Structural Study of Human Antibody Fragments with Specificity for Mucin-1 Antigen

Von der Fakultät für Mathematik, Informatik and Naturwissenschaften der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grade eines Doktors der Naturwissenschaften genehmigte Dissertation

targeted by

vorgelegt von

Master of Technology
Rajneesh Kumar Gaur
aus
Neu-Delhi, Indien

Berichter: Universitätsprofessor Dr. rer. nat. Rainer Fischer
Universitätsprofessor Dr. rer. nat. Fritz Kreuzaler

Tag der mündlichen Prüfung: 27. September 2004

Diese Dissertation ist auf Internetseiten der Hochschulbibliothek online verfügbar
‘Man has never learned to simplify, only to complicate. Often he does not say what he feels but he thinks– and sometimes he thinks too much and, in the process, gains much knowledge but also larger confusion and greater grief’

From ‘Songs of Ganga’-5455 BC

Dedicated to India
## CONTENTS

I  INTRODUCTION  
   I.1  Antibodies  
      I.1.1  Overview  
      I.1.2  Antibody engineering  
      I.1.3  Therapeutic potential of antibodies  
      I.1.4  Recombinant small antibody fragments  
      I.1.5  Importance of antibody fragments  
   I.2  The human V\(_H\) family  
      I.2.1  Variable heavy chain domain  
         I.2.1.1  General structure  
         I.2.1.2  Diversity of the V\(_H\) domain  
         I.2.1.3  Conformations of the hypervariable loops  
         I.2.1.4  V\(_H\) domain stability  
      I.2.2  Comparison between human V\(_H\) and camelids V\(_{HH}\) domains  
         I.2.2.1  Sequence differences  
         I.2.2.2  Structural differences  
      I.2.3  Camelization of the V\(_H\) domain  
   I.3  Mucin  
      I.3.1  Overview  
      I.3.2  Human mucin-1  
      I.3.3  MUC1 and cancer  
      I.3.4  Antibodies against MUC1  
   I.4  Research objectives  

II  MATERIALS AND METHODS  
   II.1  Materials  
      II.1.1  Chemicals  
      II.1.2  Laboratory consumables  
      II.1.3  Laboratory equipment  
      II.1.4  Chromatography columns  
      II.1.5  Media, stock solutions and buffers  
      II.1.6  Antibodies, antiseras and peptides  
      II.1.7  Bacterial strains  
      II.1.8  Software  
   II.2  Methods  
      II.2.1  Recombinant DNA techniques  
         II.2.1.1  Analytical gel electrophoresis  
         II.2.1.2  Preparative gel electrophoresis  
         II.2.1.3  PCR amplification of DNA segments  
            II.2.1.3.1  Native M12-V\(_H\) domain  
            II.2.1.3.2  Camelized M12-V\(_H\) domain  
         II.2.1.4  DNA digestion  
         II.2.1.5  Ligation of DNA inserts  
         II.2.1.6  PCR based analysis of recombinant bacterial clones  
         II.2.1.7  Isolation of plasmid DNA from E. coli  
         II.2.1.8  DNA sequencing and sequence analysis  
         II.2.1.9  Preparation of E. coli competent cells  
         II.2.1.10  Preparation of heat shock E. coli competent cells  
            II.2.1.10.1  Transformation of heat shock E. coli competent cells  

1  INTRODUCTION  

I  INTRODUCTION  

I.1  Antibodies  

I.1.1  Overview  

I.1.2  Antibody engineering  

I.1.3  Therapeutic potential of antibodies  

I.1.4  Recombinant small antibody fragments  

I.1.5  Importance of antibody fragments  

I.2  The human V\(_H\) family  

I.2.1  Variable heavy chain domain  

I.2.1.1  General structure  

I.2.1.2  Diversity of the V\(_H\) domain  

I.2.1.3  Conformations of the hypervariable loops  

I.2.1.4  V\(_H\) domain stability  

I.2.2  Comparison between human V\(_H\) and camelids V\(_{HH}\) domains  

I.2.2.1  Sequence differences  

I.2.2.2  Structural differences  

I.2.3  Camelization of the V\(_H\) domain  

I.3  Mucin  

I.3.1  Overview  

I.3.2  Human mucin-1  

I.3.3  MUC1 and cancer  

I.3.4  Antibodies against MUC1  

I.4  Research objectives  

II  MATERIALS AND METHODS  

II.1  Materials  

II.1.1  Chemicals  

II.1.2  Laboratory consumables  

II.1.3  Laboratory equipment  

II.1.4  Chromatography columns  

II.1.5  Media, stock solutions and buffers  

II.1.6  Antibodies, antiseras and peptides  

II.1.7  Bacterial strains  

II.1.8  Software  

II.2  Methods  

II.2.1  Recombinant DNA techniques  

II.2.1.1  Analytical gel electrophoresis  

II.2.1.2  Preparative gel electrophoresis  

II.2.1.3  PCR amplification of DNA segments  

II.2.1.3.1  Native M12-V\(_H\) domain  

II.2.1.3.2  Camelized M12-V\(_H\) domain  

II.2.1.4  DNA digestion  

II.2.1.5  Ligation of DNA inserts  

II.2.1.6  PCR based analysis of recombinant bacterial clones  

II.2.1.7  Isolation of plasmid DNA from E. coli  

II.2.1.8  DNA sequencing and sequence analysis  

II.2.1.9  Preparation of E. coli competent cells  

II.2.1.10  Preparation of heat shock E. coli competent cells  

II.2.1.10.1  Transformation of heat shock E. coli competent cells
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.2.1.10.2 Efficiency of bacterial transformation</td>
<td>36</td>
</tr>
<tr>
<td>II.2.1.10.3 Culturing and glycerol stock preparation</td>
<td>36</td>
</tr>
<tr>
<td>II.2.2 Bacterial expression of antibody fragments</td>
<td>36</td>
</tr>
<tr>
<td>II.2.2.1 Small scale expression</td>
<td>37</td>
</tr>
<tr>
<td>II.2.2.2 Large scale expression of scFvM12</td>
<td>37</td>
</tr>
<tr>
<td>II.2.3 Purification of antibody fragments</td>
<td>38</td>
</tr>
<tr>
<td>II.2.3.1 scFvM12 purification</td>
<td>38</td>
</tr>
<tr>
<td>II.2.3.1.1 Periplasmic extraction</td>
<td>38</td>
</tr>
<tr>
<td>II.2.3.1.2 Ni-NTA chromatography</td>
<td>38</td>
</tr>
<tr>
<td>II.2.3.1.3 Cation-exchange chromatography</td>
<td>38</td>
</tr>
<tr>
<td>II.2.3.1.4 Anion-exchange purification</td>
<td>39</td>
</tr>
<tr>
<td>II.2.3.1.5 Size-exclusion chromatography</td>
<td>39</td>
</tr>
<tr>
<td>II.2.3.2 Purification of the cytoplasmic M12-V_{H} fragment</td>
<td>39</td>
</tr>
<tr>
<td>II.2.3.3 M12-V_{H} fragment inclusion bodies</td>
<td>39</td>
</tr>
<tr>
<td>II.2.3.3.1 Soluble M12-V_{H} fragment purification</td>
<td>40</td>
</tr>
<tr>
<td>II.2.4 Protein analysis</td>
<td>40</td>
</tr>
<tr>
<td>II.2.4.1 Concentration and dialysis</td>
<td>40</td>
</tr>
<tr>
<td>II.2.4.2 Protein quantification</td>
<td>40</td>
</tr>
<tr>
<td>II.2.4.3 SDS-PAGE and IEF</td>
<td>41</td>
</tr>
<tr>
<td>II.2.4.4 Western blotting</td>
<td>42</td>
</tr>
<tr>
<td>II.2.4.5 Silver staining</td>
<td>42</td>
</tr>
<tr>
<td>II.2.4.6 Discontinuous non-denaturing electrophoresis</td>
<td>43</td>
</tr>
<tr>
<td>II.2.5 Stability analysis</td>
<td>43</td>
</tr>
<tr>
<td>II.2.6 Mass spectrometry</td>
<td>43</td>
</tr>
<tr>
<td>II.2.7 Flow cytometric analysis</td>
<td>44</td>
</tr>
<tr>
<td>II.2.8 M12-V_{H} fragment generation</td>
<td>45</td>
</tr>
<tr>
<td>II.2.9 Crystallization</td>
<td>46</td>
</tr>
<tr>
<td>II.2.9.1 Hanging drop method</td>
<td>46</td>
</tr>
<tr>
<td>II.2.9.1.1 M12-V_{H} domain crystallization</td>
<td>46</td>
</tr>
<tr>
<td>II.2.9.1.2 scFvM12 crystallization</td>
<td>48</td>
</tr>
<tr>
<td>II.2.9.1.3 M12-V_{H} and MUC1 complex crystallization</td>
<td>50</td>
</tr>
<tr>
<td>II.2.9.1.4 scFvM12 and MUC1 complex crystallization</td>
<td>50</td>
</tr>
<tr>
<td>II.2.9.2 Gel crystallization</td>
<td>51</td>
</tr>
<tr>
<td>II.2.9.3 Streak seeding</td>
<td>52</td>
</tr>
<tr>
<td>II.2.9.4 Crystal crystallization in meta-stable zone</td>
<td>52</td>
</tr>
<tr>
<td>II.2.9.5 Macroseeding</td>
<td>53</td>
</tr>
<tr>
<td>II.2.10 Crystal mounting</td>
<td>53</td>
</tr>
<tr>
<td>II.2.10.1 M12-V_{H} domain alone and in complex with MUC1</td>
<td>53</td>
</tr>
<tr>
<td>II.2.10.2 scFvM12 antibody fragment</td>
<td>53</td>
</tr>
<tr>
<td>II.2.10.3 scFvM12 antibody fragment in complex with MUC1 antigen</td>
<td>54</td>
</tr>
<tr>
<td>II.2.11 Data collection and data processing</td>
<td>54</td>
</tr>
<tr>
<td>II.2.12 Phasing</td>
<td>55</td>
</tr>
<tr>
<td>II.2.12.1 Molecular replacement</td>
<td>55</td>
</tr>
<tr>
<td>II.2.12.1.1 M12-V_{H} domain (1.8Å)</td>
<td>55</td>
</tr>
<tr>
<td>II.2.12.1.2 M12-V_{H} domain in complex with MUC1 peptide (1.5Å)</td>
<td>55</td>
</tr>
<tr>
<td>II.2.13 Model building and refinement</td>
<td>55</td>
</tr>
<tr>
<td>II.2.13.1 M12-V_{H} domain (1.8Å)</td>
<td>55</td>
</tr>
<tr>
<td>II.2.13.2 M12-V_{H} domain in complex with MUC1 peptide (1.5Å)</td>
<td>56</td>
</tr>
<tr>
<td>II.2.14 Structure validation</td>
<td>57</td>
</tr>
<tr>
<td>II.2.15 Structure interpretation</td>
<td>57</td>
</tr>
<tr>
<td>III RESULTS</td>
<td>58</td>
</tr>
</tbody>
</table>
III.1 Bacterial expression of scFvM12 .................................................. 58
  III.1.1 Small scale expression ............................................................. 58
  III.1.2 Large scale expression ........................................................... 58

III.2 Purification of the scFvM12 ......................................................... 60
  III.2.1 Ni-NTA chromatography ......................................................... 60
  III.2.2 Ion-exchange chromatography ................................................. 61
  III.2.3 Size-exclusion chromatography .............................................. 62
  III.2.4 Iso-electric focusing and mass spectrometry ............................ 63

III.3 Crystallization of antibody fragments ........................................ 65
  III.3.1 Crystallization of scFvM12 ..................................................... 65
  III.3.2 Limited proteolysis of scFvM12 .............................................. 66
  III.3.3 M12-V_H domain crystallization ............................................ 68
    III.3.3.1 By subtilisin treatment of scFvM12----------------------------- 69
    III.3.3.2 By thermolysin treatment of scFvM12 ................................ 72
  III.3.4 Crystallization of M12-V_H in complex with MUC1 antigen ............ 73
  III.3.5 Co-crystallization of scFvM12 in complex with MUC1 peptide ........ 74

III.4 Data Collection ............................................................................. 75
  III.4.1 scFvM12 .................................................................................... 75
  III.4.2 scFvM12 in complex with MUC1 peptide .................................... 75
  III.4.3 M12-V_H domain ..................................................................... 75

III.5 Quality of M12-V_H domain dataset ............................................ 77

III.6 Model building and refinement for the M12-V_H domain ................ 78

III.7 Overall structure of the M12-V_H domain .................................... 79

III.8 Structure description of the M12-V_H domain ............................... 81
  III.8.1 Antigen binding loops of the M12-V_H domain ......................... 83
    III.8.1.1 CDR-H1 loop ................................................................. 84
    III.8.1.2 CDR-H2 loop ................................................................. 85
    III.8.1.3 CDR-H3 loop ................................................................. 87
      III.8.1.3.1 CDR-H3 torso ......................................................... 87
      III.8.1.3.2 CDR-H3 apex ......................................................... 88
  III.8.2 Hydrogen bonding .................................................................. 88

III.9 M12-V_H structure quality assessment ....................................... 91

III.10 Crystal packing ............................................................................ 94

III.11 M12-V_H domain crystallization in complex with MUC1 peptide .... 95

III.12 Stability analysis of scFvM12 ..................................................... 96

III.13 M12-V_H domain engineering .................................................... 98
  III.13.1 Cloning strategy ................................................................. 99
  III.13.2 Amplification of the M12-V_H domain .................................... 99
  III.13.3 Restriction and ligation ....................................................... 100
    III.13.3.1 M12-V_H domain ......................................................... 101
    III.13.4 M12-V_H domains ............................................................ 102

IV DISCUSSION ....................................................................................... 104
  IV.1 Expression and purification of scFvM12 ...................................... 105
    IV.1.1 Bacterial expression ............................................................. 105
    IV.1.2 scFvM12 purification ........................................................... 106
  IV.2 Crystallization of antibodies ..................................................... 107
    IV.2.1 Crystallization of scFvM12 .................................................... 107
    IV.2.2 Single antibody domain crystallization .................................... 108
    IV.2.3 Preference for the M12-V_H domain crystallization over M12-V_L domain ................................................................. 109
    IV.2.4 Diffusion over co-crystallization .......................................... 110
<table>
<thead>
<tr>
<th>IV.3</th>
<th>Structural deviations of the M12-V&lt;sub&gt;H&lt;/sub&gt; domain</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV.4</td>
<td>Structural basis of the M12-V&lt;sub&gt;H&lt;/sub&gt; domain stability</td>
<td>111</td>
</tr>
<tr>
<td>IV.5</td>
<td>The V&lt;sub&gt;H&lt;/sub&gt;/V&lt;sub&gt;L&lt;/sub&gt; interface of the M12-V&lt;sub&gt;H&lt;/sub&gt; domain</td>
<td>112</td>
</tr>
<tr>
<td>IV.6</td>
<td>Comparison between the SM3-V&lt;sub&gt;H&lt;/sub&gt; domain and the M12-V&lt;sub&gt;H&lt;/sub&gt; domain</td>
<td>113</td>
</tr>
<tr>
<td>IV.7</td>
<td>Camelization of the M12-V&lt;sub&gt;H&lt;/sub&gt; domain</td>
<td>116</td>
</tr>
<tr>
<td>IV.8</td>
<td>Comparison between the M12-V&lt;sub&gt;H&lt;/sub&gt; domain and the camelized human V&lt;sub&gt;H&lt;/sub&gt;-P8 antibody</td>
<td>117</td>
</tr>
<tr>
<td>IV.9</td>
<td>M12-V&lt;sub&gt;H&lt;/sub&gt; domain cloning</td>
<td>119</td>
</tr>
<tr>
<td>IV.10</td>
<td>Biological implications of single domain antibody</td>
<td>120</td>
</tr>
<tr>
<td>IV.11</td>
<td>Conclusion</td>
<td>121</td>
</tr>
<tr>
<td>V</td>
<td>REFERENCES</td>
<td>123</td>
</tr>
<tr>
<td>VI</td>
<td>APPENDIX</td>
<td>142</td>
</tr>
<tr>
<td>VI.1</td>
<td>scFvM12 antibody sequence</td>
<td>142</td>
</tr>
<tr>
<td>VI.2</td>
<td>DNA sequence of M12-V&lt;sub&gt;H&lt;/sub&gt; domain</td>
<td>142</td>
</tr>
<tr>
<td>VI.3</td>
<td>Primers</td>
<td>143</td>
</tr>
<tr>
<td>VI.4</td>
<td>M12-V&lt;sub&gt;H&lt;/sub&gt; domain (1.8Å)</td>
<td>144</td>
</tr>
<tr>
<td>VI.5</td>
<td>Intermolecular crystal contacts for M12-V&lt;sub&gt;H&lt;/sub&gt; domain (1.5Å)</td>
<td>146</td>
</tr>
<tr>
<td>VI.6</td>
<td>Hydrogen bonding pattern of M12-V&lt;sub&gt;H&lt;/sub&gt; domain (1.5Å)</td>
<td>147</td>
</tr>
<tr>
<td>VI.7</td>
<td>Structural studies of macromolecules</td>
<td>148</td>
</tr>
<tr>
<td>VI.7.1</td>
<td>Overview</td>
<td>148</td>
</tr>
<tr>
<td>VI.7.2</td>
<td>Methodology</td>
<td>149</td>
</tr>
<tr>
<td>VI.7.3</td>
<td>Data collection</td>
<td>150</td>
</tr>
<tr>
<td>VI.7.4</td>
<td>Phase problem and related methodology</td>
<td>152</td>
</tr>
<tr>
<td>VI.7.5</td>
<td>Ab-initio phase determination</td>
<td>152</td>
</tr>
<tr>
<td>VI.7.6</td>
<td>Molecular replacement</td>
<td>153</td>
</tr>
<tr>
<td>VI.7.7</td>
<td>Model building and refinement</td>
<td>153</td>
</tr>
<tr>
<td>VI.7.8</td>
<td>Structure validation and submission</td>
<td>154</td>
</tr>
<tr>
<td>VI.7.9</td>
<td>Vector maps</td>
<td>155</td>
</tr>
<tr>
<td>VI.8</td>
<td>List of abbreviations</td>
<td>156</td>
</tr>
<tr>
<td>VI.9</td>
<td>List of tables</td>
<td>160</td>
</tr>
<tr>
<td>VI.10</td>
<td>List of figures</td>
<td>160</td>
</tr>
</tbody>
</table>
ABSTRACT

MUC1 is overexpressed in many adenocarcinomas, which makes it a potential target for immunotherapy. A large number of MUC1-specific, monoclonal antibodies have been produced but the stability, solubility and the size of the antibody determines their effectiveness in cancer therapy. Smaller antibody fragments are advantageous over full-size antibodies, since small antigen binding molecules can efficiently penetrate solid tumors. As a first step towards the design of a clinically desirable antibody fragment, the crystallization of a human VH fragment has been achieved. The fragment was derived from the human single chain antibody scFvM12, which recognizes the cancer-specific hypoglycosylated MUC1. The human M12-VH domain has been crystallized through limited in-vitro proteolysis of scFvM12 antibody fragment. Crystals of M12-VH domain grew in approximately 12 months up to a size of 180x60x60µm in 100mM MES, pH 6.5, 5mM zinc sulphate, 25 % v/v polyethylene glycol 550 monomethylether in the presence of low concentration of subtilisin Carlsberg. The protease addition results in complete degradation of M12-VA domain, linker and purification tags. The crystal belongs to space group C2 with unit cell dimension a = 71.34Å, b = 37.97Å, c = 37.19Å and α = γ = 90°, β = 109.674°. The solvent content of the crystal was calculated as 35.94%. The structure was solved up to 1.5Å resolution with an R_cryt factor of 15.8%, an R_free of 19.7% and possesses a good stereochemistry. Dihedral angle values comparison of first and second complementarity-determining region (CDR) of M12-VH domain with an average value, for these two hypervariable regions, shows a significant deviation and therefore, it can be speculated that the M12-VH either suggests the existence of new canonical subclass or a link among the subclasses of canonical main-chain conformation in VH3 family.

The natural existence of the stabilizing mutations Glu-H6, Arg-H66 and Ser-H52 confers the intrinsic stability to M12-VH domain under reducing conditions and therefore, the framework of this antibody domain can be used for incorporation of other specificities. This result would be helpful in structure based single domain antibody designing for biotechnological and pharmaceutical applications especially against MUC1 related cancer.

The isolated VH domain undergoes aggregation due to the exposure of the hydrophobic surface involved in VH/VL interface. The camelization of the position H44, H45 and H47 can increase the hydrophilicity of the VH domain. Native and camelized M12-VH domain cloned in cytoplasmic expression vector would be helpful in assessing the immunogenic effects of camelizing the human VH domain.
I INTRODUCTION

Mucin-1 (MUC1) is a transmembrane glycoprotein, protecting the epithelia from infection, dehydration, physical or chemical injury, which is also involved in cell-cell interactions, signaling, and metastasis (Parry et al., 2001). MUC1 is polymorphic in nature due to the occurrence of variable number of repeating amino acid segments called tandem repeats (Hinoda et al., 1994). The tandem repeats of MUC1 consists of 20 amino acids ‘APDTRPAPGSTAPPAHGVTSA’. It has been observed that MUC1 expression is frequently elevated or altered in cancer and therefore, has a potential as a tumor marker (Hilkens et al., 1989). The determination of MUC1 antigen concentration in blood has been exploited as a measure of tumor burden and changing levels are used to monitor the response to therapeutic interventions (Berruti et al., 1994).

Although, the ‘APDTRP’ sequence of MUC1 is immunogenic, the observed cellular immunity was found to be very weak (Karanikas et al., 1997). Until now, the vaccine production by using either natural or synthetic MUC1 antigen did not proved to be successful. Therefore, an antibody-based therapy might be an alternative in the management of MUC1 related cancers.

Despite the availability of a large number of antibodies for passive immunization against the MUC1 antigen, potential problems remains in their use as a therapeutic agent. These include immunogenicity of murine antibodies and poor tumor–penetration as a result of their large size. The former is a particular limitation in therapies where multiple administrations are required. These problems can be circumvented by the use of fully human antibody fragments like scFv’s and diabodies (McCafferty et al., 1990). Single domain antibodies such as $\text{V}_H$ or $\text{V}_L$ domains alone are advantageous due to their even smaller size and their efficient penetration of avascular cancerous tissues (Mayer et al., 1999). Furthermore, they have reduced immunogenicity due to the absence of the Fc region of the full size antibody.

