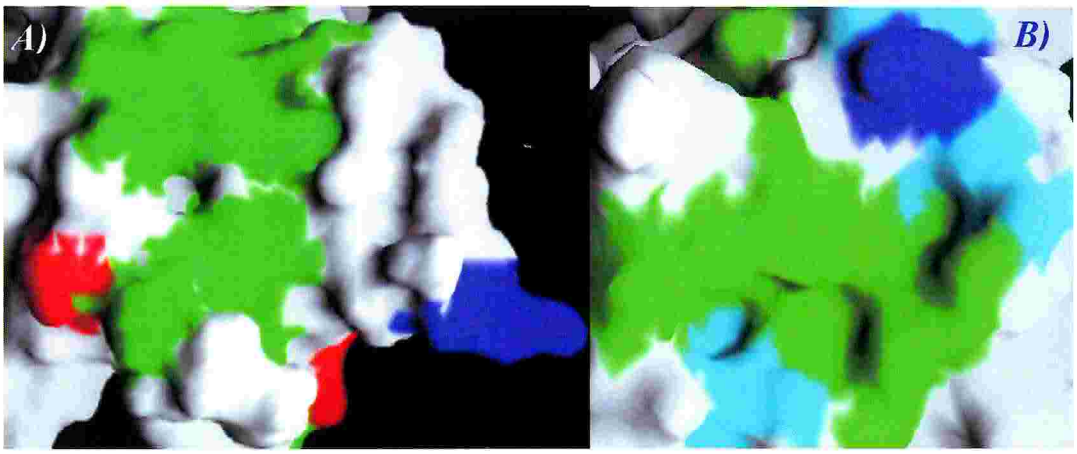


Figure 3: A) FabαD11 epitope
on NGF surface; B)FabMNAC13 binding site

This cleft can be subdivided in smaller pockets in the case of the antibody, while it is flatter and wider on NGF surface.

On the other hand a comparison between FabαD11 binding site and TrkA epitope is not so trivial, since the three dimensional structure of TrkA-d4 has not been determined. However one may possibly speculate on the coincidence of some electrostatic features of the FabαD11 site and the sequence, recognized by the FabMNAC13 on TrkA-d4:

AGWILTELEQSATVMKS

Indeed the FabαD11 binding surface shares both charged patches, present in this sequence (with two acidic residues one next to each other), together with hydrophobic residues that are clustered in this region of TrkA.

Also in this second case such a speculation seems to be confirmed comparing the physico-chemical properties of respective surfaces, as shown in Figure 4.

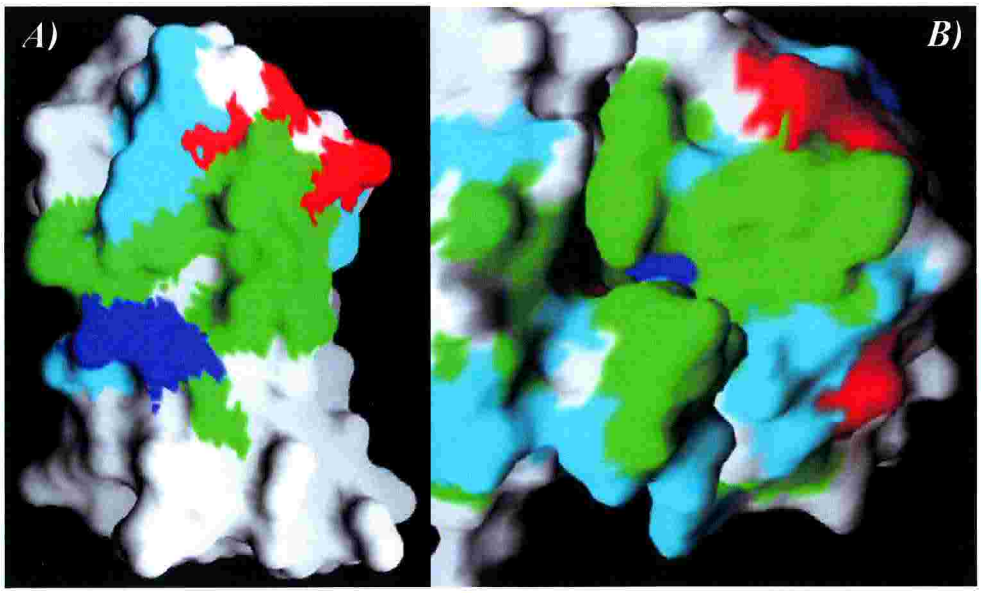


***Figure 4*: A) FabMNAC13 epitope on NGF surface; B) FabαD11 binding site**

Starting from modeled the TrkA d4 structure (described in Chapter 10), some common features of the FabαD11 binding site and the MNAC13 epitope may be underlined. Indeed they share the presence of a positive charge (colored in blu) and a negatively charged spot charge (colored in red) with a hydrophobic cleft in between.