I.1 Antibodies

I.1.1 Overview

All vertebrates have an immune system that provides a protection against invading organisms and any toxic substance produced by them. Because these immune reactions can be detrimental, it is essential that they must be triggered in response to the foreign organisms or molecules that are not host-specific. This ability to distinguish foreign molecules from
self-molecules is a fundamental feature of the immune system. Any substance that is capable of eliciting an immune response is called an immunogen or antigen. The immune system can even distinguish between very similar immunogens such as the two optical isomers of the same molecule. The cell mediated and the humoral antibody mediated immune responses are the two main arms of the immune response. The cellular immune response involves the development and proliferation of specialized cells directed against the foreign antigens. The humoral antibody response involves the production of antibodies; these polypeptides are also called immunoglobulins (Ig). Antibodies circulate in the blood stream, permeate other body fluids and bind specifically to foreign antigens. The portion of the antigen recognized by specific antibodies is called the epitope. Antibodies are synthesized exclusively by plasma cells and produced in different forms with a variety of antigen binding sites and different amino acid sequence. Antibody or immunoglobulin (Ig) molecules are responsible for two major biological functions: (1) a binding function for the recognition of foreign antigens such as toxins, viruses, exposed molecules on the surface of pathogenic organisms and differentiated cells; and (2) an effector function which results in the elimination or inactivation of the foreign antigen or the cell marked by the presence of that antigen (Roitt et al., 1993). This ability to differentiate between antigens has made antibodies an indispensable tool in diagnostic medicine and as therapeutic agents of high potential.

The IgG antibody molecule (150kDa) consists of four polypeptide chains (Figure I-1). Two identical light chains with a mass of around 25kDa and two identical heavy chains with a mass of around 50kDa. In higher vertebrates there are five classes of antibodies namely IgA, IgD, IgE, IgG, and IgM, each with its own class of heavy chain - α, δ, ε, γ, and μ respectively. In addition, there are a number of subclasses of IgG and IgA immunoglobulins; for example, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4) having γ1, γ2, γ3, and γ4 heavy chains respectively. The various heavy chains impart a distinctive conformation to the hinge and tail regions of antibodies and provide each class (and
subclass) the characteristic properties. In addition to the five classes of heavy chains, higher vertebrates have two types of light chains, \( \kappa \) and \( \lambda \), either of which may be associated with any of the heavy chains. The IgG molecule has one variable (V) and constant (C) domain for each light chain and one variable (V) and at least three constant (C) domains for each heavy chain. Disulfide bridges connect the light chain to the heavy chain and one heavy chain to the other heavy chain. Areas of variability and conserved structure are divided into segments of 110 amino acid residues. The heavy-chain V unit shows similarity to the V part of the light chain, while the three heavy chain C region units show strong homology to each other and to the C region of the light chain. Each domain has a characteristic tertiary structure consisting of two \( \beta \)-pleated sheet structures. The sandwich-like structural domain possesses a hydrophobic interior. Sheets are covalently linked by disulfide-bridge near the center of the domain. The N-terminal variable parts of both heavy (\( V_H \)) and light chain (\( V_L \)) together constitute the antigen-binding region. Comparison of amino acid sequences of the \( V_H \) and the \( V_L \) domains revealed the occurrence of three regions of high variability in each chain. These hyper-variable regions are often referred as complementarity determining regions (CDR’s) (Wu et al., 1970). Their amino acid residues define the size and shape of the antigen binding regions. An individual antibody molecule always consists of identical light chains and identical heavy chains; therefore, its two antigen-binding sites are always identical.

A classical method used in the size reduction of antibody molecules is the digestion of immunoglobulins with proteases (Figure I-2). Papain digestion of the antibody produces two Fab fragments (Fragment antigen binding) and one Fc fragment (Fragment crystallizable) (Figure down). Fab fragments (50 kDa) consist of the entire light chain and a part of the heavy chain (\( V_H \) and \( C_H \) portion). Fab fragments contain one antigen-binding site and are therefore monovalent. The Fc fragment (50 kDa) does not bind to the antigen. Pepsin digestion of the antibody produces one large bivalent fragment (100 kDa) called \( F(\text{ab}')_2 \) and degrades the remainder of the heavy chains into small pieces. When the large fragment i.e. F

![Figure I-2 Fragmentation of IgG with papain and pepsin enzyme.](image)

(ab’), is treated with reducing agents, the disulfide bridges are disrupted, giving two Fab-like
fragments, Fab’. Each of the Fab’ fragments is made up of the entire light chain and a
slightly longer part of the heavy chain (Roitt et al., 1993).

Intact antibodies also demonstrate segmental flexibility, which means that the two Fab
portions can move relative to one another on antigen binding. The angle varies from 60 to
180 degrees. This flexible region, where the arms meet the stem of the Y is called the hinge
region and is located between the C_H1 and the C_H2 domains. Only IgG, IgA, and IgD
antibody molecules have hinge regions. The heavy chains of IgM, IgA, and IgD antibody
molecules possess additional amino acid residues on the carboxyl-terminus of the last C_H
domain. These areas, called tailpieces, permit IgM and IgA to interact with like antibodies
and form multimeric molecules. Multimeric IgA also have a polypeptide called the joining
(J) chain, which is disulfide-linked to the tailpieces and stabilizes the multimeric structure.
Antibodies also contain carbohydrates; the percentage and location of the carbohydrate
molecules differ according to the class of antibody.

I.1.2 Antibody engineering

Most natural antigens have multiple epitopes, the exposure of the animal to the antigen
results in the generation of B-cells producing antibodies against different epitopes of the
antigen. A mixture of antibodies that recognize different epitopes on the same antigen are
said to be polyclonal. For many types of studies involving antibodies, monoclonal antibodies
are preferable to polyclonal antibodies. However, the biochemical purification of
monoclonal antibodies from the serum is not feasible, in part because the concentration of
any given antibody is quite low. For this reason, a methodology was developed to
immortalize the plasma cells in order to obtain usable quantities of monoclonal antibody.
Because primary cultures of normal B-lymphocytes do not grow indefinitely, such cultures
have limited use for the large-scale production of the monoclonal antibodies. This limitation
can be avoided by fusing normal B-lymphocytes with immortal transformed lymphocytes
called myeloma cells.

In 1975, Köhler and Milstein developed the ‘hybridoma technology’, which allows
the isolation and production of monoclonal antibodies having specificity for a single antigen
epitope such as a protein, a carbohydrate, and nucleic acids. Besides the importance as
research tools, the availability of murine monoclonal antibodies opened the way to the
development of diagnostics and human therapeutics. Monoclonal antibodies constitute the
most important tools for assay methods such as radiolabelled immunosorbant assay (RIA),
flow cytometry, enzyme-linked immunoabsorbant assay (ELISA) and immunocytopathology.
However, the monoclonal antibodies possess a number of technical limitations and drawbacks such as:

1. expensive production method based on a mammalian expression system;
2. inability to optimize the antibody using genetic engineering techniques;
3. the elicitation of immunogenic reaction against mouse antibodies called human anti-mouse antibody (HAMA) response after repetitive human administration (Kuus-Reichel *et al*., 1994).

Furthermore, the hybridoma technology relies on animal immunization, which are strictly controlled under new regulations. Although the hybridoma technology immortalizes the antibody producing B-cell, the advent of recombinant DNA technology provides a mean to immortalize the antibody gene (Plückthun, A., 1994). The technological advances in the recombinant DNA technology on one hand can help to solve the immunogenicity problem while on the other hand it allows us to tailor the antibodies to a desired size and specificity. Furthermore, it allows us to produce them in different expression systems.

The technique of monoclonal antibody production and recombinant DNA technology were combined to try to resolve the problems that had arisen in the application of murine antibodies for human therapy. Thus the field of antibody engineering was initiated. To circumvent the problems associated with murine monoclonal antibodies, several strategies to bring the antibody format closer to the human one have been started. The humanization process and the requirement of the smaller antibody units in diagnostics and therapeutics lead to the development of functional forms of antibodies in different size format. The different forms (Figure I-3) are as follows:

A. **Chimeric antibodies** (page 6) are made by joining the variable domains of a murine monoclonal antibody to the constant domains of a human antibody thus reducing the immunogenicity and ensuring the efficient recruitment of human effector function (Boullianne *et al*., 1984). Chimeric antibodies have a longer plasma half-life than the murine monoclonal antibodies (Billettera *et al*., 1992). However, the chimeric antibodies can elicit the human anti-chimeric antibody immune response (HACA) as a result of the murine variable domains.

B. The immunogenicity of chimeric antibodies can be reduced by grafting complementarity-determining regions (CDR) of mouse antibodies on human framework region. The resulting antibodies are called as ‘humanized or reshaped antibodies’ (Riechmann *et al*., 1988). The resulting humanized antibodies are approximately 90% human. In this approach, the six CDR regions of the murine
light and heavy chain are grafted on to the human immunoglobulin framework region (Jones et al., 1986). This method takes the advantage of conserving the structure of the antibody variable domains, while using the human framework residues as a scaffold to support the CDR loops which will determine antigen specificity. Although this approach can be successful for some antibodies, it has been shown that the reconstitution of key contacts between loop and framework residues is necessary for other antibodies.

![Figure I-3](image)

**Figure I-3 Different forms of antibodies.** Monoclonal antibodies from hybridomas are of murine origin. Recombinant chimeric antibodies comprise murine variable (V) regions attached to human constant (C) regions, whereas in humanized (or reshaped) antibodies, only the antigen-binding complementarity-determining region (CDR) loops derive from rodent antibodies. Fully human antibodies can be constructed from human V regions (isolated by phage technology) and human C regions. Fv and scFv fragments, and diabodies comprise only V regions and can be cloned from hybridomas or isolated from phage libraries.

C. **Veneered antibodies** are the non-human antibody fragments in which the exposed residues in the framework regions are replaced to match the human residues at the respective positions (Padlan et al., 1991).

D. The antibodies derived from the humans directly are called as **fully human antibodies**.

E. Although natural effector functions are powerful, antibodies can also be engineered so that the antigen-binding sites have dual specificity for a target cell and another molecule like a toxin. Such antibodies are described as ‘**bispecific antibodies**’. In these molecules, each of the two halves is derived from two parental antibodies of different specificities and the molecule therefore carries two different antigen-binding sites.
F. Intact antibodies are large molecules (150 kDa) with limitations for many therapeutic applications; therefore, smaller functional antibody fragments are prepared. The smaller antibody fragments can be generated either classically through protease treatment (I.1.1) or though recombinant DNA technology. The repertoire of smaller antibody fragments is constituted by Fab's (50kDa), variable fragments (Fv, 25kDa, the non-covalently combined variable light chain and heavy chain), single chain variable fragment (scFv, 27kDa, combination of variable light chain and heavy chain joined covalently through a linker), diabodies (50kDa, noncovalent combination of two scFv’s having shorter linker), minibodies (80kDa, scFv joined with C\_1 domain) and variable heavy (V\_H) or variable light (V\_L) chain domains alone.

Making human monoclonal antibodies poses difficulties using the classic hybridoma technology. The use of mouse myeloma as a fusion partner for human cells leads to the preferential loss of human chromosomes, and intolerable instability of the hybrids. As an alternative to fusion, immortalization of human cells by Epstein-Barr virus does lead to lines which are low producers of antibody and unstable (Winter and Milstein, 1991). Further, due to ethical reasons, human can rarely be hyper-immunized to order, especially with noxious chemicals, pathogenic viruses or cancer cells to obtain antigen specific human antibodies. Over the past decade, molecular display technologies enabled us to easily create fully human antibodies. In molecular display techniques, the genes encoding the desired protein are selected for the desired binding function. In 1990, McCarfferty et al., showed that antibody fragments could be displayed on the surface of filamentous phage particles by fusion of the antibody variable genes to one of the phage coat proteins like p8 or p3. Multiple rounds of affinity selection could subsequently enrich antigen-specific phage antibodies, because the phage particle carries the gene encoding the displayed antibody. In the early 1990’s Winter et al., showed for the first time that the phage display technology could be used to select antigen-specific antibodies derived from spleen B-cells of immunized mice. Fully human antibodies can also be obtained through phage display technologies (Clackson et al., 1991), thereby, bypassing the requirement to immortalize antigen-specific B-cells. Similarly, libraries were made from human B-cells taken from individuals immunized with antigen, exposed to infectious agents, suffering from a autoimmune diseases or cancer. Cloning Ig genes from the heterogeneous cell populations has the advantage that all the immune cells can be processed simultaneously for the amplification of the V\_H and V\_L genes but the associated disadvantage is that the original V\_H and V\_L pairing is lost upon pairing for high
affinity. A combinatorial library with only 1000 each of different \( V_H \) and \( V_L \) gene elements would necessitate the screening of \( 10^6 \) to \( 5 \times 10^6 \) clones to recover most of the original pairing. Thus the chances of recovering original pairs of \( V \) genes from a large random combinatorial library from an immunized animal is very low and even more remote for the highest affinity antibodies of hyperimmune animals. Those pairs of \( V_H \) and \( V_L \) domains that are found to be capable of binding antigen are likely to do so at lower affinity or with lower specificity than those selected by antigen (Hudson et al., 1987). Another drawback of this strategy is that the immune libraries have to be built for each antigen and the library can be biased against self-antibodies and above all, the antigen must be immunogenic in nature. The alternative to this approach is the enrichment of phage antibodies from non-immunized healthy individuals called as ‘naïve library’ (Marks, et al., 1991) or by making libraries of synthetic antibodies with variable genes consisting of germline segments artificially diversified by oligonucleotide based cloning (Barbas, et al., 1992). One of the advantages of selecting antibodies from phage libraries is that antigens can be used that are difficult for eliciting an immune response e.g. instable antigens or autoantigens or toxins.

Fully human antibodies can also be obtained by immunizing ‘Xeno-mice’ that carry the selected human immunoglobulin germline gene segments (Duchosal et al., 1992), resulting directly in the formation of human antibody producing B-cells.

### I.1.3 Therapeutic potential of antibodies

Recombinant antibody technology made it relatively easy to select, to humanize and to produce antibodies in bulk and allows the design of antibody-based reagents of any specificity for a wide variety of applications. The therapeutic potential of antibodies can be traced back more than a century. Currently, recombinant antibodies represent over 30% of the biopharmaceuticals in clinical trials. Till now, the US Food and Drug Administration (FDA) has approved 15 antibodies to treat cancer, transplant rejection and to combat autoimmune diseases (Table I-1).

At least 200 other monoclonal antibodies are worldwide in clinical trials including many fully human antibodies (Table I-2). As the development of new therapeutic agents into commercial products takes about 10 years, the FDA approved antibodies in end stage of clinical trails are mainly chimeric or humanized antibodies, since they were developed with early antibody engineering technologies.
### Table I-1 FDA approved therapeutic antibodies (Hudson et al., 2003). *Tradenames are registered trademarks.

<table>
<thead>
<tr>
<th>Trade name*</th>
<th>Target</th>
<th>Product type</th>
<th>Indication</th>
<th>Month of approval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthoclone OKT3</td>
<td>CD3</td>
<td>Mouse Transplant rejection</td>
<td>June 1986</td>
<td></td>
</tr>
<tr>
<td>ReoPro GpIIb/gpIIa</td>
<td>Chimeric Fab</td>
<td>Cardiovascular disease</td>
<td>December 1994</td>
<td></td>
</tr>
<tr>
<td>Rituxan CD20</td>
<td>Chimeric</td>
<td>Non-Hodgkin lymphoma</td>
<td>November 1997</td>
<td></td>
</tr>
<tr>
<td>Zenapax CD25</td>
<td>Humanized</td>
<td>Transplant rejection</td>
<td>December 1997</td>
<td></td>
</tr>
<tr>
<td>Remicade TNF-α</td>
<td>Chimeric</td>
<td>Crohn's disease</td>
<td>August 1998</td>
<td></td>
</tr>
<tr>
<td>Simulect CD25</td>
<td>Chimeric</td>
<td>Transplant rejection</td>
<td>May 1998</td>
<td></td>
</tr>
<tr>
<td>Synagis RSV</td>
<td>Humanized</td>
<td>Respiratory syncytial virus</td>
<td>June 1998</td>
<td></td>
</tr>
<tr>
<td>Herceptin Her-2</td>
<td>Humanized</td>
<td>Metastatic breast cancer</td>
<td>September 1998</td>
<td></td>
</tr>
<tr>
<td>Mylotarg CD33</td>
<td>Humanized</td>
<td>Acute myeloid leukemia</td>
<td>May 2000</td>
<td></td>
</tr>
<tr>
<td>Nabi-HB HBsAg</td>
<td>Human Ig</td>
<td>Acute hepatitis B</td>
<td>October 2000</td>
<td></td>
</tr>
<tr>
<td>Campath CD52</td>
<td>Humanized</td>
<td>Chronic lymphocytic leukemia</td>
<td>May 2001</td>
<td></td>
</tr>
<tr>
<td>DigiFab Digoxin</td>
<td>OvineFab</td>
<td>Digoxin toxicity</td>
<td>August 2001</td>
<td></td>
</tr>
<tr>
<td>Zevalin CD20</td>
<td>Mouse</td>
<td>Non-Hodgkin lymphoma</td>
<td>February 2002</td>
<td></td>
</tr>
<tr>
<td>Humira TNF</td>
<td>Human</td>
<td>Rheumatoid Arthritis</td>
<td>December 2002</td>
<td></td>
</tr>
<tr>
<td>Bexxar CD20</td>
<td>Mouse</td>
<td>Non-Hodgkin's lymphoma</td>
<td>June, 2003</td>
<td></td>
</tr>
</tbody>
</table>

### Table I-2 Therapeutic antibodies undergoing various phases of clinical trials (Brekke et al., 2003). *Trade names are registered trademarks.

<table>
<thead>
<tr>
<th>Tradename*</th>
<th>Molecule type</th>
<th>Phase</th>
<th>Indication(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABX-EGF Human (anti-EGF-R)</td>
<td>Phase II</td>
<td>Non-small-cell lung cancer</td>
<td></td>
</tr>
<tr>
<td>ABX-IL8 Human (anti-IL-8)</td>
<td>Phase II</td>
<td>Chronic obstructive bronchitis</td>
<td></td>
</tr>
<tr>
<td>Eculizumab Humanized (anti-C5)</td>
<td>Phase IIb</td>
<td>Rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>D2E7 Human (anti-TGF-α)</td>
<td>Phase III</td>
<td>Rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>CAT-152 Human (anti-TGF-β2)</td>
<td>Phase II/III</td>
<td>Scarring following Glaucoma</td>
<td></td>
</tr>
<tr>
<td>Antegren Humanized (anti-α-4 integrin)</td>
<td>Phase III</td>
<td>Crohn's disease, Multiple sclerosis</td>
<td></td>
</tr>
<tr>
<td>Avastin Humanized (anti-VEGF)</td>
<td>Phase III</td>
<td>Metastatic breast cancer</td>
<td></td>
</tr>
<tr>
<td>IDEC-151 Primatized (anti-CD4)</td>
<td>Phase II</td>
<td>Rheumatoid arthritis</td>
<td></td>
</tr>
<tr>
<td>MEDI-507 Humanized (anti-CD2 Receptor)</td>
<td>Phase II</td>
<td>Suppresses NK and T-cell function</td>
<td></td>
</tr>
<tr>
<td>XTL001 Human (anti-HBV)</td>
<td>Phase I/II</td>
<td>Hepatitis B virus neutralization</td>
<td></td>
</tr>
<tr>
<td>Hu1D10 Humanized (anti-MHC classII)</td>
<td>Phase II</td>
<td>Non-Hodgkin's Lymphoma</td>
<td></td>
</tr>
</tbody>
</table>


1.1.4  Recombinant small antibody fragments

The smallest portion of the entire antibody that retains affinity towards antigen is constituted by the variable light (V\textsubscript{L}) and heavy chain (V\textsubscript{H}) segment (Bedzyk \textit{et al.}, 1990). These Fv molecules are often unstable due to the non-covalent interactions between the V\textsubscript{H} and V\textsubscript{L} fragments. Only occasionally stable Fv’s are obtained (Givol, 1991). Fv’s can be stabilized by chemical cross-linking of V\textsubscript{H} and V\textsubscript{L} using glutaraldehyde or by introducing disulfide bridges between V\textsubscript{H} and V\textsubscript{L} domains (Glockshuber \textit{et al.}, 1990). Another approach is to link the V\textsubscript{H} and V\textsubscript{L} domain with a flexible peptide linker, allowing the orientation of the CDR’s in a similar fashion as in natural antibodies and stabilizing the recombinant version of the Fv fragment. Such covalently linked Fv fragment is referred to as a single chain Fv (scFv) (Bird \textit{et al.}, 1988). Although the antibody fragments such as Fv and scFv’s can be vitally used but they may require further engineering to eliminate undesirable properties such as dissociation, and aggregation. An attractive alternative to overcome these problems can be isolated single antibody domains such as V\textsubscript{H} or V\textsubscript{L}.

Other observations indicated that some heavy chains could bind antigens in the absence of the light chain (Winter \textit{et al.}, 1991 and Haber \textit{et al.}, 1966). Also V\textsubscript{H} domains with high affinities were detected in expression libraries of mouse V\textsubscript{H} genes (Ward \textit{et al.}, 1989). In contrast to the V\textsubscript{L} domains, such V\textsubscript{H} domains often retain the antigen-specificity of the parental antibody, since their CDR-H3 is often the major contributor to antigen binding. However, removing the V\textsubscript{L} domain from an Fv exposes the large interdomain hydrophobic surface of the V\textsubscript{H} to the solvent (the former interaction site with the V\textsubscript{L}), so that the isolated V\textsubscript{H} molecules becomes generally ‘sticky’ and are, therefore, difficult to produce in a soluble form.

In contrast to this, a considerable fraction (up to 75%) of functional antibodies in \textit{Camelidae} (camels, dromedaries and llamas) is constituted by variable heavy chains (15kDa) only. These antibody fragments lack the light chain (V\textsubscript{L}) and are referred to as heavy-chain antibodies (HCAb’s) (Hamers-Casterman \textit{et al.}, 1993) or variable heavy chain antibody, V\textsubscript{H}H (Renisio \textit{et al.}, 2002). V\textsubscript{H}H’s have also been documented in the spotted ratfish (Cos5-Abs) and nurse shark (NAR) (Nguyen \textit{et al.}, 2002). The characteristic structural features of V\textsubscript{H}H can be mimicked to reduce the hydrophobic surface of V\textsubscript{H} domain at the V\textsubscript{L} domain.

It has been observed that the V\textsubscript{H} domain affinity drops by one to three orders of magnitude compared to the scFv (Borrebaeck \textit{et al.}, 1992). The affinity of V\textsubscript{H} domains can be improved after site-specific secondary randomizations in CDR-H1 and CDR-H2, phage display and antigen selection (Davies \textit{et al.}, 1996).
I.1.5 Importance of antibody fragments

Reducing the size of the antibody molecule to the size of the Fv, or even smaller, would be beneficial in immunoaffinity and biosensor applications (Welling et al., 1990-91). Smaller molecules are also shown to have a better tissue penetration and faster blood clearance in in-vivo applications (George, 1994) and therefore better biodistribution. Such antibody fragments do not contain the Fc region of the antibody molecule, which is responsible for the effector functions of the immunoglobulin (Ig) molecule such as complement fixation, allergic responses and killer T-cell activation. They can be used for a variety of in-vivo applications such as enzyme inhibitors (Riechmann, L., 1999), as molecular building blocks to construct more complex molecules (bispecific or bivalent constructs), for in-vivo tumor targeting (Carter et al., 1997 and Cortez-Retamozo et al., 2002) and for immunoassays.

I.2 The human V_H family

Antibodies are encoded by a limited number of germline gene segments that undergo somatic diversification through rearrangement and mutation. In humans, the immunoglobulin heavy-chain variable locus mapped to band q32.33 of chromosome 14 and constituted by individual variable (V), diversity (D), joining (J) and constant (C_H) region segments linked in separate regions. The V_H genes in the human haploid genome are classified according to sequence homologies of 80% or above into seven families. They differ in size and contain pseudogenes, form an interspersed cluster, exhibit homologies with mouse genes and have phylogenetically developed from a single primordial gene (Lukowsky et al., 1990). The different subgroups are designated as V_H1 to V_H7. The V_H3 subgroup is the dominant subgroup followed by V_H4, V_H1, and V_H5. The remaining subgroups represented only small contributions. The subgroup classification is based on DNA analysis and mRNA analysis

![Figure I-4](image)

Figure I-4 Schematic representation of IgG heavy chain unarranged gene and mRNA. L: Leader segment, V: Variable segment, D: Diversity segment, J: Joining segment, C: Constant segment, C_H1, C_H2 and C_H3 are different constant segments, H: Hinge region.

from B-cells (Cook et al., 1994). The number of V_H gene segments per individual has most recently been estimated as about 76 (24 V_H1 segments, 5 V_H2 segments, 28 V_H3 segments, 14 V_H4 segments, 3 V_H5 segments, 1 V_H6 segments, 1 V_H7 segments) although these figures are likely to be underestimated (Walter et al., 1990). Some human V_H gene segments are
remarkably stable such as human $V_H$ gene segments from the $V_H$1, $V_H$3 and $V_H$4 gene families (Sanz et al., 1989).

The human $V_H$3 family is most homologous to mouse $V_H$3 segments, suggesting that members of the $V_H$3 family may be conserved by some functional constraint. The 5'-flanking region of each family has a family-specific structure (Haino et al., 1994). Amino acid sequence data indicate that human $V_H$3 genes correspond to only a small subset of mouse $V_H$3 genes. Human $V_H$3 genes contain a shorter intron and are two codons shorter than most mouse $V_H$3 genes. Both contain approximately 50% silent substitution (Rechavi et al., 1982). During somatic development, coding segments flanked by characteristic short recombination signal sequences, separated by intervening sequence regions that may exceed 2000kb, are recombined. Combinatorial joining of different segments as well as imprecision in this process contribute to the diversity of the primary antibody response; subsequent mutation further alters functionally rearranged genes (Hinds et al., 1986). Each variable heavy chain gene segment (V) encodes a 5' hydrophobic leader peptide and between 95 and 101 amino acid residues of the mature domain flanked at 3' end by two recombination signals consisting of a highly conserved heptamer (5'-CACAGTG-3'), a 23-base-pair spacer and a less-conserved nonamer (Figure I-4).

### I.2.1 Variable heavy chain domain

#### I.2.1.1 General structure

Variable domains are formed by two beta-sheets, packed face to face, and the inter-strand turns. Comparison of the different known structures shows that they have a core of 76 residues, which has the same main-chain conformation in all structures. This common core contains almost all of the beta-sheet structure and three inter-strand turns. The regions that differ in conformation are the three hyper-variable regions, three other inter-strand turns and a few adjacent residues. Around the deep structure there are buried hydrophobic residues that, in different variable domains, can differ greatly in volume. These differences in volume are accommodated by conformational changes in turn regions that are outside the common core. On the surface they have strongly conserved hydrophilic or neutral residues (Chothia et al., 1998).

#### I.2.1.2 Diversity of the $V_H$ domain

The structural variability of the antibodies is the result of combinatorial genetic diversity. For the $V_H$ domain, the variability generated by the assembly of the variable (V), Diversity (D) and Joining (J) gene segments. Two of the CDR’s (1&2) are encoded by the variable (V)
gene segment while CDR-H3 is formed by the 3’end of the V gene segment, the D gene segment and the 5’end of the J gene segment (refer to I.1.5). With nucleotide addition (N-region diversity at the V-D and D-J joins), the use of different reading frames in the D segment, and the combination of different rearranged heavy and light chains, results in huge diversity of the primary libraries (Tonegawa et al., 1983). During an immune response, the antibody variable regions are further diversified by somatic hypermutation, leading to higher affinity towards the antigen (Berek et al., 1988). The V_{H} fragment has evolved by unequal crossing-over, conversion, duplication and deletion (Wysocki et al., 1989).

I.2.1.3 Conformations of the hypervariable loops

In antibodies, the six antigen-binding loops CDR-H1 to -H3 and CDR-L1 to -L3 can be defined by structural variability (Al-Lazikani et al., 1997) or sequence hypervariability (Wu et al., 1970). Sequence hypervariability can also be used to identify CDR’s for antigen binding by the correlation of loop residues that have a higher relative variability than others with residues that are known from crystal structures of antibody complexes to be important in antigen binding (Padlan et al., 1995). In human and mouse, the first two antigen-binding loops of a V_{H} domain, CDR1 and CDR2, can be assigned to a limited number of possible conformations referred to as canonical structures (Chothia et al., 1987). The conformation of these loops depends both on their length and on the presence of specific residues at key positions. H3 plays the most important role in antigen binding (Padlan et al., 1994). High degree of sequence variations within the hypervariable loops or complementarity-determining regions (CDR) is responsible for the specificity and affinity of the antibodies. Different CDR structures form according to the antigen structure. These CDR structures can create flat, extended binding surfaces for protein antigens, a specific groove for a peptide, DNA or carbohydrate, or deep-binding cavities for small molecules called haptens. It has been pointed out that the distribution of amino acids in variable domains seems to be biased, and certain residues (Tyr, Trp and Asn) seem to have a propensity for being in the CDR’s and for participating in antigen recognition. It seems that the aromatic side-chains (Tyr, Trp) are more exposed to the solvent than in usual water-soluble proteins, and they are frequently found to be involved in the interaction with the antigen. This is explained by their large size (hydrophobic interaction), large polarizability (van der Waals interaction), ability to form hydrogen bonds, and rigidity (reduced loss of conformational entropy upon complexation). Thus, the concentration of aromatic rings would give certain ‘stickiness’ to the CDR’s and give diverse specificity to antibodies. Specificity for a particular antigen therefore arises
from the complementarity of the shapes of the interacting surfaces created by the proper positioning of the aromatic rings and the correct location of polar and/or charged groups.

Analysis of human and mouse immunoglobulins has shown that five of the six hypervariable loops that form the antigen binding site have a small repertoire of main chain conformations called canonical structures. The residue patterns in the V\textsubscript{H} segments imply that the first hyper-variable region (CDR-H1, H26-H32) adopts one of three described canonical structures and that the second hyper-variable regions (CDR-H2, H52-H56) adopt one of the five described canonical structures (Al-Lazikani \textit{et al.}, 1997). The specificity of the loops is produced by (1) sequence differences in their surface residues, particularly at sites near the combining site, and (2) sequence differences in the hypervariable and framework regions that modulate the relative positions of the loops (Chothia \textit{et al.}, 1992). The prediction of the third hypervariable loop (CDR-H3) structure of V\textsubscript{H} has been a bottleneck for a long time, although much progress was made in recent years (Al-Lazikani \textit{et al.}, 1997; Morea \textit{et al.}, 1998 and Shirai \textit{et al.}, 1996). The third hypervariable region of the heavy chain is far more variable in length, sequence and structure than the others, and therefore, it is difficult to classify them into canonical-structures.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Typical bulged (A) and non-bulged (B) CDR-H3 torso region (Morea \textit{et al.}, 1998). The choice of bulged versus non-bulged torso conformation is dictated primarily by the sequence, through the formation of a salt bridge between the side-chains of an Arg or Lys at position 94 and an Asp at position 101. Thus the torso region appears to have a limited repertoire of conformations, as in the canonical structure model of other antigen-binding loops. The heads or apices of the loops have a very wide variety of conformations. In shorter CDR-H3 regions, and in those containing the non-bulged torso conformation, the heads follow the rules relating sequence to structure in short hairpins (Chothia \textit{et al.}, 1998).}
\end{figure}

The CDR-H3 region comprising the residues between Cys-H92 and Gly-H104 can be divided into a torso region (comprising residues proximal to the framework, four residues from the N terminus and six residues from the C terminus), and a head/apex region. Two distinct types of torso regions (Figure 1-5 A&B) have been observed for the CDR-H3 loop having more than
ten residues: (1) the conformation of the torso has a beta-bulge at residue 101, and (2) the torso does not contain a bulge, but continues the regular hydrogen-bonding pattern of the $\beta$-sheet hairpin.

I.2.1.4 $V_H$ domain stability

The folding of immunoglobulin domains almost invariably requires the formation of a conserved structural disulfide bond between Cys22 and Cys92 in the $V_H$ domain; their soluble expression in the reducing cellular cytoplasm has failed in general (Glockshuber et al., 1992). Few exceptions have been reported, notably scFv's against HIV proteins (Maciejewski et al., 1995 and Wu et al., 1996), but low stability, low expression rates, and unpredictable behavior currently limit these perspectives (Gargano & Cattaneo, 1997). Elegant screening methods and random mutagenesis have been employed in few cases, but it is not clear how these approaches can be generalized, because they have either depended on a specific immunoglobulin framework, naturally lacking a disulfide bridge, or a metabolically selectable ligand binding activity. The following mutations Lys03Glu, Pro07Ser, Ala16Gly, Arg43Glu, Ile58Thr, Lys66Arg, Pro74Ser and X52Ser were reported to have stabilizing effect on the cytoplasmic soluble expression of the $V_H$ fragment (Wirtz et al., 1999, Wörne, 1998 and Proba et al., 1998). X refers to any amino acid.

I.2.2 Comparison between human $V_H$ and camels $V_{HH}$ domains

In *Camelidae*, the major portion (up to 75%) of the functional antibodies is constituted by the variable “heavy chain antibodies” ($V_{HH}$). The heavy chain antibodies are homodimers, with each chain consisting of an unpaired variable domain ($V_{HH}$) immediately followed by a hinge region, $C_{H2}$ and $C_{H3}$ domain. The cDNA sequence analysis illustrated that the equivalent of the conventional IgG’s $C_{H1}$ domain is clearly missing. Nevertheless, $V_{HH}$ domains possess some remarkable properties, which clearly distinguish them from the variable domain ($V_H$) of conventional antibodies.

I.2.2.1 Sequence differences

The comparison between the human $V_H3$ gene family and camelid $V_{HH}$ gene family shows a high degree of sequence homology. The sequences were obtained from cDNA libraries and RT-PCR clones. It is unlikely that dromedaries possess other $V_{HH}$ gene families, as a representative database of germline $V_{HH}$ sequences revealed the presence of some 40 different $V_{HH}$ genes that all evolved within the $V_H$ subgroup 3 (Nguyen et al., 2000). However, it was noted that the $V_{HH}$’s carry a number of remarkable amino acid substitutions
in the framework-2 region (FR2). In conventional antibodies, the \( V_H \) residues in FR2 region normally interact with the \( V_L \) domain and are well conserved throughout the evolution (Kabat et al., 1991). The amino acid substitutions Val37Phe or Val37Tyr, Gly44Gln, Leu45Arg and Trp47Gly (Figure I-6) in \( V_H \) render the \( V_H/V_L \) interface more hydrophilic and an overall enhanced solubility in comparison to \( V_H \) fragment is observed (Muyldermans et al., 1994 and Vu et al., 1997). Another noteworthy substitution is the Leu11Ser. This amino acid normally interacts with the \( C_H1 \) domain, a domain missing in the camel heavy chain immuno-globulins (V\(_H\)H).

### I.2.2.2 Structural differences

The crystallographic structure of four \( V_H \) (Desmyter et al., 1996; Decanniere et al., 1999 and Spinelli et al., 1996) and the NMR structure of the camelized human \( V_H \) (Riechmann, 1996) have been reported. All these structures confirm that the identical scaffold architecture of the \( V_H \) and the \( V_H \) domain i.e. consisting of two anti-parallel beta-sheets organized in a beta-barrel (Renisio et al., 2002). The root mean square deviation of the framework-C\( \alpha \) atoms between a \( V_H \) and a human \( V_H \) domain is approximately 0.75Å (Decanniere et al., 2000). However, due to the presence of amino acid substitutions in the FR2 region of \( V_H \), the surface formed by the FR2 residues is rendered much more hydrophilic.

![Figure I-6 Illustration of the differences between human \( V_H \) and camelid \( V_H \) based on the sequence comparison (Mulyysdermans & Lauwereys, 1999).](image)

Three hypervariable regions can be clearly distinguished in the \( V_H \) sequences (Figure I-6). The CDR-H1 and CDR-H2 loops adopt new structures that fall outside the canonical structures described for human or mouse \( V_H \) loops (Decanniere et al., 2000). Furthermore, CDR-H1 loop of \( V_H \) fragments may begin closer to the N-terminus in comparison to human or mouse CDR's (Harmsen et al., 2000). The high incidence of amino acid insertions or deletions in, or adjacent to the first and second antigen-binding loops of \( V_H \) will undoubtedly diversify the possible antigen-binding loop conformations (Nguyen et al.,
The CDR-H3 loop of V\textsubscript{H}H fragments is on average longer (17 residues) than the human (12 residues) or mouse (9 residues) CDR-H3 loop. To stabilize the long CDR-H3 loop, the V\textsubscript{H}H domains often contain a second intra-domain disulfide bond, which connects the CDR-H3 with the end of the CDR-H1 or with a core residue between CDR-H1 and CDR-H2 loop (Muyldermans et al., 1994). The CDR-H3 loop in V\textsubscript{H}H lacks the salt bridge between Arg94 and Asp101. The V\textsubscript{H}H with known structure have a tendency to turn their CDR-H3 loop outwards from the beta-barrel, towards the region that would interact with the V\textsubscript{L} in an Fv structure. Therefore, the location of the long CDR-H3 within the V\textsubscript{H}H is incompatible with the association of a V\textsubscript{L}. In summary, the structural analyses provide evidence that the antigen-binding loops of V\textsubscript{H}H exhibit a much larger structural repertoire than the one observed for the V\textsubscript{H}’s. It is possible that the absence of the V\textsubscript{L} domain generates the additional freedom that makes these changes possible and necessary.

I.2.3 Camelization of the V\textsubscript{H} domain

Single domain antibodies are an attractive alternative to scFv’s because of their much smaller size and their comparable affinities (Spinelli et al., 1996). However, low levels of soluble expression and solubility problems have hampered the development of such molecules (Muyldermans et al., 2001). The discovery of soluble and functional camelid heavy chain antibodies ‘V\textsubscript{H}H’ (Hamers-Casterman et al., 1993) opened up new opportunities for the development of single domain antibodies. The V\textsubscript{H}H possess distinct sequential and structural features responsible for their greater solubility in comparison to conventional V\textsubscript{H} domains (refer to I.1.7.1 & I.1.7.2). The process of incorporating features of V\textsubscript{H}H domains into human V\textsubscript{H} frameworks is called camelization of V\textsubscript{H} domains. In 1994, Davies & Riechmann observed that isolated human V\textsubscript{H} domains undergo aggregation at protein concentrations above 1 mg/ml. The aggregation behavior of human V\textsubscript{H}’s after introducing the mutations G44E, L45R, W47G (to mimic the theoretical V\textsubscript{H}/V\textsubscript{L} interface of V\textsubscript{H}H), was significantly reduced. It might therefore be possible to create a human single V\textsubscript{H} domain that behaves well in solution by mutating a very limited number of residues involved in the ‘former’ V\textsubscript{L} interaction to mimic those residues present in V\textsubscript{H}H (Davies et al., 1996). Camelization of human V\textsubscript{H}’s is a promising technology for the generation of small antigen-binding fragments that should be useful for therapeutic purposes in humans.
I.3 Mucin
I.3.1 Overview

Mucins (Greek: slimy) are secreted or transmembrane heterogeneous glycoproteins of high molecular weight (>200kDa) that apically protects the epithelia from infection, dehydration, physical or chemical injury and are also involved in cell-cell interactions, signaling, and metastasis (Parry et al., 2001 and Hilgers et al., 1989). In addition, mucins can inhibit or block the cytotoxic activity of neutrophils, natural killer (NK) cells, and cytotoxic T cells (Hayes et al., 1990). Mucins also have antiadhesive properties (Lightenberg et al., 1992).

The human mucin family is constituted by six membrane-bound or tethered mucins (MUC1, MUC3, MUC4, MUC11, MUC12, MUC13) and five secreted (MUC2, MUC5AC, MUC5B MUC7, MUC9) mucins (Perez-Villar et al., 1999). Genes encoding these mucins are found on different chromosomes except MUC2, MUC5 and MUC6, which are linked on chromosome 11. The mucin genes are highly polymorphic (Swallow et al., 1987). The structure of the cDNA’s and the protein core of different mucins is similar in exhibiting regions encoding N-terminal protein sequences followed by sequences containing a variable number of tandem repeats (VNTR). Following this is the transmembrane region and the cytoplasmic tail (Figure general MUC1 structure). Variation of this general pattern, caused by exon splicing, leads to the secreted and truncated forms of mucins (Apostolopoulos et al., 1994).

The tandem repeat region of each mucin consists of a different number and composition of amino acids: 20 amino acids (MUC1), 26 amino acids (MUC2), 17 amino acids (MUC3), 16 amino acids (MUC4), 8 amino acids (MUC5), 169 amino acids (MUC6), and 23 amino acids (MUC7). In addition, the number of repeats gives many different alleles; for example, for MUC1 there can be between 40 to 80 repeats in different individuals (Apostolopoulos et al., 1994). Therefore, the repeat units of the different mucins show no similarity to each other either in respect to their sequence or the number of amino acids (Table I-3).

<table>
<thead>
<tr>
<th>Mucin</th>
<th>Amino acid residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUC 1</td>
<td>SAPDTRPAPGSTAPPAHGVT</td>
</tr>
<tr>
<td>MUC 2</td>
<td>PTTTPISTTTMTPTPTPTPTGQT</td>
</tr>
<tr>
<td>MUC 3</td>
<td>HSTPSFTSSSITTTETTS</td>
</tr>
<tr>
<td>MUC 4</td>
<td>TSSASTGHATPLPVTD</td>
</tr>
<tr>
<td>MUC 5</td>
<td>PTTSTTSA</td>
</tr>
<tr>
<td>MUC 6</td>
<td>169 amino acid repeat unit</td>
</tr>
<tr>
<td>MUC 7</td>
<td>TTAAPPTPSATTPAPPSSSAPPE</td>
</tr>
</tbody>
</table>
The VNTR and the whole protein are rich in O-linked glycosylation sites (have a high content of serine, threonine, and proline residues). The 50% of the mucins molecular weight derives from the O-linked carbohydrate residues (Strouss et al., 1992). The average number of O-glycans added to each tandem repeat is around 2.5 (Muller et al., 1997). There is a great variety in the structure of the carbohydrate side chains, which may be composed of only one type of disaccharide or may be complex containing several sugars and branched chains.

### I.3.2 Human mucin-1

Mucin-1 (MUC1) is also known, among many other denominations, as human milk fat globule (HMFG) antigen, DF3 antigen, epithelial membrane antigen (EMA), H23, PAS-0 (positive staining for carbohydrate with periodic and acid-schiff’s reagents), epitectin, polymorphic epithelial mucin (PEM), peanut-reactive urinary mucin (PUM) and episialin (epithelial glycoprotein of high sialic acid content). MUC1 mucin is membrane bound and is not a secretory or gel-forming mucin (Van Klinken et al., 1995).

The MUC1 gene (Ligtenberg et al., 1990) is located on chromosome 1q21-24 (Swallow et al., 1987). The human MUC1 gene is unusually compact, spanning approximately 4 to 7kb of genomic DNA and contains seven exons, the second of which contains the entire tandem repeat sequence (Lancaster et al., 1990) (Figure I-7). The genomic sequence is unusual in composition. The overall genomic sequence is 65% G+C rich, whereas, the tandem repeat region is 82% G+C, which is in marked contrast to the 40% G+C composition usually observed in mammalian genomic DNA.

The number of repeats varies from 20 to 125 per allele and results in an extensive genetic polymorphism, which is observed in the DNA, RNA and protein products. The regulation of MUC1 expression and the genetic mechanisms responsible for its increased expression in tumors is largely unknown.

The cDNA encodes a protein with a very large extracellular domain (1000 to 2200 amino acids) made up largely of 20 amino acid tandem repeats (TR’s). The extracellular domain was found to made up of 20-125 tandem repeats of 20 amino acids with the highly
conserved sequence ‘HGVTSAPDTRPAPGSTAPP’. The high degree of polymorphism observed in the protein was based on differing numbers of tandem repeats present in the core protein (Gendler et al., 1987). Although many of the repeats were identical, variant sequences have been observed in cancer cell lines (MCF-7, T47D and HPAF) and in normal human mammary gland cells from milk (Hanisch et al., 2000). Signal peptide, a hydrophobic transmembrane domain and a phosphorylated cytoplasmic tail of 72 amino acids, containing signalling motifs (Figure I-8).