Antibody's humanization

For medical, diagnostic and therapeutic applications human antibodies are preferred, but it is usually difficult to obtain them. The production of rat or mouse antibodies is usually straightforward and thus antibodies against specific targets are often obtained in animals. However the major drawback to the use of animal antibodies in man is their potential antigenicity. Several studies have shown that "humanization" of non-human antibodies provides a way to minimize this non-specific antibody response and fast clearance rates, while retaining the highly specific and tight binding characteristics of the antigen-antibody interaction (Winter & Milstein, 1991; Winter & Harris, 1993). Typically humanization involves grafting the intact antigen recognizing CDRs from a donor animal antibody to an acceptor human framework. However it has been demonstrated that this step alone is not always sufficient, because other key residues outside the hypervariable regions may play an important role in antigen binding, whether through direct antigen interaction (Amit *et al.*, 1986; Padlan, 1989) or by providing structural support for a CDR (Chothia *et al.*, 1989, Novotny *et al.*, 1990, Tramontano *et al.*, 1990; Tempest *et al.*, 1991; Foote & Winter, 1992; Presta *et al.*, 1993; Schildbach *et al.*, 1993). Therefore humanization by CDR grafting is not a straightforward procedure.

Considerable progress in the successful humanization of an antibody can be made by prior knowledge of its 3D structure, based on present

understanding of other antibody structures. Success is even more likely if the precise structure of the target antigen binding site is known. Therefore the determination of the crystal structure of both neutralizing antibodies is expected to aid in our efforts to derive a humanized version of each of them. In particular the structure should allow a more accurate interpretation of the mutagenesis experiments designed to graft the high affinity antigen binding site into an acceptor human antibody framework. In addition, on one hand the knowledge of the NGF 3D structure and on the other epitope mapping experiments may allow the prediction of which residues in the CDR loops respectively of FabαD11 and FabMNAC13 are important for antigen recognition and binding. Once selected appropriate human Fv templates for both neutralizing antibodies on the basis of the aminoacidic sequence, the crystal structure can drive in the determination of the residues of the donor antibody to be retained in the humanized versions. The donor residues selected for modification may include not only residues belonging to the hypervariable regions, but also those supporting or proximal to the CDR loop.

In conclusion, these structures provide the basis by which framework residues influencing CDR conformations may be determined and together with the tools of protein engineering these structural data offer the potential to extend the medical, diagnostic and therapeutic uses of these blocking antibodies.

Design of CDRs peptides

Despite humanization the use of intact antibodies as clinical tools may still remain problematic, largely because of their size, that limits tissue penetration (including tumors). These unfavorable features reduced their applications in development of drugs and their use as diagnostics.

The application of a reductionist approach to protein design and structural biology has enabled the minimal functional domains of various proteins to be determined. This strategy can be particularly useful for the design of antibody-derived small molecules for a further reduction of Fab and Fv fragments, both obtained by proteolytic cleavage (Figure 1).

As a result synthetic peptides derived from CDRs sequences have shown in many cases to possess similar properties to the intact antibody (Kang *et al.*, 1988; Taub *et al.*, 1989; Welling *et al.*, 1991, Levi *et al.*, 1993, Jarrin *et al.*, 1994, Smith *et al.*, 1994, Igarashi *et al.*, 1995, Zhang *et al.*, 1996 and 1997, Laune *et al.*, 1997).

These peptides can inhibit idiotype-anti-idiotype interactions, bind specific antigens, interact with cellular receptors and stimulate biological processes. However these biological and therapeutic applications are often precluded by the high flexibility of these molecules, whose structure is strongly influenced by environment (Hrubi *et al.*, 1990; Tobias *et al.*, 1990) leading to random conformations in solution. Moreover this high conformational flexibility, together with the volume to surface ratio, appear to be unfavorable for proper folding (Mutter, 1988). As a consequence linear peptides often aggregate and do not acquire an intramolecular folded state. Analyzing key structure adopted by proteins in their native states, loops and turn in α-helices are crucial in permitting suitable orientations of binding groups as well as bio-availability and stability (Rose *et al.*, 1985; Martin *et al.*, 1989).