MUC1 exists in alternatively spliced forms. MUC1/SEC is a secreted form and lacks the hydrophobic transmembrane and cytoplasmic domains required for anchorage to the cell membrane and the MUC1/Y lacks the tandem repeat domains (Zrihan-Licht et al., 1994). MUC1/Y is preferentially expressed in breast carcinomas. An additional alternative splice form lacking the TR domain called MUC1/X or MUC1/Z has been described. The significance of different splice forms is unknown. Different types of epithelia in different

Figure I-9 The single tandem repeat peptide showing in hatched circles the residues to which O-linked glycans can become attached and in gray circles the epitopic sequence against which the majority of murine MUC1 monoclonal antibodies are directed. The circles enclose, in numerical order of frequency, the epitopic sequences against which natural human MUC1 antibodies are directed (von Mensdorff-Pouilly et al., 2000).

Figure I-8 General MUC1 structure.
physiologic states heterogeneously express MUC1. Glycosylation levels vary from about 50% in the normal mammary gland to 80% in colon carcinoma cells or pancreas, resulting in MUC1 proteins ranging in size from approximately 500 to >1000 kDa (Baeckstrom et al., 1991). With in each tandem repeat, two serines and three threonines represent five potential O-glycosylation sites (Figure I-9). The carbohydrate chain may be comprised of up to 16 or more monosaccharide units. At the termini, these may be fucosylated or sialylated (Hanisch et al., 1989).

The determination of hydrophilicity profiles and secondary structures of the MUC1 repeated VNTR sequence indicate that the protein core is dominated by a hydrophilic region extending through residues ‘PDTRPAP’ and bound by two S and T potential glycosylation site (Chou et al., 1978). Initially, the MUC1 mucin tandem repeat protein core was believed to exist in random-coil conformation, attaining the secondary structure solely by the addition of carbohydrates to serine and threonine residues. In 1993, Fontenot et al. showed that the MUC1 protein core is not in a random-coil secondary structure but adopts a polyproline β-turn helix conformation extending 200 to 500 nm above the plasma membrane (Hilkens et al., 1992). The immunodominant region of MUC1 is the ‘PDTRP’ sequence, which forms the tip of a protruding “knob” exposed to solvents forming a stable type II β-turn. Extended spacer regions connect the knobs whereas only multiple tandem repeats can enhance their stability and presentation (Fontenot et al., 1995). The crystal structure of the SM3 Fab antibody fragment in complex with the peptide epitope ‘PDTRP’ has shown that the MUC1 peptide is bound in an extended conformation in an elongated groove in the antibody-combining site, with no discernible secondary structure (Dokurno et al., 1998). Therefore, the correct MUC1 structure determination needs further structural studies of the native protein.

**I.3.3 MUC1 and cancer**

In cancer, MUC1 is overexpressed on cells and has increased circulation levels (Zotter et al., 1988). Its distribution is no longer restricted to the apical surfaces of ducts and glands but is rather found throughout the tumor mass and on the entire surface of tumor cells. There is a correlation between MUC1 overexpression and the high metastatic potential of the cancer as well as the outcome (Takao et al., 1999). In fact, statistics show that cancers expressing MUC1 accounted for about 72% of new cases and for 66% of death rate (Greenlee et al., 2000). Therefore, it can be concluded that the most immunogenic molecule of carcinomas appears to be the MUC1 mucin (Burchell et al., 1989).
Tumour-associated MUC1 has aberrated carbohydrate side chains (Figure I-10) (Hull et al., 1989) resulting in unmasking of the peptide core to allow binding of the antibodies (such as SM3 and DF-3P), which have been valuable in detecting the aberrant presentation of the MUC1 protein in carcinomas (Burchell et al., 1987). On one hand, the overexpression of MUC1 in cancer cells interferes with cell adhesion, favouring metastasis (Hilkens et al., 1995) and also shield the tumour cells from immune recognition; on the other hand, MUC1, in essence a self-antigen, is displaced and altered in malignancy and induces immune responses. In carcinomas, mucin molecules shed into the circulation (Hilkens et al., 1985). The variations in the levels of circulating MUC1 are used to monitor therapy and allow an early detection of reemergence of the tumors (Graves et al., 1998).

The exposed protein core of MUC1, resulting of aberrant glycosylation, offers a unique opportunity to act as an immunotherapeutic target for the specific tumors (Barratt-Boyes et al., 1996).

### 1.3.4 Antibodies against MUC1

Immunization with MUC1 peptides or fusion proteins generally induces an antibody response rather than a CTL response in mice and humans (Karanikas et al., 1997). Natural antibodies against MUC1 are found in patients with malignant tumors. Circulating MUC1 IgM antibodies are present in breast, colon and pancreatic cancer patients (Kotera et al., 1994) and in women with benign and malignant ovarian tumors (Richards et al., 1998). MUC1 IgG antibodies have been detected in colorectal cancer patients (Nakamura et al., 1998). A recently described assay for free IgG and IgM antibodies directed against the MUC1 tandem repeat showed that high levels of free antibodies might inhibit natural metastasis (von
Mensdorff-Pouilly et al., 1998). It seems that humoral immunity is the only possible way to treat, reduce or eradicate MUC1 induced malignancies.

Numerous mucin specific antibodies have been derived following immunization of animals with normal or malignant epithelial cells. The majority of the antibodies recognize the tumour associated MUC1 molecule. The monoclonal antibodies against MUC1 have been available since the early 1980s and are used in clinic for diagnostic purposes e.g. HMFG1 antibody used mainly for ovarian cancer (Maraveyas et al., 1994). Recent epitope mapping has established that many of these monoclonal antibodies react with the tandem repeat core, in particular with epitopes contained within the amino-acid sequence APDTRPA (Burchell et al. 1989).

In particular, large number of murine monoclonal antibodies (~58) against MUC1 have been raised and studied (Price et al., 1998). The repetitive use of these monoclonal antibodies in cancer therapy initiates a human anti-mouse antibody (HAMA) response in humans. Therefore, there is an urgent need either to humanize or to generate human antibodies against MUC1. Human antibodies of low affinity have been detected in serum after immunizing the patients with intact or fragmented MUC1 antigens. Unfortunately, very few recombinant human antibodies specific for the MUC1 have been developed (Andersson et al., 1999; Henderikx et al., 1998; Söderlind et al., 2000; Wong et al., 2001 and Jirholt et al., 2002) and several of these have been shown to be of low affinity and to display a rapid dissociation rate constant, probably making them less suitable for tumor therapy or diagnosis. Furthermore, no attempt has been made to reduce the size of the antibodies needed to achieve a better drug distribution. The use of small recognition units in cancer therapy demands for their improved affinity towards the antigen and their increased stability. These requirements can be met by rational protein design based on the knowledge of the molecular basis of the antibody and its interaction with the antigen.

I.4 Research objectives

Numerous murine monoclonal antibodies have been generated against the MUC1 antigen, but the repetitive dosage of these antibodies can elicits the human anti-murine antibody (HAMA) response in humans. Humanized or human antibodies provide an effective alternative. The fully human scFvM12 antibody has reactivity against adenocarcinoma-associated mucin MUC1 (Wong et al. 2001). The scFvM12 antibody fragment recognizes the hydrophilic epitope ‘PDTRP’ of the MUC1 core protein, considered as an immunodominant region.
Over the last decades, large numbers of peptide-antibody Fab structures have been reported (Wilson & Stanfileld, 1994). A number of uncomplexed tumor specific antibodies against the same antigen have been analyzed (Banfield et al., 1996). But only few structures of either murine or humanized Fab antibody fragments directed against MUC1 have been solved (Dokurno et al., 1998; Banfield et al., 1997). The specificity of the scFvM12 antibody towards the MUC1 antigen makes it a potential and useful tool in the diagnosis and treatment of MUC1 associated adenocarcinomas. It is therefore of significant interest to understand at the molecular level, the specificity of the mucin/antibody interaction, which might allow higher affinity antibodies to be obtained through the use of rational antibody engineering (Rees et al., 1994). The structure would be useful in the generation of antibody with improved functionality and physico-chemical properties such as improved stability, reduced immunogenicity, better tissue distribution and faster blood clearance when used in vivo.

Single domain antibodies are more efficient in solid tumor penetration (Mayer et al., 1999). Although, single domain antibodies have faster clearance rate, simultaneously in comparison to Fab or scFv’s, they achieve better biodistribution in avascular tissues. The crystallisation and structure determination of isolated V_H and V_L domains of the scFvM12 antibody would provide the basis for developing stable small antigen recognizing units. In scFvM12, the V_H and the V_L domains are derived from two different human libraries (Sheets et al., 1998). The structural study of isolated domains and the conformational changes, which may occur upon their association in the scFv format, can be used to derive the rational of the V_H/V_L interaction occurring between two unrelated partners.

In 1996, Riechmann reported that the camelization of the V_H/V_L interface of the V_H domain could be helpful in increasing its solubility. Therefore, the final objective of this thesis included the camelization of the V_H domain of the scFvM12 antibody fragment at positions H44, H45 and H47.

The goals of this thesis can be summarized as follows:

1. Optimisation of expression in *Escherichia coli* and purification of single chain antibody fragment M12 (scFvM12).
2. Crystallization of scFvM12 alone as well as in complex with MUC1 antigen.
4. Generation of the single antibody domain(s) of scFvM12 through limited proteolysis and their crystallization alone as well as in complex with MUC1 antigen.
5. X-ray analysis of the crystal(s) and structure determination for the corresponding dataset. Furthermore, the deposition of the determined structure to the Research Collaboratory for Biological Structures/Protein Data Base (RCSB/PDB).

6. Protein engineering of the M12-V$_{H}$ domain based on the crystal structure.
II MATERIALS AND METHODS

II.1 Materials

II.1.1 Chemicals

Biochemical grade chemicals were purchased from Fluka (Taufkirchen, Germany), ICN (Eschwege, Germany), Sigma (Deisenhofen, Germany) or Pierce (Rockford, IL, USA) and Roth (Karlsruhe, Germany).

II.1.2 Laboratory consumables

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCA protein assay kit</td>
<td>Pierce (Rockford, USA)</td>
</tr>
<tr>
<td>Bradford protein assay kit</td>
<td>Bio-Rad Laboratories Inc., USA</td>
</tr>
<tr>
<td>Dialysis tubing (MWCO- 3.5kDa &amp; 10kDa)</td>
<td>Roth, Kleinfeld, Hannover, Germany</td>
</tr>
<tr>
<td>Falcon tubes (15ml &amp; 50ml)</td>
<td>Becton Dickinson, Heidelberg, Germany</td>
</tr>
<tr>
<td>Hybond™ C nitrocellulose membrane (0.45µm)</td>
<td>Amersham Life Science, Sweden</td>
</tr>
<tr>
<td>Plasmid isolation Kit</td>
<td>Qiagen, Germany</td>
</tr>
<tr>
<td>Parafilm</td>
<td>American National, Canada</td>
</tr>
<tr>
<td>Pipette tips (10µl, 200µl, 1000µl)</td>
<td>Gilson, Villiers-le Bel, France</td>
</tr>
<tr>
<td>Pipettes (2.0µl, 10µl, 20µl, 200µl, 1000µl)</td>
<td>Gilson, Germany</td>
</tr>
<tr>
<td>Plasmid isolation kit (Mini)</td>
<td>Qiagen (Hilden)</td>
</tr>
<tr>
<td>Plastic disposable pipettes (1ml, 5ml, 10ml)</td>
<td>Falcon, Germany</td>
</tr>
<tr>
<td>Sterile filters (0.22µm, 0.45µm)</td>
<td>Millipore, Molsheim, Germany</td>
</tr>
<tr>
<td>Syringes (1ml, 2ml, 10ml, 20ml)</td>
<td>Braun, Melsungen, Germany</td>
</tr>
<tr>
<td>Whatman paper (No.1)</td>
<td>Whatman (Maidstone, UK)</td>
</tr>
</tbody>
</table>

II.1.3 Laboratory equipment

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoclave</td>
<td>Bachofer, Reutlingen, Germany</td>
</tr>
<tr>
<td>Bio-bench in-situ bioreactor</td>
<td>Applikon dependable Instruments, Schiedam, Holland</td>
</tr>
<tr>
<td>Centrifuge</td>
<td></td>
</tr>
<tr>
<td>Avanti™ 30 and Avanti™ J-25</td>
<td>Beckman, California, USA</td>
</tr>
<tr>
<td>Biofuge A</td>
<td>Heraeus, Hanau</td>
</tr>
<tr>
<td>Sigma 3-10 and Sigma 4-10</td>
<td>Sigma, St. Louis, Missouri, USA</td>
</tr>
</tbody>
</table>
II-MATERIALS AND METHODS

RC5C and RC5B plus | Sorvall instruments, Du Pont, Bad Hamburg
Chambers for SDS-PAGE and western blotting | Bio-Rad Laboratories GmbH
Gel air dryer | Bio-Rad Laboratories Inc., USA
Gel electrophoresis equipment (Mini PROTEAN II™) | Bio-Rad Laboratories GmbH
Innova™ 4340 incubator shaker | New Brunswick Scientific, Nürtingen, Germany
Magnetic stirrer | Heidolph M2000 Bachofer, Germany
MAR-Research image plate, mar345 | MARresearch, Hamburg, Germany
Mass spectrometer | Micromass UK Limited, Manchester, UK
Octane workstation | Silicon Graphics Incorporation, USA
pH-meter | WTW, Germany
Quartz cuvettes | QS Hellma, Germany
Rotors
0650, F2402H, JLA 10.500 and JA 25.50 | Beckman, California, USA
#1140 and #11222 | Sigma, USA
RLA-300, SS-34 and GS-3 | Du Point, Hamburg
Single beam spectrophotometer | Eppendorf, Germany
Sonicator | Braun Biotech, Melsungen, Germany
X-ray generator RU2000 | Rigaku, Tokyo, Japan

II.1.4 Chromatography columns

Pre-packed columns (Superdex 75 HiLoad 16/60, MonoQ HR 10/10, MonoS HR 10/10) and resins used were from Amersham-Pharmacia Biotech (Freiburg), self-pack column from Millipore (Eschborn) and Bio-Processing (Consett, Co. Durham, England). For protein purification, the Äkta-FPLC (Amersham Phramcia Biotech) set up was used.

II.1.5 Media, stock solutions and buffers

Standard media and buffers were prepared according to standard procedures (Ausubel et al., 1998; Coligan et al., 1998; Sambrook et al., 1996) using deionized water (R-18.3Ωcm⁻¹). Heat-sensitive components such as antibiotics were prepared as stock solutions, filter-sterilized (0.2µm) and added to the medium/buffer after cooling to room temperature. Compositions of non-standard solutions or buffers are listed at the end of the respective
II-MATERIALS AND METHODS

method section. Media for cultivating bacteria were sterilized by autoclaving (121°C/1-2 bar/20min).

The following bacterial growth media were used:

I. **Small scale (shake flask):**

   **Luria-Bertini (LB) medium**
   - Bacto-peptone 1.00% (w/v)
   - Yeast extract 0.50% (w/v)
   - NaCl 1.00% (w/v)

   The LB agar plate was prepared with LB medium containing 100µg/ml of ampicillin.

   **2xTY medium**
   - Bacto-peptone 1.60 % (w/v)
   - Yeast extract 1.00 % (w/v)
   - NaCl 0.50 % (w/v)

   **SOC medium**
   - Bacto-peptone 2.00% (w/v)
   - Yeast extract 0.50% (w/v)
   - NaCl 0.05% (w/v)
   - Glucose 55.55 mM
   - MgCl₂ 10.00 mM
   - MgSO₄ 10.00 mM

II. **Large scale (fermenter):** The composition of the minimal medium used for high cell density fermentation (II.2.2.2) is as follows:

   - KH₂PO₄ 1.660 % (w/v)
   - NH₄PO₄ 0.400 % (w/v)
   - CaCl₂.2H₂O 0.007 % (w/v)
   - FeSO₄.7H₂O 0.014 % (w/v)
   - MgSO₄.7H₂O 0.150 % (w/v)
   - Glucose 2.500 % (w/v)
   - Citric acid 0.210 % (w/v)
   - L-arginin-HCl 0.020 % (w/v)
   - Trace metal solution* 0.400 % (v/v)
   - Ampicillin 100.00 mg

*The composition of trace metal solution (PTM₁, Invitrogen) is as follows:

   - Biotin 0.82 mM
II-MATERIALS AND METHODS

H$_3$BO$_3$  0.32 mM  
CoCl$_2$·6H$_2$O  2.10 mM  
CuSO$_4$  24.00 mM  
FeSO$_4$  230.00 mM  
NaI  0.53 mM  
MnSO$_4$  19.87 mM  
Na$_2$MoO$_4$·2H$_2$O  0.83 mM  
ZnCl$_2$  0.15 mM

II.1.6 Antibodies, antisera and peptides

Mouse anti-c-myc tag monoclonal antibody (9E10) (ATCC clone number CRL-1729) was used for detection of scFv-fragments by immunoblot (Figure III-4B, III-6B etc). Mouse anti-His$_6$ (Qiagen, Germany) antibodies were also used for detection of scFv fragments. Alkaline phosphatase (AP) or horseradish peroxidase (HRP)-conjugated to goat anti-mouse IgG (H+L, Fc) (Dianova) antibodies were used as secondary antibody in immunoblot analysis (Figure III-4B). NBT/BCIP (Bio-Rad Laboratories Inc., USA) were used as substrate for detection of immobilized proteins in immunoblot (Figure III-4B).

The C-terminal biotinylated MUC1 peptide APDTRAPGSTAPPAHGVTSK-ε-biotin was purchased from Anaspec, San Jose, USA (II.2.9.1.3 & II.2.9.1.4). The peptide was dissolved in 20mM sodium acetate buffer, pH 5.0 and stored at –20°C.

II.1.7 Bacterial strains

E. coli strain DH5α was used as a host cells for subcloning. BL21 and TG1 strains were used for expression of the soluble periplasmic scFv-fragments and V$_H$ domains.

Table II-1 Names, suppliers and genotypes of E. coli strains used.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Source</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>Stratagene</td>
<td>F-,Φ80dlacZΔM15, endA1, recA1, hsdR17 (r_k ,m_i’), supE44, thi-1, gyrA96, relA1, ∆ (lacZYA-argF)U169,λ’</td>
</tr>
<tr>
<td>TG1</td>
<td>Stratagene</td>
<td>F’ traD36 lacFΔ (lacZ)M15 proA’B’/supE ∆ (hsdM-mcrB)5 (r_k,m_i’McrB) thi ∆ (lac-proAB)</td>
</tr>
<tr>
<td>BL21 (DE3)</td>
<td>Novagen</td>
<td>B F– dcm ompT hsdS(r_f– m_i–) gal λ (DE3)</td>
</tr>
<tr>
<td>BL21- Codon plus (DE3)-RIL</td>
<td>Stratagene</td>
<td>B F– ompT hsdS(r_f– m_i–) dcm+ Tet’ gal λ (DE3) endA Hte [argU ileY leuW Cam’]</td>
</tr>
</tbody>
</table>

II.1.8 Software

I. General - Windows NT 4.0 operating system (Microsoft Corporation, USA), Adobe Photoshop 6.0 (Adobe Systems GmbH, Germany); Origin 6.0 (Data analysis and
technical graphics, Microcal Software, Inc.), Bioexpert NT 3.0 beta (Applikon, Schiedam, Holland).

II. Crystallography – DENZO/SCALEPACK (Otwinowski & Minor, 1997); CCP4 program suite (version 4.2.2, Collaborative Computational Project, Daresbury, UK, 1994); PHASES (William Furey, University of California, San Francisco, USA); O (Jones et al., 1991); CNS & X-PLOR (Cooke et al., 1991), Insight II (Biosym Technologies/MSI, San Diego, CA, USA) and PROCHECK (Laskowski et al., 1993).

II.2 Methods

All experiments related to the genetic engineering were performed according to the Regulations of “S1-Richtlinien” and were officially approved by the “Regierungspräsidium des Landes NRW” (RP-Nr.: 23.203.2 AC 12, 21/95) and “BGA” [AZ 521-K-1-8/98:AI3-04/1/0866/88 (S1)].

II.2.1 Recombinant DNA techniques

General DNA recombinant techniques like restriction enzyme analysis, DNA ligation and DNA agarose gel electrophoresis were performed according to the standard protocols described in Sambrook et al. (1989) and Ausubel et al. (1995) or according to the manufacturers instructions when using kits. The recombinant DNA work was supported by Melanie König (RWTH Aachen, Institute for Molecular Biotechnology, Aachen, Germany).

II.2.1.1 Analytical gel electrophoresis

Undigested plasmid DNA, restriction enzyme fragments and PCR products were electrophoretically separated on 1.2% (w/v) agarose gel prepared in TBE buffer containing 0.1µg/ml ethidium bromide. The gel was run at 120V for approximately 30 min in minigel chambers (Bio-Rad). Known amount of DNA molecular marker such as 1kb ladder and PstI digested lambda marker were used to evaluate the size, integrity and to determine DNA concentration. The DNA was visualized on an UV transilluminator at 302 nm and documented using a black and white E.A.S.Y.429K camera (Herolab).

II.2.1.2 Preparative gel electrophoresis

PCR amplified DNA segments or restriction enzyme digested DNA fragments were isolated through preparative gel electrophoresis prior to cloning. After electrophoresis, DNA fragments were excised from the gel using an UV transilluminator. The ‘Qiaquick Gel extraction kit’ (Qiagen) was used to isolate DNA from excised gel pieces according to the
manufacturers instructions. The concentration of the isolated DNA was determined either spectrophotometrically or by agarose gel electrophoresis.

II.2.1.3 PCR amplification of DNA segments

Polymerase chain reaction (PCR) was used for the amplification of DNA segments of interest (Saiki et al., 1988). The reaction was performed in 0.2ml PCR reaction tubes (Biozyme Diagnostik GmbH, Oldendorf, Germany) using a DNA thermal cycler (MWG Biotech). The optimal annealing temperature ($T_A$) was determined either by empirical optimization or by the following formula (Wu et al., 1991):

$$T_A = [22 + 1.46 \{2x(G+C) + (A+T)\}]$$

PCR amplified products were resolved on 1.2% (w/v) agarose gel with appropriate marker to confirm their integrity and amplification. The primers used are listed in the appendix (VI.3). The PCR amplification of the different DNA fragments and the conditions used for amplification were as follows:

II.2.1.3.1 Native M12-V$_H$ domain

The pSynI vector containing the gene for the scFvM12 antibody was purified from the _E. coli_ strain BL21 (DE3) by using the ‘Qiagen mini spin kit’. The 1x master mix used for the PCR reaction was as follows:

- PCR buffer (10x) 5.00 µL
- dNTP mixture (10mM) 1.00 µL
- M12-V$_H$ forward primer (10pmol/µL) 1.00 µL
- M12-V$_H$ backward primer (10pmol/µL) 1.00 µL
- Taq DNA Polymerase (5U/µL) 0.65 µL
- Template DNA (~100ng) 1.00 µL
- MgCl$_2$ variable
- distilled H$_2$O added to 34.35 µL

The master mix was optimized for 0.5mM, 1.0mM, 2.0mM, 2.5mM, 3.0mM, 3.5mM, 4.0mM and 4.5mM of MgCl$_2$. The amplification was performed for each MgCl$_2$ concentration under the following condition:

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>95°C</td>
<td>3.0 min</td>
<td>Initial denaturation</td>
</tr>
<tr>
<td>95°C</td>
<td>30.0 sec</td>
<td>Denaturation</td>
</tr>
<tr>
<td>35 cycles</td>
<td>55°C</td>
<td>30.0 sec</td>
</tr>
<tr>
<td>72°C</td>
<td>1.0 min</td>
<td>Primer extension</td>
</tr>
</tbody>
</table>
II.2.1.3.2 Camelized M12-V<sub>H</sub> domain

To mimic the camelids V<sub>H</sub>H, the M12-V<sub>H</sub> domain was simultaneously mutated at position Gly44Glu, Leu45Arg and Trp47Gly/His (VI.3). The 1x master mix used was as follows:

- PCR buffer (10x) 5.00 µL
- MgCl<sub>2</sub> (50mM) 3.00 µL
- dNTP mixture (10mM) 1.00 µL
- Forward primer (10pmol/µL) 1.00 µL
- Backward primer (10pmol/µL) 1.00 µL
- Taq DNA Polymerase (5U/µL) 0.65 µL
- Template DNA (~100ng) 1.00 µL
- Distilled H2O added to 37.50 µL

The PCR reaction was setup, for the amplification of the mutated product, under the following conditions:

- 95°C 3.0min Initial denaturation
- 95°C 30.0sec Denaturation
- 20 cycles 40°C 30.0sec Primer annealing
- 72°C 1.0min Primer extension
- 72°C 5.0min Final extension

The camelized M12-V<sub>H</sub> domain containing the mutation Trp47Gly and Trp47His were designated as M12-V<sub>H</sub>camG and M12-V<sub>H</sub>camH. Both of these domains also contain Gly44Glu and Leu45Arg mutations respectively.

II.2.1.4 DNA digestion

DNA isolated through plasmid DNA minipreparation (II.2.1.7) was restriction digested according to the manufacturers instructions in the respective buffer(s). 500ng to 1500ng of plasmid DNA was restricted with 16U of enzyme for 3-4 hrs at 37º C. Restriction fragments used for ligation (II.2.1.6), 0.5-1µg of vector DNA and 3-8µg of insert DNA depending on the size of the insert were digested with 10U of restriction enzyme for 2hrs at the recommended temperature. Fragments were subsequently separated by agarose gel
electrophoresis (II.2.1.1) and photo-documented (II.2.1.1) or purified. The restriction of the genes of interest was performed using the following conditions:

A. The PCR amplified native M12-V\textsubscript{H} fragment corresponding to the MgCl\textsubscript{2} concentration 1.0mM, 2.5mM, 3.0mM and 4.0mM was extracted from the agarose gel by using the ‘Qiagen quick PCR purification kit’ (Qiagen). The buffer mix used for digestion with N\textit{de}l and \textit{Xho}I at 37\textdegree C was as follows:

<table>
<thead>
<tr>
<th>Components</th>
<th>Native M12-V\textsubscript{H}</th>
</tr>
</thead>
<tbody>
<tr>
<td>M12 vector DNA (50ng/µL)</td>
<td>30.40 µL</td>
</tr>
<tr>
<td>Restriction buffer 4 (10x, NEB)</td>
<td>4.00 µL</td>
</tr>
<tr>
<td>N\textit{de}l (20,000U/ml, NEB)</td>
<td>0.80 µL</td>
</tr>
<tr>
<td>\textit{Xho}I (20,000U/ml, NEB)</td>
<td>0.80 µL</td>
</tr>
<tr>
<td>BSA (NEB, 10x)</td>
<td>4.00 µL</td>
</tr>
<tr>
<td>distilled H\textsubscript{2}O added to</td>
<td>40.00 µL</td>
</tr>
</tbody>
</table>

B. The restriction digestion of the camelized M12-V\textsubscript{H}\textit{camG}\textsubscript{H} (Trp47Gly/Trp47His) DNA was performed under the following conditions:

<table>
<thead>
<tr>
<th>Components</th>
<th>M12-V\textsubscript{H}\textit{camG}</th>
<th>M12-V\textsubscript{H}\textit{camH}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA (50ng/µL)</td>
<td>10.00µL (50ng/µL)</td>
<td>13.0µL (40ng/µL)</td>
</tr>
<tr>
<td>Restriction buffer 2 (10x, NEB)</td>
<td>4.00µL</td>
<td>4.00µL</td>
</tr>
<tr>
<td>N\textit{de}l (20,000U/ml, NEB)</td>
<td>0.80µL</td>
<td>0.80µL</td>
</tr>
<tr>
<td>\textit{Xho}I (20,000U/ml, NEB)</td>
<td>0.80µL</td>
<td>0.80µL</td>
</tr>
<tr>
<td>BSA (NEB, 10x)</td>
<td>4.00µL</td>
<td>4.00µL</td>
</tr>
<tr>
<td>distilled H\textsubscript{2}O added to</td>
<td>40.00µL</td>
<td>40.00µL</td>
</tr>
</tbody>
</table>

In all the restriction analysis, pK3C (90ng/µL) vector (appendix) restricted with N\textit{de}l and pK3C vector (90ng/µL) restricted with N\textit{de}l and \textit{Xho}I was used as positive and negative control respectively.

**II.2.1.5 Ligation of DNA inserts**

Purified DNA from restriction digest (II.2.1.4) was ligated via their sticky ends (Sambrook, \textit{et al.}, 1989). Generally, a vector-insert ratio of 5:1 was used. The ligation mixture was centrifuged at 10,000xg for 1min and ligation was carried out at 14\textdegree C for 12hrs. PCR products were ligated into bacterial vector pK3C (appendix) digested with \textit{Xho}I and N\textit{de}l restriction enzymes. The ligation of the M12-V\textsubscript{H} fragments into the pK3C vector was performed under the following conditions:
Components

<table>
<thead>
<tr>
<th>Components</th>
<th>Native M12-V&lt;sub&gt;H&lt;/sub&gt;</th>
<th>M12-V&lt;sub&gt;H&lt;/sub&gt;camG</th>
<th>M12-V&lt;sub&gt;H&lt;/sub&gt;camH</th>
</tr>
</thead>
<tbody>
<tr>
<td>pK3C vector DNA (50ng/µL)</td>
<td>2.0µL</td>
<td>10.0µL</td>
<td>10.0µL</td>
</tr>
<tr>
<td>(5ng/µL)</td>
<td>(5ng/µL)</td>
<td>(10ng/µL)</td>
<td>(10ng/µL)</td>
</tr>
<tr>
<td>Template DNA (1x)</td>
<td>2.0µL</td>
<td>2.0µL (20ng)</td>
<td>2.0 µL (20ng)</td>
</tr>
<tr>
<td>(3ng)</td>
<td>(20ng)</td>
<td>(20ng)</td>
<td></td>
</tr>
<tr>
<td>T4 Ligation buffer (1x)</td>
<td>2.0µL</td>
<td>2.0µL</td>
<td>2.0µL</td>
</tr>
<tr>
<td>distilled H&lt;sub&gt;2&lt;/sub&gt;O added to</td>
<td>20.0µL</td>
<td>20.0µL</td>
<td>20.0µL</td>
</tr>
</tbody>
</table>

In all the sticky end ligation reactions, the pK3C vector digested with Ndel and XhoI was used as negative control while the single-cut vector was used as positive control. The M12-V<sub>H</sub> DNA fragment was diluted 10x in water. 100µL of *E. coli* DH5α competent cells were transformed with 20µL of the ligated product(s) using the heat shock method (II.2.1.9).

**II.2.1.6 PCR based analysis of recombinant bacterial clones**

Bacterial clones harboring the plasmid having the gene of interest were identified by PCR (Jesnowski *et. al.*, 1995). Single colonies were picked with sterile toothpicks and used to inoculate a ‘masterplate’ and then dipped into a PCR tube containing 10µL of sterile water. 15µL of the PCR mix was added to the PCR tube. Specific primers annealing to the 5’ and 3’ ends of the cloned gene or primers for the specific vector backbone were used for the PCR amplification reaction. The reaction conditions were the same as before (II.2.1.3.1), except initial denaturation of the DNA was performed for 10 minutes. 10µL of the PCR amplified product was analyzed on a 1.2% (w/v) agarose gel (II.2.1.1). The recombinant bacterial clones were checked by PCR (II.2.1.3) using the following buffer:

- **PCR buffer (10x)**: 5.0 µL
- **Forward primer (10pmole/µL)**: 1.0 µL
- **Backward primer (10pmole/µL)**: 1.0 µL
- **dNTP’s (10mM)**: 1.0 µL
- **MgCl<sub>2</sub> (50mM)**: 3.0 µL
- **Taq polymerase (5U/µL)**: 0.5 µL
- **distilled H<sub>2</sub>O** added to: 38.5 µL

The reaction cycle was same as mentioned in II.2.1.3.2.
II.2.1.7 Isolation of plasmid DNA from *E. coli*

Mini-preparations of plasmid DNA were performed using the ‘Spin mini-prep’ kit (Qiagen), based on the alkaline lysis method (Sambrook et al., 1989), according to the manufacturers instructions. DNA concentration and purity were determined spectrophotometrically. An O.D.260nm of 1 corresponds to ~50 µg/ml of dsDNA, ~40 µg/ml of ssDNA and RNA and ~20 µg/ml of ss-oligonucleotides. The purity of the nucleic acid was ascertained by determining the ratio O.D.260/ O.D.280 according to Müller et al., (1993) and Sambrook et al., (1996). An O.D.260/ O.D.280 = 1.8 for DNA and 2.0 for RNA are acceptable. Contaminants like phenol and proteins lower the ratios. Isolated plasmid DNA was stored at -20°C.

II.2.1.8 DNA sequencing and sequence analysis

DNA sequencing was performed with the dideoxy chain termination method of Sanger et al., (1977). Fluoresceins labeled oligonucleotides were used for DNA sequencing of the gene of interest using ‘Thermosequenase fluorescence labeled cycle sequencing kit’ and a LICOR 4200 IR2 automated DNA sequencer (MWG Biotech). The sequencing reaction mix was resolved on denaturating (6.0M) polyacrylamide gel (4.0 – 6.0% (w/v)).

The ‘Chromas software package’ (Technelysium Pty Ltd., Australia) was used to display the chromatogram files from the LICOR automated sequencer. The files were exported to ‘DNAsis’ (Hitachi Software Engeenering Company Limited, Kanagawa, Japan) and ‘GCG software package’ (Accerlys Incorporation, USA) for evaluation of sequencing data.

The PCR amplified product was subjected to sequence analysis by using the T7 promoter and termination primer in the following buffer mix:

- PCR amplified DNA (80ng/µL) 3.0 µL
- Forward primer (10pmole/µL) 1.0 µL
- Backward primer (10pmole/µL) 1.0 µL
- distilled H₂O added to 11.0 µL

II.2.1.9 Preparation of *E. coli* competent cells

II.2.1.10 Preparation of heat shock *E. coli* competent cells

Heat-shock competent cells were prepared according to the procedure of Inoue et al. (1990). The heat-shock competent cells were prepared from 100mL of a log-phase culture (O.D.600nm ~ 0.5) in LB medium. The bacterial suspension was centrifuged (10min,
4500xg, 4°C) after cooling on ice at 4°C for 10min. The cell pellet was gently re-suspended in 40mL of filter sterilized (0.2µm) ice-cold TB buffer [10mM PIPES, pH6.7; 15mM CaCl₂; 250mM KCl; 55mM MnCl₂]. After cooling on ice for 10min, the suspension was centrifuged (10min, 2300xg, 4°C). The pellet was gently re-suspended in 8ml of ice-cold TB buffer and pure ice-cold DMSO was added to 7.0 % v/v (final concentration). After cooling the suspension on ice for 10min, aliquots of 200µL were shock-frozen in liquid N₂ and stored at –80°C.

II.2.1.10.1 Transformation of heat shock \textit{E. coli} competent cells

\textit{E. coli} was transformed by the heat-shock method as described (Hanahan \textit{et al.}, 1985). For transformation, an aliquot of competent cells was thawed on ice, mixed with 100ng of plasmid DNA or a ligation product (Sambrook \textit{et al.}, 1989). After incubating for 30min on ice, a 2min heat-shock was given at 42°C. The heat-shock cells were incubated for 5min on ice. The volume of the aliquot was made up to 1ml with sterilized SOC medium (II.1.5) and allowed to grow (1h, 150 r.p.m., 37°C) and subsequently plated on LB agar supplemented with ampicillin (100µg/mL) for the selection of recombinant clones.

II.2.1.10.2 Efficiency of bacterial transformation

Efficiency of transformation of each new batch of competent cells was measured by test transformations with known concentrations of supercoiled plasmid pUC18. The transformation rates of greater than 10⁹/µg pUC18 and 10⁸/µg pSyn & pK3C vectors were obtained.

II.2.1.10.3 Culturing and glycerol stock preparation

Individual colonies of all strains were obtained by plating the respective strain on a LB agar plate containing 100µg/ml of ampicillin with subsequent incubation at 28°C for overnight. The plates were stored at 4°C for short periods (less than 2 weeks). LB medium (II.1.5) containing 100µg/ml of ampicillin and 1% (w/v) glucose was inoculated with a single recombinant colony of \textit{E. coli} and grown overnight at 28°C with vigorous shaking (200r.p.m.). Glycerol stocks were prepared by suspending the pellet of 1ml fresh culture in 500 µl of sterile buffer [25mM Tris, pH 7.5; 100mM MgSO₄; Glycerol 50% (w/v)]. Glycerol stocks were stored at -80°C.

II.2.2 Bacterial expression of antibody fragments
II.2.2.1 Small scale expression

A freshly transformed colony of *E. coli* TG1 (II.1.7) harboring recombinant plasmid, having scFvM12 gene, was inoculated in 5.0 ml of 2xTY medium (II.1.5) containing 1% (w/v) glucose and 100 µg/ml ampicillin and cultivated overnight at 37°C with vigorous shaking (180 r.p.m.). The following day, 500 ml (4x) of sterilized 2xTY containing 0.1% (w/v) glucose, dipotassium hydrogen phosphate (3g/L), 100µL of MgSO₄ (25.0%, w/v) and 100 µg/ml of ampicillin was inoculated with 1.0 ml of overnight culture and allowed to grow at 37°C/200r.p.m. to an O.D.600nm of 0.7-0.8. Expression of recombinant proteins was then induced by the addition of 1mM IPTG (final concentration) after reducing the cultivation temperature to 28°C. At the time of induction, 50ml of sterile yeast extract (3 g/L) was also added to the culture. After 4-5hrs of incubation, the cell suspension was harvested by centrifugation (10 min/4000 r.p.m. /4°C).

The M12-V₁ fragments were expressed under the same conditions as the scFvM12, except 2xTy medium was used for the expression.

II.2.2.2 Large scale expression of scFvM12

Few recombinant colonies of freshly transformed *E. coli* TG1 were inoculated overnight at 37°C in 5ml of LB medium containing 1.0% (w/v) glucose and 100µg/mL ampicillin. One milliliter of the pre-culture was used to inoculate 150ml of LB medium containing 1.0% (w/v) glucose and 100 µg/mL ampicillin and allowed to grow at 37°C for 12hrs. This culture was used to inoculate 4L of sterilized minimal medium in a 7L working volume stirred-tank reactor (Biobench fermenter, Applikon, Holland). Glycerol-fed-batch phase was carried out prior to induction using a 50% (v/v) glycerol feed containing 14.7g/L MgSO₄ and 4ml/L PTM₁ trace salts. The aeration rate was set at 0.5vvm (vessel volume per minute) or 2ml/min. The rate of aeration was controlled through synchronized stirrer speed (approx. between 500-1100 r.p.m.). The growing culture was maintained at pH 6.8 and the pH was controlled by using 25% (v/v) NH₄OH. Three six-bladed Ruston impellers of one-third the vessel diameter provided agitation. The culture was grown at 37°C until an O.D.600nm of 50 was reached and then the temperature was reduced to 28°C prior to before induction with 1.0mM of IPTG (final concentration). The culture was harvested after 6-8 hrs of induction by centrifuging the suspension at 5000xg for 15 min. The pellet was stored at –80°C. The supernatant (1L) was processed through a Ni-NTA column (II.2.3.1.2) after re-centrifugation at 7000xg for 30 min.
II.2.3 Purification of antibody fragments

II.2.3.1 scFvM12 purification

The entire purification procedure was performed at 4°C.

II.2.3.1.1 Periplasmic extraction

The periplasmically expressed scFvM12 was purified by osmotic shock (Ausubel et al., 1989) to *E. coli* TG1 suspended in 30mM Tris, pH 8.0, 20% (w/v) sucrose and 1mM (final concentration) of EDTA. The suspension was gently stirred for 15min and then centrifuged at 5000xg. The pellet was re-suspended in 30mM Tris, pH 8.0, 5mM of ice-cold MgSO₄ and a cocktail of protease inhibitors (Roche, Germany). The suspension was gently stirred for 20min and centrifuged twice at 15000xg for at least 20min.

II.2.3.1.2 Ni-NTA chromatography

An appropriate volume of chelating sepharose fast flow resin [nitrilo-acetic acid (NTA), Amersham Pharmacia Biotech] was packed in a column and charged with Ni²⁺ ions. The excess metal ions were removed by passing 5 column volumes (CV) of weak acetic acid (0.02M). The column was equilibrated with 10 CV of binding buffer (50mM Tris pH 8.0, 500mM NaCl, 20mM imidazole and 2mM benzamidine-HCl). The clear supernatant containing the periplasmic protein was loaded on the Ni-NTA column at a flow rate of 1ml/min. After washing with 40CV of binding buffer, the column was connected to the Äkta-FPLC (Pharmacia Biotech) and the protein was eluted with step gradients of 45mM, 65mM, 80mM and 500mM of imidazole in the elution buffer (50mM Tris pH 8.0, 500mM NaCl, 500mM imidazole and 2mM benzamidine-HCl). The protein fractions were pooled and dialyzed overnight against at least 5L of 30mM sodium succinate pH 4.4, 20mM NaCl, 2mM EDTA and 2mM benzamidine-HCl. The dialyzed protein fraction was filtered through 0.2µm filter.

II.2.3.1.3 Cation-exchange chromatography

The dialyzed and filtered Ni-NTA purified protein (II.2.3.1.2) was loaded on a MonoS HR 10/10 column (Amersham Pharmacia Biotech) pre-equilibrated with at least 5 CV of binding buffer (30mM sodium succinate pH 4.4, 20mM NaCl, 2mM DTT, 2mM EDTA and 2mM benzamidine-HCl). After washing the column with 10 CV of binding buffer, the protein was eluted with a 20 CV linear gradient from 0.02M to 0.7M NaCl (30mM sodium succinate, pH 4.4, 2M NaCl, 2mM DTT, 2mM EDTA and 2mM benzamidine-HCl) at a flow rate of 0.6ml
/min. The desired protein fractions were pooled and dialyzed against 5L of 30mM Tris, pH 8.0, 20mM NaCl, 2mM EDTA and 2mM benzamidine-HCl–HCl.

II.2.3.1.4 Anion-exchange purification

Cation exchange purified, dialyzed and filtered protein (II.2.3.1.3) was loaded at 1ml/min on a MonoQ HR 10/10 column (Amasham Pharmacia Biotech) pre-equilibrated with at least 5 CV of binding buffer (30mM Tris, pH 8.0, 20mM NaCl, 2mM DTT, 2mM EDTA and 2mM benzamidine-HCl). After washing with 8 CV of binding buffer, the protein was eluted with 20 CV linear gradient from 0.02M to 0.6M NaCl (30mM Tris, pH 8.0, 2M NaCl, 2mM DTT, 2mM EDTA and 2mM benzamidine-HCl) at a flow rate of 0.6ml/min.

II.2.3.1.5 Size-exclusion chromatography

Anion-exchange purified protein (II.2.3.1.4) was concentrated to an appropriate volume using Centricon 20 plus concentrators (Amicon) and loaded on a Superdex 75 Hiloa 16/60 column (Pharmacia). The isocratic elution was performed at 0.5ml/min in 20mM Tris, pH 8.5, 20mM NaCl, 0.5mM EDTA and 1mM DTT. The column was calibrated in the respective buffer with aprotinin (6.5 kDa), cytochrome C (12.4 kDa), carbonic anhydrase (29 kDa), BSA (66 kDa) and dextran blue (2000 kDa) as molecular weight standards (Fluka).