Thus, starting from CDRs sequences, it has been shown that it is possible to develop constrained peptidic loops that retain biological and antigen binding activity (Williams *et al.*, 1988 and 1989). Cyclic peptide analogs, whose constrained structure is similar to that in the intact antibody (such a reverse turn), can have enhanced activity. (Williams *et al.*, 1991). Peptides can be circularized by incorporation of cystein residues and

oxidation (or incorporation of organic linkers): this intramolecular covalent disulphide bond restricts the conformation of the peptide in order to mimic more accurately the secondary structure.

However the successful strategy of utilizing peptide analogs to retain pharmacophoric information has resulted in the notion that the spatial arrangement of the active groups of the amino acid side chains is the most important feature of biological specificity (Williams *et al.*, 1991 and 1989, Ishida *et al.*, 1988). Then the specificity of biological interactions depends on two factors: the ensemble of pharmacophoric information present in the primary sequence, and the three dimensional presentation of the active groups of the amino acid side chains. These two structural features successfully combine to produce a molecule, whose active groups are exposed in the appropriate orientation and space. As a consequence the ideal starting point for the rational design of small molecule analogs is the structural analysis.

From this point of view structural information on both FabMNAC13 and FabαD11 is expected to provide indications on the crucial group residues involved in the interaction. Indeed the key point for specificity and affinity of the synthetic looped CDRs is the appropriate constraint, that can be precisely established in terms of residue positioning using structural data.

Rational drug design and development of peptides mimics

However because of their proteinaceous nature peptide analogs tend to be subjected to many of the major drawbacks of intact antibodies. As a consequence the usefulness of peptide analogs *in vivo* tends to be limited (Williams *et al.*, 1991).

Structural information are the basis to identify relevant contact residues and conformations involved in antigen binding and a first and preliminary

step in the design and synthesis of organic compounds with biological activity and without the limitations of peptides. Therefore such structurally mimics should be water-soluble and, due to their synthetic nature and size (less than 1KDa), should be non-immunogenic, permeable to blood-brain barrier and resistant to protease degradation because they contain only one or two peptide bonds.

A simple strategy has been developed to synthesize organic CDR structures on macrocyclic ring structure, that can be altered to mimic the predicted and biologically active conformation of hypervariable loop structures.

Structural information on both FabMNAC13 and FabαD11 together with computer aided molecular modeling may facilitate the design of these nonpeptidic CDRs, that have been shown to be non toxic, resistant to peptidases and soluble (Saragovi et al., 1991). The construction of CDR mimetics is achieved by a retrosynthetic modular component synthesis strategy, which introduces natural and synthetic amino acid side chains into a ring structure. The mimetic is conformationally restricted in a loop configuration by a "bridge" (the X group linker Figure 3) and the possibility to alter it permits the variation of side chain orientations, backbone distances and dihedral angles.

Besides this approach these structural information of both neutralizing antibodies are expected to be the basis for a wide range computer aided molecular modeling to developing small and proteolytically stable molecules with agonistist or antagonist activity.

References

Amit, A. G., Mariuzza, R. A., Phillips, S. E., Poljak, R. J. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. *Science*; **233**, 747-753

Bach, J. F., Fracchia, G. N., Chatenoud, L. (1993). Safety and efficacy of therapeutic monoclonal antibodies in clinical therapy. *Immunol Today*; **14**, 421-425

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M., and Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature* 342, 877-883

Foote, J. & Winter , G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol.;* **224**, 487-499.

Hruby, V. J., al-Obeidi, F., Kazmierski, W. (1990). Emerging approaches in the molecular design of receptor-selective peptide ligands: conformational, topographical and dynamic considerations. *Biochem J*; **268**, 249-262

Jarrin, A., Andrieux, A., Chapel, A., Buchou, T., Marguerie, G. (1994). A synthetic peptide with anti-platelet activity derived from a CDR of an anti-GPIIb-IIIa antibody. *FEBS Lett*; **354,** 169-172

Kang, C. Y., Brunck, T. K., Kieber-Emmons, T., Blalock, J. E., Kohler, H. (1988). Inhibition of self-binding antibodies (autobodies) by a VH-derived peptide. *Science*; **20**;2401034-2401036

Igarashi, K., Asai, K., Kaneda, M., Umeda, M., Inoue, K. (1995). Specific binding of a synthetic peptide derived from an antibody complementarity determining region to phosphatidylserine. *J Biochem (Tokyo)*; **117**, 452-457.