II.2.3.2 Purification of the cytoplasmic M12-VH domain

II.2.3.2.1 M12-VH domain inclusion bodies

The bacterial pellet (II.2.2) was suspended in resuspension buffer 50mM Tris, pH 8.0; 500mM NaCl; 20mM MgSO4; 2mM CaCl2; 2mM benzamidine-HCl containing 0.5mg/ml lysozyme and 1U/ml of benzonase-II. The suspension was incubated on ice for 1hr. Triton-X100 was added to the final concentration of 1% (v/v) and the suspension again incubated for 1hr on ice. The suspension was sonicated (Braun Biotech Sonicator) with 0.5 repeating cycle using a power level of 50W. The sonication was performed for 30 sec on ice bath with intermittent cooling of 5-6min. After sonication, the suspension was centrifuged at 25,000xg for 30min at 4°C and the supernatant kept separately. The pellet was washed twice with washing buffer (resuspension buffer containing 1% (v/v) Triton-X100 but no lysozyme). Then, the pellet was resuspended in 50mM Tris, pH8.0; 100mM NaCl; 6M guanidine-HCl and incubated for 1hr on ice with intermittent vortexing. The insoluble material was removed by centrifugation at 25,000xg for 30min at 4°C. The supernatant was loaded on Ni-NTA column, pre equilibrated with equilibration buffer (50mM Tris, pH 8.0; 500mM NaCl; 5mM imidazole; 6M guanidine-HCl; 2mM benzamidine-HCl). The column washed with equilibration buffer followed by washing with a linear gradient of 25CV of equilibration
buffer without guanidine-HCl. The protein was eluted with 15CV of 50mM Tris, pH 8.0; 500mM NaCl; 250mM Imidazole; 2mM benzamidine-HCl.

**II.2.3.2.2 Soluble M12-V<sub>H</sub> domain purification**

The bacterial pellet (II.2.2.1) was thawed on ice and subjected to lysis in buffer made up of 25ml of 25mM Tris, pH7.5; 500mM NaCl; 5mM imidazole; 2mM benzamidine-HCl; mixture of protease inhibitors (Roche, 1mg/ml each of aprotinin, leupeptin, pepstatin, 100mM PMSF), 1U/ml benzonase-II and 1mM of lysozyme. After incubation for 4hrs on ice, the suspension was sonicated as in II.2.3.2.1. After sonication, the suspension was centrifuged at 18000xg for 30min at 4°C. The supernatant was loaded on 0.5ml of Ni-NTA resin column (II.2.3.1.2). The column was washed with 8 CV of wash buffer (25mM Tris, pH7.5; 500mM NaCl; 20mM imidazole; 2mM benzamidine-HCl). The protein was eluted with 5 CV of elution buffer (25mM Tris, pH7.5; 500mM NaCl; 500mM imidazole; 2mM benzamidine-HCl).

**II.2.4 Protein analysis**

**II.2.4.1 Concentration and dialysis**

Small samples were concentrated using ‘NanoSep or MicroSep’ concentrators (Pall Filtron, Dreieich, Germany). Larger samples were concentrated in a stirred cell (Millipore) under N<sub>2</sub> gas pressure (6bar). In all cases, a membrane with molecular weight cut-off (MWCO) of 10kDa was used.

Buffer exchange between chromatographic steps or after purification was performed in a classical way by using the Servapor™ (Serva Electrophoresis GmbH, Heidelberg) dialysis tubing (diameter 16mm and 26mm, MWCO-10 kDa).

**II.2.4.2 Protein quantification**

The concentrations of the proteins were determined in duplicate on the assumption that an O.D.280nm of 1.0 is approximately equal to 0.7mg/ml of scFv using a single beam spectrophotometer (Eppendorf, Germany).

The BCA protein assay Kit (Pierce) was also used to determine the concentrations of purified proteins according to the manufacturers protocol using bovine serum albumin (BSA) as standard. The Pierce BCA protein assay is a detergent-compatible formulation based on bicinechoninic acid (BCA) for the colorimetric detection and quantification of proteins (Smith et al., 1985; Sorensen and Brodbeck, 1986). This method combines the reduction of Cu<sup>2+</sup> to Cu<sup>+</sup> by proteins in an alkaline medium with a sensitive and selective colorimetric detection.
II-MATERIALS AND METHODS

of cuprous cations (Cu$^{+1}$) coupled with bicinchoninic acid (Wiechelman et al., 1988). The standard curve was prepared in an appropriate solvent for every assay.

II.2.4.3 SDS-PAGE and IEF

Proteins were electrophoretically separated by SDS-PAGE (Laemmli, 1970, Ausubel et al., 1998) by using the Mini Protean II system (BioRad) and self-cast 12% gels at 150V for 60 min. Before loading onto the gel, the samples were boiled with 4µl of 3x Laemmli loading dye buffer for 3-5min [for 10 ml buffer - 2.4 ml Tris-Cl (1M) pH 6.8, 3 ml SDS (20.0% w/v), 3 ml Glycerol (100.0%), 1.6 ml β-mercaptoethanol, 0.0006 % (w/v) bromophenol blue dye (stored at –20°C)]. The protein bands were detected after staining the gel with Commassie brilliant blue R250 (0.1% (w/v) in destaining solution) for 30 min at room temperature under constant rocking. The gel was destained until the protein bands become clearly visible in destaining solution [methanol (10% v/v); glacial acetic acid 10% (v/v)]. The SDS electrophoresis buffer used was 25mM Tris, pH 8.3; 192mM glycine; 0.1% (w/v) SDS. Protein Marker M12 (Invitrogen) was used as reference.

For iso-electric focusing (IEF), Phast IEF gels (pH 3-9) were used. Electrophoresis and staining procedures for the Phast-system were used according to protocols supplied by the manufacturer (Amersham Pharmacia Biotech). The following buffers were used for casting the gels:

4x Separating gel Buffer* (100ml)
2M Tris-HCl, pH 8.8 75.0 ml
10% (w/v) SDS 4.0 ml
de-ionized water 21.0 ml

4x Stacking gel Buffer* (100ml)
1M Tris-HCl, pH6.8 50.0 ml
10% (w/v) SDS 4.0 ml
de-ionized water 46.0 ml

* Both buffer solutions are stable for months in the refrigerator.

The calculation for an X% separating gel was performed using the formula:

\[
\text{Acrylamide solution} \times \frac{X}{3} \text{ ml}
\]

[30% (w/v) acrylamide, 0.8% (w/v) bis-acrylamide]

4x Separating gel buffer 2.50 ml
deionised water \((7.5- \times \frac{X}{3})\) ml
10% (w/v) ammonium persulphate 50.0 µL
TEMED 5.0 µL
The 5% stacking gel (2x) was prepared as follows:

- Acrylamide solution: 0.67 ml
- 4x Stacking gel buffer: 1.00 ml
- Deionised water: 2.30 ml
- 10% (w/v) ammonium persulphate: 30.0 µL
- TEMED: 5.0 µL

**II.2.4.4 Western blotting**

Immuno-blotting was performed essentially as described by Towbin et al., (1979). Protein samples were subjected to SDS-PAGE (II.2.4.3) and electro-blotting transferred protein to a Hybond C™ nitrocellulose membrane (0.45µm) (II.1.2) at a constant current of 200mA for 55 minutes by using the transfer buffer [25mM Tris-HCl, pH8.3; 92mM glycine; 20% (v/v) methanol]. The blot was blocked with 1xPBS (pH 7.4; 137 mM NaCl; 2.7mM KCl; 8.1mM Na₃PO₄×2H₂O; 1.5mM KH₂PO₄) containing 5% (w/v) skimmed milk powder (Marvel) for 30min. The blot was probed with a primary antibody for 45min (murine monoclonal antibody anti 9E10, which recognizes the c-myc tag or murine monoclonal anti-Histidine antibody which recognizes the histidine tag; each in 1:4000 dilution of a 1.5mg/ml stock solution in 1xPBST). Bound primary antibody was detected with alkaline phosphatase labeled secondary antibody for 45min (goat anti-mouse mAb in 1:4000 dilution of 1.5mg/ml stock solution; Dionova). Alkaline phosphatase activity was visualized upon addition of NBT substrate (Pierce) prepared in AP-buffer. Washing and dilutions were performed using the 1x PBST, pH 7.4.

Sufficient amount of crystals were collected from the hanging drops and centrifuged at 500xg for 1min. The crystal pellet was washed twice with 10µl of 50mM Tris, pH 8.0. The crystal pellet and supernatant were stored overnight at 4°C and then boiled for 3-5 min with an appropriate volume of protein loading dye buffer before loading on SDS-PAGE (II.2.4.3) for western blotting.

**II.2.4.5 Silver staining**

For silver staining, the silver stain Plus kit (Bio-Rad Laboratories Inc., USA) was used. Briefly, the manufacturers procedure is as follows: after gel electrophoresis, the gel was washed in 200ml of de-ionized water for 15min with gentle agitation and then incubated for 20 minutes in fixative enhancer solution [reagent grade methanol, 50% v/v; reagent grade acetic acid, 10% (v/v); fixative enhancer concentrate, 10% (v/v) and deionized water, 30% (v/v)]. The gel was rinsed twice for 20min with 400ml of deionised water with gentle
agitation. The gel staining solution was prepared as follows: 35ml of deionised water was taken in a clean beaker and 5ml of silver complex solution was added, followed by the addition of 5ml of reduction moderator solution and 5ml of image development reagent. All solutions were mixed well. In this mixture, 50ml of development accelerator solution was added just before gel staining. Then, the mixed solution was added to the gel-containing vessel. The gel was allowed to develop in the dark until the protein bands were stained. The staining reaction was stopped by placing the gel in 5% (v/v) acetic acid solution for at least 15min. Finally, the gel was rinsed twice in deionised water for 5 minutes.

**II.2.4.6 Discontinuous non-denaturing electrophoreshis**

Native gel electrophoresis was performed as SDS-PAGE (II.2.4.3) without SDS in separating and stacking gel buffer (Schägger *et al.*, 1994). The sample-loading buffer included 25ml Tris-HCl (0.5M), pH 6.8; 20ml glycerol (100% v/v); 1mg (w/v) bromophenol blue dye. The separation of the protein was performed at a constant current of 100mA for 4-5 hrs using the native gel electrophoresis buffer (25mM, Tris base; 192mM, glycine).

The crystals were collected under the microscope, centrifuged at 400xg for 3min and the supernatant collected separately. The crystallized pellet was dissolved in an appropriate volume of buffer 50mM Tris, pH8.5. 8µL of dissolved crystal pellet were mixed with 5µL of 2x non-denaturing sample loading buffer dye and then electrophoresed on a 12% non-denaturing discontinuous gel at 100mA constant current for 5hrs.

**II.2.5 Stability analysis**

The stability of the gel filtration (II.2.3.1.5) purified scFvM12 in 20mM Tris, pH (7.5, 8.0 and 8.5); 20mM NaCl; 0.5mM EDTA and 1mM DTT was studied as a function of pH, temperature and salt concentration. Sodium chloride was added to scFvM12 solution to a final concentration of 100mM. Different pH solutions with and without added salt were incubated at 4°C and room temperature for 60-120 days. The stability of the protein was checked through silver staining and MALDI-TOF mass spectrometry.

**II.2.6 Mass spectrometry**

The protein sample was dialysed against 5mM Tris, pH 8.0 and the dialyzed sample was mixed with the MALDI-TOF matrix solution [10mg/ml sinapinic acid, 50% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid], so that the final concentration of the protein should be in the range of 5-10µmole. A 1.0µL of the mixture was placed on the MALDI-TOF target and allowed to dry at room temperature. The molecular mass was analyzed using
a MALDI-TOF mass spectrometer (Tof spec 2E, Micromass UK Limited, UK) in its linear mode.

For electrospray mass spectrometric (ESI-MS/MS) analysis, the SDS-PAGE separated bands were excised from the gel. The excised bands were reduced with 10mM dithiothreitol in 0.1M NH$_4$HCO$_3$. After reduction the bands were alkylated with 55mM iodoacacetamide in 0.1M NH$_4$HCO$_3$ at room temperature for 20min in dark. The alkylated protein was washed with 200µL of 0.1M NH$_4$HCO$_3$ for 15 min. The gel particles were suspended by spinning. The suspended gel particles were rehydrated at 4°C in the digestion buffer 50mM NH$_4$HCO$_3$, 5mM CaCl$_2$ and 12.5ng/µL of trypsin for 30-35min. The sample left at 37°C overnight after covering the gel pieces with the rehydration buffer without trypsin. After overnight typtic digestion, the gel particles washed in 10-15µL of 25mM NH$_4$HCO$_3$ at 37°C. After spinning, acetonitrile equal to 1-2 times volume of gel particles was added and the mixture incubated at 37°C with shaking for peptide extraction. In the supernatant 40-50µL of 5% formic acid was added and the mixture vortexed for 15min. at 37°C. The extraction was repeated twice and the resulting supernatant mix was dried in a vacuum centrifuge. The vacuum dried powder dissolved in distilled water having 5% formic acid and 30% acetonitrile. 1-2µL of the dissolved peptide mixture was subjected to ESI-MS/MS analysis (Micromass UK Limited, UK).

II.2.7 Flow cytometric analysis

The stability and qualitative functionality of the scFvM12 was determined by flow cytometric (FACS) analysis. Human breast cancer cell lines COS and MDA MB231 were purchased from the American Type Culture Collection (ATCC, Bethesda, USA). The cell lines were maintained at 37°C with 5% (v/v) CO$_2$ in RPMI 1640 medium (Gibco BRL) supplemented with 10% (v/v) foetal calf serum (FCS).

The scFvM12 (10µg) and scFvH10 (10µg, as negative control) was incubated with approximately 5x10$^6$ cells. The scFvM12 and the scFvH10 (as negative control) bearing the c-myc tag were detected with the murine monoclonal anti9E10 antibody (7.5µg), followed by an anti-mouse Fc specific phycoerythrin conjugated (Fab’)$_2$ antibody fragment (Jackson Immunoresearch). The fluorescence intensity was analyzed with a flow cytometer (FACScan instrument from Becton-Dickinson) and the mean fluorescence intensity was determined on a log scale. All incubations were carried for 30 min out on ice. Every incubation step was followed by a wash step, using ice-cold 1xPBS and cells sedimented at 500xg at 4°C.
II.2.8 M12-V\textsubscript{H} domain generation

Proteins are often composed of several individual domains connected by relatively unstructured linker domain. The reduction of a multidomain protein into its individual domains through proteolysis provides a mechanism for structural characterization of the protein. Limited proteolysis has been used either in the improvement of protein crystal quality (Nieves-Alicea et al., 1998) or to crystallize the number of other conformationally heterogeneous proteins (McPherson et al., 1990 and Chitarra et al., 1995). Proteolytic clipping of the scFv linker region is a useful technique for the crystallization of scFv’s domains. It would be more challenging and interesting to generate the single antibody domains for crystallization without using the recombinant DNA technology. In this thesis, the in-drop limited proteolysis was used to generate and crystallize the single antibody domain of scFvM12.

Single antibody domain (M12-V\textsubscript{H}) of scFvM12 antibody have been generated through limited in-vitro proteolysis using low concentration of broad-spectrum and non-specific proteases at 4°C and 17°C. Three proteases were screened to digest the scFvM12 antibody at 1:1000 molar ratio (protease:protein) for generating the isolated single antibody domain. The enzymes used in the study were subtilisin Carlsberg [from \textit{B. licheniformis} (Sigma), which has preference for large uncharged residues], thermolysin [from \textit{B. thermoproteolyticus rokko} (Sigma), which cleaves preferentially before Leu residues] and ProteinaseK (from \textit{Tritirachium album}, Qiagen), which cleaves peptide bonds mostly after the carbonyl group of N-substituted hydrophobic aliphatic and aromatic amino acids].

For M12-V\textsubscript{H} domain generation, as shown later on by mass spectrometric analysis, the scFvM12 (in 50mM Tris, pH 8.5; 20mM NaCl; 1mM DTT and 0.5mM EDTA) was buffer exchanged to 20mM Tris, pH7.5. Buffer exchanged scFvM12 solutions (8 to 13.5 mg/ml), were supplemented with as follows:

A. subtilisin Carlsberg was added in 1:100 molar ratio (protease:protein) just before setting the hanging drops.

B. thermolysin was added in 1:50 molar ratio (protease:protein). Zinc chloride (ZnCl\textsubscript{2}) was added to a final concentration of 1µM to the mixture, since thermolysin activity is Zn\textsuperscript{2+} ion dependent. The mixture was incubated at room temperature for 30min before setting up the trials using the sparse matrix screen (Hampton Research) at 22°C.
II-MATERIALS AND METHODS

II.2.9 Crystallization

II.2.9.1 Hanging drop method

The general aspects of the crystallography are dealt briefly in appendix (page 158). Conditions for the crystallization of the antibody fragments were screened at 4°C by the sparse matrix approach (Cudney et al., 1994) using the hanging-drop method (McPherson, A., 1982) in 24-well VDX Linbro plates (Hampton Research). A 3µL protein drop was suspended on silanized glass cover slips (22mm diameter, Hampton Research) and inverted over 500µL of reservoir solution and sealed using silicone grease coated on the periphery of the wells.

II.2.9.1.1 M12-V\textsubscript{H} domain crystallization

Preliminary crystallization for M12-V\textsubscript{H} domain was screened by using the ‘Crystal Screen I & II’ (Hampton Research) and sodium malonate screen. The M12-V\textsubscript{H} domain was generated in the hanging drop as described in II.2.8. The initial crystal conditions were optimized as follows:

A. The M12-V\textsubscript{H} crystals generated after subtilisin treatment in 100mM MES, pH 6.5, 10mM ZnSO\textsubscript{4}, 25% (v/v) PEG 550monomethylether, 4mM DTT and 0.04% (v/v) NaN\textsubscript{3}. The condition was optimized and improved by changing the following variables:

- **Protein concentration** (at 4 °C): The protein to well buffer (v/v) ratio was changed from 1:1 to 1:2 & 2:1 to observe the effect of the difference in the rate of vapor diffusion on the crystal morphology.

- **pH screening**: The pH is a major variable in the crystallization of proteins; therefore, for the desirable pH screening, the buffers having their pKa value close to the desirable pH were chosen.

  At 4°C: The screening was performed around the initial screen pH value of 6.5 at pH intervals of 0.5 units. Sodium citrate at pH 4.0, 4.5, 5.0, 5.5, MES at pH 6.0, 6.5, MOPS at pH 7.0, HEPES at pH 7.5 and Tris at pH 8.0 and 8.5 were used respectively at a concentration of 100mM.

  At 17°C: The pH screening was performed using MES (100mM) at pH 6.0 to 6.5 and Tris (100mM) at pH 8.0 to 8.5 with an interval of 0.1 units respectively.

- **PEG screening** (at 4°C): PEG is an important precipitant in the development of the crystals. Therefore, by varying the composition of PEG, it is possible to manipulate the crystal morphology and its quality.
The percentage of PEG 550MME was changed to observe its effect on crystal formation. The concentration (% v/v) was changed from 5.0 to 60.0 at 5 unit intervals.

Higher molecular weight PEG’s was tested instead of PEG 550MME to obtain bigger mono-crystals. The PEG’s tested were PEG 750MME, PEG 1000, PEG 1500, PEG 3000 PEG 5000 and PEG 8000.

**Zinc sulfate screening** (at 4°C): The effect of zinc sulphate effect at concentrations ranging from 0 to 100 mM was screened at 10mM intervals.

**Effect of subtilisin**: Since the cleavage of the scFvM12 antibody is a function of both the incubation time and the concentration of the protease, these variables were considered for improvement of the crystal morphology and their quality.

**Effect of metal compounds**: It has been observed that different metal ions not only induce the crystal formation but also have the ability to change their appearance. Therefore, zinc sulfate was substituted with other divalent metal ions such as lithium sulphate, magnesium chloride, cadmium chloride, and cobalt chloride to study their influence on crystal morphology.

**Effect of additives**: In addition to their inherent influence, viscous additives like glycerol might be helpful in the production of more symmetrical crystals by slowing down the rate of vapor diffusion and drop super saturation. The generation of crystals in the presence of glycerol can be further useful due to the cryo-protectant effect of glycerol. The glycerol concentrations added to the reservoir solution were ranged from 5%-50% (v/v) at 5% (v/v) intervals.

N-octyl-β-D-Glucopyranoside is considered a mild detergent, which has been proven to be especially useful in the crystallization of hydrophobic proteins by exerting solubilizing effects. The M12-V$_H$ fragment has a strong hydrophobic interface, which usually interacts with the V$_L$ domain in the scFv’s. This detergent was used at 20mM concentration to improve the quality of the M12-V$_H$ fragment crystals.

B. The M12-V$_H$ crystals generated by *thermolysin* treatment of scFvM12 in 100mM HEPES pH7.5, 2.0M ammonium sulphate, 2.0% (v/v) PEG 400 and 0.04% (v/v) NaN$_3$. This condition was optimized as follows:

**pH**: HEPES buffer (100mM) of variable pH i.e. 7.2, 7.4, 7.6, 7.8 was used. Sodium acetate (100mM) was used for pH 6.2 and 6.8.
**Precipitant type and concentration**: The PEG 400 concentration was varied from 1.0 to 6.0 at 2.0% (v/v) intervals to improve the crystal quality. The effect of PEG 550MME at 0.5%, 1.0%, 2.0%, 5.0% (v/v); instead of PEG 400 was also investigated.

**Salt concentration**: Ammonium sulphate concentration was varied from 1.0M to 1.8M at 0.2M intervals.

**Protein to protease ratio**: The thermolysin was added to scFvM12 solution in 100:1, 200:1 and 400:1 molar ratio (protein:protease) and incubated for 30 min at room temperature before setting the hanging drops at 17°C.

**Buffer**: Tris-HCl (100mM) was used at pH 7.2, 7.4 and 7.8 instead of HEPES buffer to study the effect of buffer on the crystal quality.

C. **Sodium malonate screen**: Sodium malonate has been recently shown to be as an effective reagent in the crystallization of many different proteins (McPherson, A., 2001). M12-VH fragment crystallization has been tested in sodium malonate at pH 6.5, 7.5 and 8.5. Since the saturation of sodium malonate is temperature and pH dependent, the sodium malonate solution was separately saturated at room temperature for different pH values.

**II.2.9.1.2 scFvM12 crystallization**

Preliminary crystallization screening of the scFvM12 using the Crystal Screen I, II and the ammonium sulphate grid screen (Hampton Research, CA, USA) produced either micro crystals or micro needles. The following conditions:

I 100mM Tris pH 8.0, 1.5M LiSO4, 4mM DTT, 0.04% (w/v) NaN3.
II 100mM Sodium citrate pH 5.0, 1.6M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN3.

were optimized to get the improved crystals at 4°C, 17°C and 25°C.

A. **Optimization of condition II.2.9.1.3.1**

Initially at 4°C, few sea urchins like crystals were observed after 4-5 days of incubation. The following parameters were varied to improve the crystal quality -

**Effect of pH and LiSO4 concentration**: Trials were setup at different pH values pH 5.0 (100mM sodium citrate), pH 6.5 (100mM MES), pH 7.0 (100mM MOPS), pH 7.5 (100mM HEPES), pH 7.8, 8.0, 8.2 and 8.5 (100mM Tris) and salt combinations. The LiSO4 concentration was varied from 0.8M to 1.6 M at 0.1M unit intervals. The pH value was also optimized by using different buffers such as MES (6.0 and 6.5), MOPS (7.0), HEPES (7.2, 7.4 and 7.6) and Tris (7.8 and 8.5) for the condition: 1.5M LiSO4; 10mM hydroxyectoine; 4mM DTT; 0.04% (w/v) NaN3 at 4°C.
Effect of compatible solutes: In general, compatible solutes enhance the solubility of the proteins. Since, the scFv’s have the tendency to undergo multimer formation during storage, compatible solutes were added to reduce this phenomenon in hanging drops. The compatible solutes added to the hanging drops in buffer [100mM Tris pH8.0, 1.5M LiSO₄, 4mM DTT, 0.04% (w/v) NaN₃] were betaine (0.25, 0.5, 0.75 & 1M), mannose (0.25, 0.5, 0.75 & 1M) and hydroxyectoine (25, 50, 75 & 100mM). The LiSO₄ salt concentration was also optimized for 100mM HEPES pH7.6, LiSO₄ [0.5M, 1.0M, 2.0M and 2.5M], 10mM hydroxyectoine, 4mM DTT, 0.04% (w/v) NaN₃ at 4°C.

B. Optimization of condition II.2.9.1.3.II
The ammonium sulphate screening (Hampton Research) of scFvM12 (~5.00 or 10.0 mg/ml) resulted in micro-crystals at 4°C and 25°C. The following parameters were optimized to improve the crystal quality:

Alteration of pH and salt: Different combination of pH and ammonium sulphate concentration was used. The pH of sodium citrate buffer was varied from 4.6 to 5.2 at 0.2 unit intervals, while the concentration of ammonium sulphate were altered up to 0.6, 1.0, 1.2, 1.4, 1.6, 1.8 and 2.0M by changing one variable at one time.

Effect of buffer: Sodium citrate buffer was replaced with HEPES (100mM) at pH 7.4, 7.6, 7.8 and Tris (100mM) at pH 8.0, 8.2 and 8.4 with 0.6, 0.8, 1.0 and 1.2M ammonium sulphate.

Effect of the sodium citrate concentration: The sodium citrate concentration at pH 5.6 was varied from 0 to 100mM at 10mM interval.

Effect of additives: Glycerol (88% v/v) at 2% and 5% (v/v), hydroxyectoine at 10, 50, and 100mM, PEG 550MME at 1%, 5%, 10%, 15% and 20% (v/v) and n-octyl-β-glucopyranoside at 0.5% (w/v) were used as additives in 100mM sodium citrate pH 5.6, 1.2M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃. Different PEG types (300, 600, 750, 1000,1500, 2000, 3000, 5000 and 8000) at 1% (v/v), glycerol (88% v/v) at 0.5%, 1.0% and 1.5% (v/v), ethanol (100% v/v), dioxane at 1%, 2% and 5% (v/v) and ethylene glycol at 1%, 2% and 5% (v/v) were used in a 100mM sodium citrate pH 5.6, 1.2M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃ buffer condition to improve the quality of the crystals.

Effect of divalent metal ions: CdCl₂ (10mM), ZnCl₂ (10mM), CoCl₂ (10mM) and MgCl₂ (10mM) were added in 100mM sodium citrate pH 5.0, 1.2M ammonium sulphate, and 4mM DTT and 0.04% (w/v) NaN₃ at 25°C to improve the crystals.
Effect of temperature: The trials were setup at 30°C and 37°C with 100mM sodium citrate pH5.0, 1.2M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃.

Effect of the Alan D’arcy oil: To reduce the rate of vapor diffusion and therefore slowing down the rate of drop super saturation, Al’s oil (Hampton Research) was layered over the well buffer (500 µL) using different volumes (100, 200, 300, 400 µL). The oil was layered over the buffer after the buffer was mixed with the protein solution for setting the hanging drops.

Sodium malonate screen: The hanging drops were setup at 17°C and 25°C in sodium malonate at pH 7.5 (50% saturated at RT) and 8.5 (60% saturated at RT).

II.2.9.1.3 M12-V₁ and MUC1 complex crystallization

Crystals of M12-V₁ in complex with MUC1 peptide were obtained by soaking method at 4°C. Firstly, the M12-V₁ domain crystals were generated through hanging drop method by mixing the scFvM12 and subtilisin in 250:1 and 500:1 (protein:protease) molar ratios (II.2.9.1.1). This resulted in the generation of suitable M12-V₁ domain crystals after one year (III.3.4). The drop containing M12-V₁ domain crystals supplemented with 2µL of MUC1 (5mg/ml in 20mM sodium acetate, pH 5.0) and the drop was examined for any change in morphology for another two months.

II.2.9.1.4 scFvM12 and MUC1 complex crystallization

For crystallizing the scFvM12 in complex with the MUC1 peptide, the gel purified scFvM12 (~10mg/ml, in 20mM Tris, pH 8.5) was mixed with synthetic c-terminal biotinylated MUC1 peptide (20mg/ml, in 20mM sodium acetate, pH 5.0) in a molar ratio of 1:5 (protein to peptide) and the protein was allowed to saturate with the peptide overnight at 4°C. After 24 hrs, the hanging drops were set up at 17°C with the mixture by using the sparse matrix crystal screen I and II (Hampton Research). After 5 months, the potential complex crystal producing conditions were identified and optimized as follows:

A. 100mM HEPES pH 7.5, 2% (v/v) PEG 400, 2M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃: The condition was optimized at 17°C in terms of pH, ammonium sulfate concentration and polyethylene glycol types. The pH of the HEPES buffer was varied from 6.9 to 7.8 at 0.1 unit intervals. The ammonium sulphate concentration was varied from 1.0M to 2.4M at 0.2 unit intervals. The different PEG types used at 1% (v/v) were PEG 400, 550MME, 1000, 1500 and 3000. Effect of few additives like glycerol (% v/v) 1.0, 2.0, 4.0, 6.0, 8.0, 10.0, 12.0, 15.0 and dioxane (% v/v) 0.1, 0.2, 0.5, 1.0, 1.5, 2.0, 4.0, 6.0, 8.0 and 10.0 were also studied.
C. 100mM sodium acetate pH 4.6, 10mM CoCl$_2$, 1.0M 1,6-hexane-diol, 4mM DTT, 0.04% (w/v) Na$_3$N: The condition was optimized in terms of sodium acetate buffer pH from 4.0 to 5.2 at 0.2 unit intervals, the concentration of sodium acetate at pH 4.6 from 10 to 100mM at 10mM intervals, cobalt chloride concentration at 0.2, 0.5, 1.0, 2.0, 5.0, 8.0, 10.0mM and the concentration of 1,6-hexane-diol from 0.7M to 1.4M at 0.1M unit interval with and without cobalt chloride (10mM). Effect of additives like glycerol at 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 % (v/v) and dioxane at 0.1, 0.2, 0.5, 0.8, 1.0, 1.5, 2.0 % (v/v) were also studied.

D. 100mM MES pH6.5, 50mM CsCl$_2$, 30% (v/v) jeffamine M600, 4mM DTT, 0.04% (w/v) Na$_3$N: The condition was optimized in terms of MES buffer pH from 5.8 to 7.0 at 0.2 unit intervals, the concentration of MES buffer at pH 6.5 from 20mM to 100mM at 10mM unit interval, cesium chloride concentration from 0mM to 80mM at 10mM unit interval and the concentration of jeffamine from 5 to 50 at 5% (v/v) intervals.

E. Sodium formate concentration: The sodium formate concentration was varied from 1.5, 2.0, 2.5, 3.0 and 3.5M for a 5:1 molar ratio between MUC1 and scFvM12. MUC1 and the scFvM12 in 3:1 molar ratio trials were also setup at 3.1, 3.2, 3.3 and 3.4M concentration of sodium formate.

F. Sodium malonate condition: The scFvM12 and the MUC1 peptide were mixed in a 10:1 molar ratio and kept overnight at 4°C. Subtilisin Carlsberg was added to the mixture in a 100:1 molar ratio (protein mixture to protease). The 3 µL hanging drop was set up with 1.5µL of protein and 1.5µL of well buffer i.e. sodium malonate (50% saturated at room temperature) at pH 8.0 and 8.5.

II.2.9.2 Gel crystallization

Gels are very efficient media for growing macromolecular crystals (Robert et al., 1988). Silica gel is particularly advantageous in terms of stability and usability over a wide range of temperature (0-60°C) and compatible with a wide variety of precipitants and additives used for crystal growth. Depending upon the macromolecule and selected conditions, gels can reduce nucleation (heterogeneous and secondary) and sedimentation, provide additional stability and allow crystals to larger sizes.

For gel crystallization of scFvM12 (~8.5mg/ml), the ‘Silica Hydrogel Kit’ (Hampton Research) was used. The sitting drops (5µL, Figure VI-4A) were set up according to the manufacturer instructions with slight a modification of the protocol at 4°C and 17°C. The gel was cured at room temperature for 1-2 hrs and the precipitant was allowed to absorb for 7-8
II-MATERIALS AND METHODS

hrs by the silica gel. The precipitant and protein volume was 5µL and 10 µL respectively. The drops were setup in the following buffer conditions:

A. 100mM HEPES pH7.5, 1.5M LiSO₄, 4mM DTT, 0.04% (w/v) NaN₃
B. 100mM Tris pH8.0, 1.5M LiSO₄, 4mM DTT, 0.04% (w/v) NaN₃
C. 100mM Tris pH8.5, 1.4M LiSO₄, 4mM DTT, 0.04% (w/v) NaN₃
D. 100mM Tris pH8.5, 1.5M LiSO₄, 4mM DTT, 0.04% (w/v) NaN₃
E. 100mM Sodium citrate pH 5.0, 1.6M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃

The gel crystallization of the MUC1 and scFvM12 complex in 5:1 molar ratio was performed at 17°C, using the procedure described above in the following buffer condition:

A. 100mM HEPES pH 7.5, 2.2M ammonium sulphate, 5% (v/v) PEG 550MME, 4mM DTT, 0.04% (w/v) NaN₃
B. 100mM Tris pH 7.5, 2.2M ammonium sulphate, 2% (v/v) PEG 550MME, 4mM DTT, 0.04% (w/v) NaN₃

II.2.9.3 Streak seeding

To improve the quality of the crystals, the streak seeding technique described by Ducruix A. and Giege’ R. (1999) (in ‘Crystallization of Nucleic acids and proteins: A practical approach’, text book, page no. 185-186) was used. Cat whiskers were defatted with 70% (v/v) ethanol, washed with deionised water and dried. Seeds were taken from the drop (40mM sodium citrate pH 5.0, 1.2M ammonium sulphate, 4mM DTT, 0.04% w/v NaN₃). Seeds were transferred to the hanging drops, pre-equilibrated (72hrs) against 100mM sodium citrate pH 5.0, 0.5M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃. The source and the receiving wells were resealed immediately after the transfer.

II.2.9.4 Crystal growth in meta-stable zone

It has been argued that the meta-stable zone is an optimum zone for crystal growth to avoid excessive nucleation and to facilitate the slow growth rate (Stura et al., 1992). Recently, a simple method for growing the large crystals in the meta-stable zone using the vapor diffusion technique was described (Saridakis et al., 2000). This method induces crystal growth in meta-stable zone by decoupling it from nucleation. The hanging drops of scFvM12 (~8mg/ml) were setup at 25°C in 40mM sodium citrate pH 5.0, 1.2M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃. After different time intervals i.e. 0, 24, 36, 48, 56, 72, and 84 hrs, the drops were transferred to the wells containing 40mM sodium citrate pH 5.0, 1.0M ammonium sulphate, 4mM DTT, 0.04% (w/v) NaN₃ as well buffer.
II.2.9.5 Macroseeding

For macroseeding, the protocol described in ‘Crystallization of Nucleic acids and proteins: A practical approach’ (text book, page no. 191-196) by Ducruix A. and Giege’ R. (1999) was used. Maintaining the humid environment around the source and recipient drops performed the seeding. Macroseeding is done under a dissecting microscope where the small amount of heat generated from the microscope stage light bulb may actually be slightly beneficial in increasing the humidity level around the drop. Seeds are washed in a slightly dissolving solution (reservoir solution having slightly low concentration of precipitant than the seed reservoir solution) to remove the top layer of protein of the seed; the top layer might contain possible defects. From the wash solution, each seed is then transferred to the protein-precipitant drop. The seeds of the scFvM12 and MUC1 peptide complex were obtained from the drop grown at 17°C in 4M sodium formate buffer. 2µL of MUC1 peptide and scFvM12 protein complex solution were mixed with 2µL of well buffer (4M sodium formate) and allowed to equilibrate at 17°C against 1 ml of well buffer. After 24 hrs, these drops were opened one at a time and seeded with few washed crystals by using a microwick.

II.2.10 Crystal mounting

II.2.10.1M12-V_{H} domain alone and in complex with MUC1

Crystals were morphologically examined and manipulated under a stereomicroscope (Leica). The thin plates of the M12-V_{H} domain were mounted on small diameter rayon loops after series of brief washing in reservoir solution [100mM MES pH 6.5, 10mM ZnSO_{4}, 25% (v/v) PEG 550MME, 4mM DTT, 0.04% (w/v) NaN_{3}] containing 5%, 10%, 15%, 20%, 25% and 30% (v/v) of glycerol as cryo-protectant. The crystal was transferred in approximately 10-20sec on goniometer head (45°) and immediately flash cooled at 100K using a nitrogen cryostream (Oxford Cryosystems Ltd, UK). The crystals remained frozen in liquid nitrogen for one month.

The crystals of M12-V_{H} in complex with MUC1 peptide (II.2.9.1.3) were washed and flash cooled using 40% (v/v) glycerol as cryo-protectant. These crystals remained frozen in liquid nitrogen for ten months before data collection.

II.2.10.2 scFvM12 antibody fragment

Crystals of scFvM12 (II.2.9.1.2) from the following conditions were flash cooled using PEG 400 (40% v/v) and glycerol (30% v/v) as cryo-protectant (II.2.10.1) and remained frozen in liquid nitrogen for approximately 13 months before testing with X-rays:
A. 100mM NaH$_2$Citrate pH 5.0, 1.2M ammonium sulfate and 2% v/v glycerol (88% v/v). Crystals started appearing after approximately one week at 25°C.

B. 100mM NaH$_2$Citrate pH 5.0, 1.2M ammonium sulfate and 10mM hydroxyectoine. Crystal started appearing after approximately one week at 25°C.

C. 100mM NaH$_2$Citrate pH 5.0, 1.2M ammonium sulfate and 1% (v/v) PEG 550MME. Crystal started appearing after approximately 15-20 days at 25°C.

D. 40mM NaH$_2$Citrate pH 5.0, 1.2M ammonium sulfate and 0.5mM ZnCl$_2$. Crystal started appearing after approximately 15-20 days at 25°C.

E. 40mM NaH$_2$Citrate pH 5.0, 1.2M ammonium sulfate and 0.5mM CdCl$_2$. Crystal started appearing after approximately 15-20 days at 25°C.

II.2.10.3 scFvM12 antibody fragment in complex with MUC1 peptide

Tiny crystals of the scFvM12 and MUC1 complex (II.2.9.1.4) were flash cooled using 40% (v/v) glycerol as cryo-protectant in the reservoir solution (3.5M sodium formate) as above (II.2.10.1). The crystals remained frozen for ten months before being analyzing on synchrotron beamline X13 (DESY, Hamburg).

II.2.11 Data collection and data processing

The frozen thin plate like M12-VH domain crystal (Figure III-14) was mounted on an X-ray generator (Center for Protein Engineering, University of Liege, Belgium). 125 images were collected on Mar345 image plate using X-ray radiation from a rotating copper anode (CuK$_\alpha$ radiation at 1.542Å, Rikagu, Japan) operated at 40kV and 100mA. The spindle was rotating for 1° with 3 oscillations for each image in 30 minutes and a crystal to film distance was 130cm. The crystal diffracted to a high resolution of 1.8Å. The diffraction data was collected using a single crystal.

The data for frozen crystal of M12-VH in complex with MUC1 peptide (Figure III-19) was collected at the synchrotron beam line X13 (λ$_{\max}$ - 0.802 Å, DESY, Hamburg), having a beam polarization of 0.900. In total, three crystals were tested on the beamline. The first crystal showed a twinning effect, while the diffraction pattern of the second crystal had distinct ice-rings, which indicates problems during flash cooling. The third crystal was found well ordered, defectless and data from this crystal was collected to 1.5Å resolution using
CCD detector (Bruker Nonius, Germany). The data collection was started at 55° and terminated at 196° using an oscillation angle of 0.5°. Completeness of the data insured by collecting 210 images. Due to overexposures for the low-resolution diffractions, 100 images at low resolution ~ 50Å were collected with a 2° oscillation angle. The crystal to detector distance was 130.00cm and the radiation dose was 1500 for the high-resolution data and 500 for the low-resolution data collection.

The diffraction data were indexed and integrated using DENZO and SCALEPACK (Otwinowski & Minor, 1997). The data from the different images were integrated separately and optimized for mosaicity, resolution, spot ellipticity and background. After optimization, the two datasets were merged together. The cell dimensions were refined by least squares methods. The collected data set for both the M12-V H fragment alone and its complex with MUC1 revealed that the crystal belongs to space group C2.

II.2.12 Phasing

II.2.12.1 Molecular replacement

II.2.12.1.1 M12-V H domain (1.8Å)

The tertiary structure of M12-V H domain was solved (by Dr. Kurt M.V. Hoffmann, refer to VI.4) through molecular replacement method by using the AMoRe program (Navaza, J., 1994) incorporated into the CCP4 program suite (II.1.8). The V H fragment of the theoretical model (PDB code - 1H0U, Hougs et al., 1999) was used as a search model for determining the phases for M12-V H (1.8Å).

II.2.12.1.2 M12-V H domain in complex with MUC1 peptide (1.5Å)

The M12-V H domain structure (1.8Å) solved in-house (by Dr. Hoffmann, 2001, RWTH-Aachen, Aachen, Germany) was used as a search model for determining phases for the M12-V H domain in complex with MUC1 peptide (1.5Å) through molecular replacement.

II.2.13 Model building and refinement

II.2.13.1 M12-V H domain (1.8Å)

The initial model obtained as a result of molecular replacement was subjected to reciprocal refinement and simulated annealing (Brünger et al., 1987) by using the CNS Program suite. The percentage of reflections used for the calculation of the R_free was 10.0%. The refinement was carried out within the resolution range of 8.0-2.0Å. Atoms were removed from the search model if they did not correspond to atoms in the M12-V H fragment or did not fit to the
electron density. After this, the model was subjected to initial rounds of simulated annealing, using harmonic restraints. Simulated annealing was performed at 2500K with a cooling rate of 25K. After rigid body refinement, 2F₀-Fₐ and F₀-Fₐ electron density maps contoured at 1σ were prepared and visualized in the ‘O’ program (Jones et al., 1991). The resulting model was not suitable for electron density interpretation; therefore, the phases of the model was further improved by using the PHASES program (Furey et al., 1997). The resulting model was visualized in ‘O’ and modified accordingly by using electron density as template. Refinement continued for an additional ten cycles with improvements of stereochemistry, following the addition of the water molecules. Refinement was terminated when no peaks greater than 3σ were observed in an F₀-Fₐ difference Fourier map.

II.2.13.2 M12-V₉ domain in complex with MUC1 peptide (1.5Å)

The model building and refinement for the M12-V₉ in complex with MUC1 peptide complex was carried out by using the automated protein model building and structure refinement package wARP/ARP (Perrakis et al., 1999). The wARP/ARP package is based on a unified approach to the structure solution process by combining electron density interpretation using the concept of a hybrid model, pattern recognition in an electron density map and maximum likelihood model parameter refinement. The major drawback of the package is its requirements to provide the diffraction data at 2.5Å or higher resolution. The automated model building was performed using the initial model of M12-V₉ in complex with MUC1 peptide complex as a result of molecular replacement. Initially, 5% of the reflections were kept aside to observe the progress of refinement. The autotraced chain of the protein was docked after 7 cycles of autobuilding. Autotracing was carried out by using the limited depth-first alpha carbon recognition algorithm. Difference Fourier maps 2m|F₀|-|D|Fₐ| and m|F₀|-|D|Fₐ| were used to guide manual fitting of the model in the program ‘O’ (Jones et al., 1991). After manually observing the partial model in ‘O’, the electron density for the MUC1 peptide was not observed; therefore, the resulting model is regarded, treated and presented as M12-V₉ domain structure in the following results and discussion chapters. The refinement was performed with the REFMAC program (Murshudov et al., 1997) within the CCP4 suite using the restrained maximum likelihood method with a bulk solvent correction. In the final set of refinement, all data were included. Water molecules are located automatically with the program wARP/ARP (Perrakis et al., 1999).
II.2.14 Structure validation

The overall quality of the structure was checked by using the program PROCHECK (Laskowski et al., 1993) and SFCHECK (Vaguine et al., 1999).

II.2.15 Structure interpretation

CDR residues were defined according to Kabat et al., (1991). Structural superpositions were performed with the Insight II software (Biosym Technologies/MSI, San Diego, CA, USA). The calculation of the accessible surface areas was performed by using the Connolly (1993) algorithm implemented in the Insight II software. The V\textsubscript{H} germline family assignment was performed by using the V-Base database (www.mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901). For the comparison of sequential and structural features of the antibody structures (< 2.5Å) having PDB codes: 1DLQ (Vetting et al., 2000), 1IGM (Fan et al., 1992), 1AQK (Faber et al., 1998), 1F2X (Decanniere et al., 2000), 1HCV (Spinelli et al., 1996), 1JTP (Decanniere et al., 2001), 1MEL (Desmyter et al., 1996) and 1VHP (Reichmann, L., 1996) were utilized.
III RESULTS

III.1 Bacterial expression of scFvM12

III.1.1 Small scale expression

The scFvM12 antibody expression vector was made according to Wong et al., (2001), using pSynI (see appendix) for the periplasmic bacterial expression. The recombinant plasmid was isolated and transformed into \textit{E. coli} BL21 (DE3) & TG1 strains. After identifying recombinant \textit{E. coli} TG1 colonies for optimum yield of scFvM12, the conditions for bacterial periplasmic expression of the protein was optimized in terms of temperature, pH, medium type, additives, IPTG concentration to obtain maximized yields (data not shown). The optimum expression of the protein was observed at 28°C in 2xTy medium supplemented with yeast extract (3g/L) and the pH (~7.0) of the medium regulated by the addition of the dipotassium hydrogen phosphate (3g/L). Induction was induced with 1mM IPTG. To obtain the higher yield of scFvM12, the best combination of parameters in shake flasks was found as pH 6.8, 28°C, 1mM IPTG (final concentration), and 4-5hrs of incubation after induction. The yield of the protein after purification was estimated to be 0.4-0.6 mg scFvM12 per liter of the culture medium.