Ishida, T., Yoneda, S., Doi, M., Inoue, M., Kitamura, K. (1988). Molecular-dynamics simulations of [Met5]- and [D-Ala2,Met5]-enkephalins. Biological implication of monomeric folded and dimeric unfolded conformations. *Biochem J*; **255**, 621-628

Laune, D., Molina, F., Ferrieres, G., Mani, J. C., Cohen, P., Simon, D., Bernardi, T., Piechaczyk, M., Pau, B., Granier, C. (1997). Systematic exploration of the antigen binding activity of synthetic peptides isolated from the variable regions of immunoglobulins. *J Biol Chem*; **272**, 30937-30944

Levi, M., Sallberg, M., Ruden, U., Herlyn, D., Maruyama, H., Wigzell, H., Marks, J., Wahren, B. (1993). A Complementarity-Determining Region Synthetic Peptide Acts as a Miniantibody and Neutralizes Human Immunodeficiency Virus Type 1 *in vitro*. *PNAS*; **90**, 4374-4378.

Martin, A. C., Cheetham, J. C., Rees, A. R. (1989). Modeling antibody hypervariable loops: a combined algorithm. *Proc Natl Acad Sci U S A*; **86**, 9268-9272

Mutter. M. (1988). Nature's rules and chemist's tools: a way for creating novel proteins. *Trends Biochem Sci*; **13**, 260-265

Novotny. J., Bruccoleri, R. E., Haber, E. (1990). Computer analysis of mutations that affect antibody specificity. *Proteins*; **7**, 93-98

Kohler, G. and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*; **256**, 495-497

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J., Davies, D. R. (1989). Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *PNAS U S A*; **86**, 5938-5942

Presta, L. G., Lahr, S. J., Shields, R. L., Porter, J. P., Gorman, C. M., Fendly, B. M., Jardieu, P. M. (1993). Humanization of an antibody directed against IgE. *J Immunol*; **151**, 2623-2632

Rose, G. D., Gierasch, L. M., Smith, J. A. (1985). Turns in peptides and proteins. *Adv Protein Chem*; 37, 1-109

Saragovi, H. U., Fitzpatrick, D., Raktabutr, A., Nakanishi, H., Kahn, M., Greene, M. I. (1991), Design and synthesis of a mimetic from an antibody complementarity-determining region. *Science*; **253**, 792-795

Schillbach, J. F., Near, R. .I, Bruccoleri, R. E., Haber, E., Jeffrey, P. D., Novotny, J., Sheriff, S., Margolies, M. N. (1993). Modulation of antibody affinity by a non-contact residue. *Protein Sci.*; **2**, 206-14.

Smith, J. W., Hu, D., Satterthwait, A., Pinz-Sweeney, S. and Barbas 3rd, C. F. (1994). Building synthetic antibodies as adhesive ligands for integrins. *J. Biol. Chem*; 269, 32788-32795

Taub, R., Gould, R. J., Garsky, V. M., Ciccarone, T. M., Hoxie, J., Friedman, P. A., Shattil, S. J. (1989) A monoclonal antibody against the platelet fibrinogen receptor contains a sequence that mimics a receptor recognition domain in fibrinogen. *J Biol Chem*, **264**, 259-265

Taub, R. and Greene, M. I. (1992). Functional validation of ligand mimicry by anti-receptor antibodies: structural and therapeutic implications. *Biochemistry*; **31**, 7431-7435. Review.

Tempest, P. R., Bremner, P., Lambert, M., Taylor, G., Furze, J. M., Carr, F. J., Harris, W. J. (1991). Reshaping a human monoclonal antibody to inhibit human respiratory syncytial virus infection in vivo. *Biotechnology*; **9**, 266-271

Tobias, D. J., Sneddon, S. F., Brooks, C. L. 3rd. (1990). Reverse turns in blocked dipeptides are intrinsically unstable in water. *J Mol Biol*; **216**, 783-796.

Tramontano, A., Chothia, C., Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol*; **215**, 175-82

Welling, G. W., van Gorkum, J., Damhof, R. A., Drijfhout, J. W., Bloemhoff, W., Welling-Wester, S. (1991). A ten-residue fragment of

an antibody (mini-antibody) directed against lysozyme as ligand in immunoaffinity chromatography. *J Chromatogr*; **548**, 235-242

Winter, G. & Harris, W. J. (1993). Humanized antibodies. *Immunol Today*; 14, 243-246. Review.

Winter, G. & Milstein, C. (1991). Man-made antibodies. *Nature*; **349**, 293-299

Williams, W. V., Kieber-Emmons, T., VonFeldt, J., Greene, M. I., Weiner, D. B. (1991). Design of bioactive peptides based on antibody hypervariable region structures. Development of conformationally constrained and dimeric peptides with enhanced affinity. *J Biol Chem*; **266**, 5182-5190

Williams, W. V., Weiner, D. B., Kieber-Emmons, T., Greene, M. I. (1990). Antibody geometry and form: three-dimensional relationships between anti-idiotypic antibodies and external antigens. *Trends Biotechnol*; **8**, 256-63

Williams, W. V., Moss, D. A., Kieber-Emmons, T., Cohen, J. A., Myers, J. N., Weiner, D. B., Greene, M. I. (1989). Development of biologically active peptides based on antibody structure. *Proc Natl Acad Sci U S A*; **86**, 5537-5541

Williams, W. V., Guy, H. R., Rubin, D. H., Robey, F., Myers, J. N., Kieber-Emmons, T., Weiner, D. B., Greene, M. I. (1988). Sequences of the cell-attachment sites of reovirus type 3 and its anti-idiotypic/antireceptor antibody: modeling of their three-dimensional structures. *Proc Natl Acad Sci U S A*; **85**, 6488-6492

Zhang, X., Piatier-Tonneau, D., Auffray, C., Murali, R., Mahapatra, A., Zhang, F., Maier, C. C., Saragovi, H., Greene, M. I. (1996). Synthetic CD4 exocyclic peptides antagonize CD4 holoreceptor binding and T cell activation. *Nat Biotechnol*; **14**, 472-475

Zhang, X., Gaubin, M., Briant, L., Srikantan, V., Murali, R., Saragovi, U., Weiner, D., Devaux, C., Autiero, M., Piatier-Tonneau, D., Greene, M. I. (1997). Synthetic CD4 exocyclics inhibit binding of human immunodeficiency virus type 1 envelope to CD4 and virus replication in T lymphocytes. *Nat Biotechnol*; **15**, 150-154

Ringraziamenti

Sembrera' forse anacronistico in una tesi di PhD, ma tengo moltissimo a scrivere questa pagina nella mia lingua madre, l'unico modo per poter sperare di esprimere almeno vagamente la mia gratitudine. Ho l'impressione di padroneggiare l'inglese troppo poco per poter dar voce al mio cuore.

Mi risulta alquanto arduo poter sintetizzare in poche righe piu' di quattro anni di esperienze condivise con moltissime persone davvero speciali (e non solo dal punto di vista lavorativo), che mi sono state accanto a diverso titolo, ma con incredibile pazienza e disponibilita' (non ultimo il mitico autista della 51, che piu' volte ha avuto la pieta' di raccogliermi sulla strada da Basovizza a Elettra).

Naturalmente il mio piu' sentito ringraziamento va al mio supervisor Antonino Cattaneo, prima di tutto per avermi coinvolto nel mondo incredibilmente affascinanante e stimolante della cristallografia. Un grazie particolare per l'entusiasmo e gli stimoli che ha saputo infondermi durante lunghe discussioni scientifiche (anche nei momenti piu' critici) e per la fiducia (spero non mal riposta) che ha sempre dimostrato nei miei confronti. Credo sia un modello, sia per apertura mentale e approccio scientifico che per coraggio nell'abbracciare una disciplina nuova e, non sempre banale da capire (in termini di tempi e difficolta') qual e' la cristallografia. Va da se' che a questo punto il mio pensiero vada a chi mi ha pazientemente e (di recente) quotidianamente seguito in questo nuovo campo, Doriano Lamba, che mi piace definire (a suo malincuore) maestro, per la sua competenza e la sua attitudine nei confronti di noi studenti, da cui traspare sollecito interesse e affetto. Trovo che sia stato un grande privilegio per me essere seguita da una persona che, pur nella

sua grande esperienza, sa essere estremamente chiara, didattica e paziente. La sua perizia e disponibilita' mi sono state preziose non solo nell'apprendere concetti e un vero e proprio metodo, ma anche e soprattutto durante la stesura di questa tesi.

Un apporto fondamentale per la realizzazione di questo lavoro proviene da quel gruppo di persone che piu' che colleghi, sento quasi ormai membri di una famiglia allargata, prima di tutto "mamma" Gabri, che insieme alla migliore amica che potessi mai incontrare, il nostro splendido tecnico Jessica, ha saputo piu' volte raccogliere i cocci spronandomi a continuare. Devo ringraziarle per saputo, insieme a Giada, Sergio e Federica, esserci e condividere il bello e il brutto di quattro anni di vita insieme, non solo scientificamente ma soprattutto emotivamente. Un grazie di tutto cuore a Monica, Francesca, Teresa e Marie, splendide amiche dentro e fuori il lab. Un caro ringraziamento anche a Luisa per la dolcezza e l'entusiasmo con cui ha seguito i miei progressi, a Roberto con I suoi preziosi insegnamenti anche fuori delle aule universiarie, a Kevin e Michela per le utili discussioni e "dritte" scientifiche, a Miranda, Silvia e Mauro, tutte persone davero speciali.

Nell'ultimo anno mi sono poi sentita adottata anche dagli splendidi ragazzi del biolab di Elettra, fra cui in ordine sparso ricordero' sempre con grande affetto la dolcezza di Ivette, la simpatia di Stefano e Maurizio, la disponibilita' di Cristina e le discussioni scientifiche, sporadiche ma non meno illuminanti, con Kristina Dijnovic e Massimo Degano. Un enorme debito di gratitutine e' quello che mi lega a Kristina e Jochen, che hanno raccolto in mia vece il set di dati ad alta risoluzione per MNAC13 a ESRF, per non parlare di Khristian, incomparabile compagno (provvidenziali le sue sopraciglia per separare dal fascio un cristallo singolo per MNAC13), insieme a Doriano della mia prima esperienza sulla beamline di Elettra, e naturalmente a Cecilia, la persona che piu'

cose ha saputo insegnarmi per unita' di tempo (visto l'unico weekend passato insieme sulla beamline). Passando poi allo step successivo negli ultimi mesi ho avuto la fortuna di essere affiancata da Alberto Cassetta, che ha saputo essere un degno successore di Doriano sia scientificamente che umanamente (il che e' tutto dire) nel processamento dei dati di αD11. Accanto a Jessica, un'altra persona che spero di non perdere mai e' Elena, factotum non solo nel suo lab, ma anche come amica (ex-coinquilina, taxista per lunghe distanze, psicologa e all'occorrenza, psichiatra, per non parlare di…. Cupido!).

Anche se sembrera' banale il debito piu' profondo e' quello che mi lega alla mia famiglia, non solo ai miei genitori ma soprattutto alla mia inossidabile e indistruttibile nonna Ilde, mia prima instancabile sostenitrice e degna sostituta del mio adorato nonno Arnaldo (che troppo presto mi ha lasciato) nell'entusiasmo e nell'affetto per ogni mio piu' banale gesto, amorosa parzialita' che e' cibo per la mia anima perennemente insicura. Non credo di poter esprimere l'amore che provo per la mia famiglia, a cui nonostante l'eta' sono legata indissolubilmente: la mia isola di pace al di la' di tutto, una specie di Arca di Noe' (vista la copiosa rappresentanza animale: solo due gatti, Beauty e Morello, purtroppo visto che Minou non c'e' piu', la mia adorata cricetina San Souci, la tartaruga Kevin e le new entries, due bruchi stecco) sulla quale posso rifugiarmi e sapere di trovare la migliore amica che si possa avere: mia madre.

Ultimo in questo excursus sicuramente incompleto e magari anche cronologicamente, ma in pole position nel mio cuore, Paolo che ha saputo ridarmi la prospettiva del mondo, cambiandomi la vita:

Tu, perso nelle distanze interstellari

Io, persa nelle distanze interatomiche.

In noi gli estremi del cerchio dell'universo si uniscono.