III.1.2 Large scale expression

In general, the crystallographic study needs a fairly large amount of pure protein (at least 5mg), therefore, it was essential to scale-up the scFvM12 expression using fermentation technology. The large-scale bacterial expression of the scFvM12 antibody was based on the parameters optimized for shake flasks (III.2.1.1). For large-scale fermentation, a fixed-volume fed-batch high-cell-density fermentation (HCDF) strategy was selected because this is a good alternative for the production of large quantity of recombinant proteins in small volume of culture maintained at high density (O.D.600nm ~60).

For the HCDF fermentation of the antibody, synthetic modified minimal medium (II.1.5.II) was used. During the growth phase, the bacterial culture was maintained at 37°C. The fermenter was inoculated with 4% (v/v) \textit{E. coli} TG1 strain harboring the pSynI bacterial expression vector containing the scFvM12 gene and the ampicillin resistance gene. After inoculation, the culture remained in the latent phase for about 2-3hrs until an increase in the metabolic activity of the bacteria was indicated by enhanced oxygen consumption. The culture was grown to an O.D.600nm ~25 (Figure III-1). From 18hrs to 23hrs, a reduced O$_2$
requirement of the culture with a slight increase of the pH was observed which indicated the depletion of the carbon source. After 23hrs, additional carbon was fed in the form of a 50% glycerol solution. As soon as the feeding started, a sudden surge in O$_2$ requirement was observed. The initial feeding of the culture resulted into significant increase of the culture density up to ~50 within 6hrs. At O.D.600nm~52, the culture was induced by adding 1mM of IPTG (final concentration) and the temperature was reduced to 28°C. Half an hour after induction, the feed rate was changed to maintain the culture in a stationary phase. The continuous slow increase in the culture density after induction indicates the presence of suitable conditions for the culture growth. The rise of the pH just after induction was due to the culture adaptation after a long starvation.

![Graph showing HCDF fermentation profile of scFvM12](image)

**Figure III-1 HCDF fermentation profile of scFvM12.** scFvM12 expressed in modified minimal medium (II.1.5.II) at 28°C and induced with 1mM IPTG.

![Western blot analysis](image)

**Figure III-2 Western blot analysis (II.2.4.4) for scFvM12 expressed in bacterial periplasm and its leakage into the bacterial supernatant during HCDF (II.2.2.2).** scFvM12 was separated on 12% SDS-PAGE (II.2.4.3) and transferred to nitrocellulose membrane. The blot was developed with anti-his antibody (1:4000 dilution). M: Pre-stained molecular weight marker (10µL). Lanes 1 and 2 indicates that the scFvM12 protein is present in the supernatant (20µL) and in the pellet (1mg, wet weight) respectively.
The oxygenation of the culture was controlled automatically by the synchronization of the stirrer speed throughout the fermentation. Foaming was controlled by the addition of antifoaming agent. The culture was harvested 16 hrs after induction by centrifuging for 20min at 5000xg. The pellet was stored at –80°C while the supernatant was kept at 4°C for protein purification, as the expression of the scFvM12 is leaky from the periplasm. Figure III-2 indicates the presence of significant amounts of protein in the bacterial supernatant.

HCDF work was carried out in co-operation with Dr. Stephan Hellwig (Fraunhofer IME, RWTH, Aachen, Germany) as when the scFvM12 expression was needed.

### III.2 Purification of the scFvM12

Since the protein quality required for crystallization should be of highest purity (above 99%), three techniques namely affinity-, ion exchange and size-exclusion chromatography were used to purify the scFvM12.

#### III.2.1 Ni-NTA chromatography

The protein was extracted from the *E. coli* TG1 HCDF pellet by osmotic shock and loaded on a Ni-NTA column, where protein binding was based on the c-terminal His₆ tag in scFvM12. The binding of the protein to the affinity column occurs as a result of formation of a co-ordination complex between the histidine amino acids and immobilized transition metal ions such as Ni²⁺ and Zn²⁺. However, the His₆-tag binding strength is dependent on experimental conditions and can vary from protein to protein.

![Ni-NTA purification profile of scFvM12](image)

**Figure III-3A** Ni-NTA purification profile of scFvM12. Ni-NTA (II.2.3.1.2) loaded scFvM12 was eluted in 50mM Tris pH 8.0, 500mM NaCl, 500mM imidazole and 2mM benzamidine-HCl at a flow rate of 1ml/min by using 2ml of fast flow chelating sepharose resin column having immobilized Ni²⁺ ions.

![SDS-PAGE analysis of scFvM12 eluates](image)

**Figure III-3B** SDS-PAGE (II.2.4.3) analysis of scFvM12 eluates from the Ni-NTA column (II.2.3.1.2). Lane 1: Periplasmic extract, Lane 2: Flow-through, Lane 3: Wash, Lane 4-9 are the various fractions collected during 60, 70, 80, 85, 90 and 95 ml of elution volume of protein elution (20µL of each sample was loaded). M - Molecular weight standard (10µL of Mark 12).
The conditions for the affinity purification of the scFvM12 antibody were optimized (data not shown) and the protein purification was performed with step gradients of 5%, 8%, 10% and 100% of 500mM imidazole in the elution buffer (Figure III-3A). The purified protein was checked through SDS-PAGE (Figure III-3B). The eluted protein was still contaminated with low amounts of proteins of bacterial origin. These non-specific and unwanted proteins bind to the Ni-NTA column along with the desired proteins, presumably through surface-accessible histidines or other amino acids present in the native molecule, such as cysteine, aspartate, and glutamate (Janson and Ryden, 1989).

**III.2.2 Ion-exchange chromatography**

The Ni-NTA purified scFvM12 antibody was further purified through ion-exchange chromatography. The conditions for the binding of the scFvM12 to the ion-exchange column were optimized in terms of ionic strength and pH (data not shown).

The cation-exchange purification profile of scFvM12 on MonoS column at pH 4.4 shows the appearance of four peaks A, B, C, D respectively (Figure III-4A). The peaks B, C and D were found positive for scFvM12 after western blotting with anti-9E10 antibody (Figure III-4B).

The different peaks observed during the cation exchange purification were kept separately and the major peak C was loaded on a MonoQ column for anion-exchange purification. The MonoQ purification profile of scFvM12 antibody showed the presence of distinct peaks...
C & E and three shoulders A, B and D (Figure III-5A). The peak C was composed of multiple peaks, which indicated scFvM12 present in more than one form. The western-blot analysis with the anti-9E10 antibody indicated that the shoulders and the peaks belong to the scFvM12 antibody (Figure III-5B). The existence of these peaks could be explained on the basis of the presence of either different aggregated forms or different conformational forms of the protein.

![Graph showing anion-exchange profile of scFvM12](image)

**Figure III-5A Anion-exchange (II.2.3.1.4) elution profile scFvM12.** 20ml of scFvM12 was loaded on MonoQ HR 10/10 column and the bound scFvM12 was eluted with a linear gradient of NaCl at pH 8.0. Western blot analysis (Figure III-5B) of the different eluted peaks revealed that all of them belong to scFvM12.

![Western-blot analysis for different peaks](image)

**Figure III-5B Western-blot (II.2.4.4) analysis for different peaks (Figure III-5A) of scFvM12 eluted during anion-exchange purification.** The blot was developed with the anti-9E10 antibody (1:4000 dilution). Lane 1: Flowthrough, Lane 2: 100ml, Lane 3: 106ml, Lane 4: 108ml, Lane 5: 110ml, Lane 6: 117ml, Lane 7: 120ml. M: 10µL of pre-stained molecular weight standard. 20µL of each sample loaded on SDS-PAGE (12%).

The pooling of the fractions corresponding to different MonoQ purified peaks were avoided and each anion-exchange purified peak was subjected size-exclusion chromatography separately.

### III.2.3 Size-exclusion chromatography

The ion exchange purified protein fractions were further purified through size-exclusion chromatography using the isocratic elution method in low salt buffer (II.2.3.1.5). Although, the possibility of protein binding to the column matrix (Superdex) can not be ruled out, a low salt concentration in the gel purification buffer was chosen to keep the protein ready for the crystallization setups just after elution. The protein elutes as a single peak during gel filtration (Figure III-6A) and showed a single band on SDS-PAGE, which indicates that, a major fraction or the whole protein is intact. The integrity of the protein was also confirmed by...
western blotting (Figure III-6B). The yield of the protein after gel purification was 24mg per kg (wet weight) of HCDF pellet.

**Figure III-6A** Size-exclusion chromatography (II.2.3.1.5) elution profile for peak C of MonoQ purified (Figure III-5A) scFvM12. Anion exchange purified protein was concentrated (II.2.4.1) to an appropriate volume (~1ml) and loaded on a HiLoad superdex 75 16/60 column. scFvM12 was eluted 0.5ml/min flow rate. The column was calibrated with B (bovine serum globulin, 66 kDa), CA (carbonic anhydrase, 29 kDa), CY (cytochrome C, 12.4 kDa) and A (aprotinin, 6.5 kDa) as reference (Fluka).

**Figure III-6B** Western-blot (II.2.4.4) analysis of gel-purified scFvM12 (Figure III-6A). The blot was developed with anti-9E10 antibody (1:5000 dilution). Lane 1: 1µL scFvG19 (15mg/ml) as positive control, Lane 2: 2µL of gel-purified scFvM12 (8.5mg/ml). M - 10µL of pre-stained molecular weight standard.

### III.2.4 Iso-electric focusing and mass spectrometry

The identification of the purification tags through western blotting only confirms the

**Figure III-7** Iso-electric focusing (IEF, II.2.4.3) of gel-purified (Figure III–6A) scFvM12. M - 2µL of IEF Standard (pl 3-9, SERVA). scFvM12 (1µL of 13mg/ml) loaded on cathode (lane 1) and anode side (lane 2) separately. The gel was run according to the manufacturers instructions (Amasham Pharmacia Biotech).
integerity of the protein but it does not explain the existence of different peaks during ion-exchange chromatography. To reveal the presence of different conformational forms of the protein, the gel-purified protein was subjected to iso-electric focusing (IEF), which illustrates the pI of the protein and also helpful in illustrating the conformational heterogeneity of the protein. The IEF gel (Figure III-7) showed the presence of two major and two to three minor bands. The presence of the two distinct bands suggests the existence of protein in more than one conformational state. The IEF analysis indicates that scFvM12 is acidic in nature having a pI around 6.5 (theoretical pI–6.16).

The purity and the molecular weight of the gel-purified scFvM12 were determined by MALDI-TOF mass spectrometry (Figure III-8). The mass spectrum showed a distinct parent peak of 27.450 kDa, which closely corresponds to theoretical molecular weight of scFvM12 monomer (27.478 kDa).

```
Figure III-8 MALDI-TOF mass analysis (II.2.6) of gel-purified (Figure III-10A) scFvM12. 10µl of gel-purified scFvM12 (2mg/ml) in 5mM Tris, pH 8.0 was used for analysis.
```

The peak of 13729.25Da is due to the double protonation [M+2H]^{2+} of scFvM12 during measurement. The absence of the peaks corresponding to the degradation products in the MALDI spectrum indicates that the different forms observed in the IEF (Figure III-7) were based on the conformational heterogeneity.

scFvM12 were leaked into the fermentation supernatant was purified in the same manner as described for periplasmic protein (data not shown). The yield of the gel-purified protein was approximately 7.0 mg per liter of bacterial supernatant.
III.3 Crystallization of antibody fragments

III.3.1 Crystallization of scFvM12

Preliminary crystallization screening for the gel-purified scFvM12 (27.47kDa) was performed using the crystal screens I & II and ammonium sulphate grid screen (Hampton Research) at 4°C, 17°C and 25°C. The initial conditions resulted either in microcrystals or microneedles conditions and require extensive optimization to study the effect of different variables on the crystal quality. The variables of the initial conditions listed in II.2.9.1.3, II.2.9.2, II.2.9.4 and II.2.9.5 were varied either alone or in combination. The effect of these alterations did not alter the rate of nucleation and the crystal quality significantly.

The intact scFvM12 was crystallized in the form of small rods using ammonium sulphate as precipitating agent at 17°C (Figure III-9A). The integrity of the protein in the crystal form was checked through silver staining (Figure III-9B). The silver staining revealed that the protein was crystallized without any degradation.

![Figure III-9A](image)

**Figure III-9A** The rod shaped crystals of full size scFvM12 (27.45kDa) antibody grown in 100mM sodium citrate pH5.0, 1.2M ammonium sulfate, 4mM DTT and 0.04% (w/v) NaNO₃ at 17°C (II.2.91.2). The crystals appeared within 5-7 days as small needles and grew to the size suitable for diffraction data collection within 6-8months.

![Figure III-9B](image)

**Figure III-9B** Study of integrity of scFvM12 in crystal form by using the silver stained (II.2.4.5) SDS-PAGE (12%). M - Molecular weight standard (10µL of Mark12). Lane 1: 0.3 µL of gel-purified scFvM12 (8.5mg/ml) was used as positive control. Lane 2: Solubilized scFvM12 crystals.

The scFvM12 crystals diffracted very poorly (III.5.1). The probable reason could be the existence of profound conformational heterogeneity inside the crystallization drop.

The functionality of the gel-purified scFvM12 (peak C of MonoQ pure scFvM12) was tested through FACS analysis (Figure III-10) to find out the cleavage of the linker after incubating the protein for an extended period of time at 4°C. The functionality of the scFvM12 was tested in 20mM Tris pH8.5, 20mM NaCl, 1.0mM DTT and 0.5mM EDTA.
The active nature of the protein indicates towards the integrity of the scFvM12 with no linker cleavage.

Based on functionality analysis of the scFvM12 it seems that the conformational heterogeneity is the major reason for the appearance of disordered scFvM12 crystals. Trimming the highly flexible regions such as purification tags can reduce the conformational heterogeneity. Using very low concentration of proteases can facilitate the trimming.

**III.3.2 Limited proteolysis of scFvM12**

To reduce the conformational heterogeneity (Figure III-7) of the scFvM12, limited proteolysis was performed. The observed conformational heterogeneity may be either due to the linker joining the two domains or the purification tags of the scFvM12. Limited *in-vitro* proteolysis of scFvM12 was performed using three non-specific and broad-spectrum proteases. The scFvM12 digestion profile (Figure III-11) revealed that the scFvM12 is a good substrate for proteinaseK. ProteinaseK digests scFvM12 completely within 6hrs while subtilisin and thermolysin seems to be equal in their proteolytic behavior and degrades only a minor fraction of scFvM12. No band accumulation proportional to the disappearance of undigested scFvM12 protein was observed.

It can be predicted that the entire process is progressing in such a manner that firstly the protein is breaking down into its components namely the V\textsubscript{H} or V\textsubscript{L}, linker, tags and then or simultaneously these components are degrading in order of preferences.
The scFvM12 degradation rate with subtilisin and thermolysin is much slower in comparison to proteinaseK; therefore, these two proteases can be helpful in reducing the conformational heterogeneity of the protein. Irrespective of protease, there is an appearance of a single band (~12kDa) during the digestion study. This band might represent either the V\textsubscript{H} or V\textsubscript{L} domain of the scFvM12 antibody.

Since, the scFvM12 was constructed in V\textsubscript{H}- (G\textsubscript{4}S\textsubscript{3})-V\textsubscript{L}-cmyc-his\textsubscript{6} configuration, western blot analysis was performed to access the type of the fragment(s) produced as a result of scFvM12 digestion with proteases subtilisin Carlsberg and thermolysin. No band (Figure III-12) was observed after the cleavage of scFvM12 with subtilisin at different intervals of time, which indicate either the resulting fragment is V\textsubscript{H} domain or the V\textsubscript{L} domain whose
associated the marker tags had been cleaved and therefore, made it impossible to detect. Similar study performed by using thermolysin (data not shown).

The single prominent band (~12kDa) generated as result of subtilisin and thermolysin limited proteolysis of scFvM12 was extracted from the gel and subjected to ESI-MS/MS analysis. The sequence determination of the peptides obtained after the MS analysis indicates that the bands generated, as a result of the two different protease digests, belongs to the V\textsubscript{H} domain of the scFvM12 (data not shown).

Since, the scFvM12 was constructed by assembling the V\textsubscript{H} and V\textsubscript{L} domains from two different human phage libraries (Sheets \textit{et al.}, 1998) and the sequence analysis revealed that the V\textsubscript{L} gene sequence of the scFvM12 is nearly identical to one of the germline V-genes while the V\textsubscript{H} gene has undergone multiple mutations; the V\textsubscript{H} domain should dominantly contribute to the affinity of the scFvM12 towards the MUC1 antigen. Therefore, the crystallization of the V\textsubscript{H} domain of scFvM12 was performed through limited proteolysis.

\section*{III.3.3 M12-V\textsubscript{H} domain crystallization}

The cleavage activity of the proteases was found as a function of enzyme to substrate ratio (Figure III-13).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{SDS-PAGE (18\%, II.2.4.3) showing the effect of subtilisin concentration at 22°C on the cleavage pattern of scFvM12. 5\textmu L of digested sample was loaded in order of 1:100, 1:500 and 1:1000 molar ratios (protease:protein) from left to right for each time interval. M: Molecular weight marker (10\textmu L), Lane 1: 2\mu L of scFvM12 (8mg/ml) used as positive control. Lanes 2, 3 & 4: digested for 0 min, Lanes 5, 6 & 7: digested for 5 min, Lanes 8, 9 &10: digested for 60 min, Lanes 11, 12 &13: digested for 6 hrs and Lane 14: digested for 24hrs.}
\end{figure}

The ratio was chosen carefully to generate the isolated M12-V\textsubscript{H} fragment in abundance with in the crystallization drop so that the drop get supersaturated and forces the fragments to crystallize even if they are susceptible to further proteolysis. Initially, a protein to protease molar ratio of 100:1 and 50:1 was selected for subtilisin digestion.

Similar digestion study was performed to optimize the molar ratio of thermolysin and scFvM12 (data not shown).
III.3.3.1 By subtilisin treatment of scFvM12

Crystallization trials were set up at 4°C after adding the subtilisin Carlsberg protease to scFvM12 solution in 1:100 molar ratio (enzyme:protein) by using the crystal screen I and II (Hampton Research). Numerous star shaped crystals appeared in 3-5 days in 100mM MES pH6.5, 25% (v/v) PEG 550MME, 10mM zinc sulfate heptahydrate, 4mM DTT and 0.04% (w/v) NaN₃ buffer. The condition was found reproducible both at 4°C and 17°C. The optimization of the initial condition was performed as follows:

A. Effect of metal ions (II.2.9.1.1.A): It was observed that the absence of Zn²⁺ ion results in micro-crystals while the replacement of zinc ions with other divalent metal ions, such as lithium, magnesium, cobalt, cadmium certainly increases the crystal dimension after long time (approx. 30 days) but mostly resulting in the defective crystals. However, the replacement of SO₄⁰⁻ ions with chloride ions i.e. ZnCl₂ used in place of ZnSO₄, the chloride ions shows some synergistic effect with zinc ions in improving the crystal quality but the crystal growth rate was substantially slower (30-40 days).

B. Protein concentration effect (4°C, II.2.9.1.1.A): Change of the protein to precipitating buffer ratio of 1:1 in hanging drops to 1:2 and 2:1 did not affect the morphological symmetry of the crystals. The only noticeable difference was the number and the rate of appearance of the crystals. They are less in number and slower in growth in the 1:2 ratio by using a 2:1 ratio (v/v), grown within 30hrs, which is expected, due to the difference in protein concentration.

C. Effect of pH and temperature (II.2.9.1.1.A): The morphology of the crystals varies with pH and temperature. The change in pH (at 4°C) towards more acidic conditions resulted into phase separation, skin or aggregates of tiny crystals while a shift in pH (at 4°C & 17°C) towards the basic conditions resulted in to the formation of either isolated or aggregated thin plates.

D. Effect of polyethylene glycol (PEG): PEG 550MME concentration from 5 to 15% (v/v) and above 40% (v/v) produced either amorphous or gelatinous precipitates (II.2.9.1.1.A). The rise in PEG 550MME concentration above 25 % up to 40% (v/v) increases the rate of nucleation and resulted in multiple aggregates of star shaped crystals.

E. Incubation time effect: The most critical point in the development of the crystals was the incubation time of the protein with subtilisin before setting up the trials. It was observed that crystals developed only when the subtilisin
mixed with protein solution just prior to setting the hanging drops. Pre-
incubation of the protein with the enzyme resulted only either in precipitation
or skin in the drops rather than crystals of any shape. The most probable
reason for no crystal growth was the complete digestion of the protein by
subtilisin as observed in the protein digestion study (Figure III-19).

F. Effect of additives (at 4°C):
   i. Glycerol (10% v/v) resulted in the formation of isolated thin plates but
      increases the crystal growth time by a factor of ten.
   ii. n-octyl-β-D-Glucopyranoside (20mM) produced large number of
       benzene shaped small thin plates in 25-30 days.

G. Sodium malonate screen: Recently, sodium malonate was reported as one of
the best crystallizing substance for a number of different proteins (McPherson
A, 2001). The trials for the subtilisin treated scFvM12 in 10%-100%
saturation range of malonate were set up at pH6.5, 7.5 and 8.5 at 17°C at 10%
intervals (II.2.9.1.1.A). After three weeks, the 50% saturated sodium malonate
condition at pH 8.5 resulted in the formation of club shaped crystals growing
lengthwise having broom like proximal end and sharp distal end. After 5-6
weeks, the 60% saturated sodium malonate at pH 7.5 resulted in the crystals
similar in morphology as at pH 8.5.

M12-VH domain crystals were obtained after optimizing the temperature condition. The
optimization of the other parameters did not affect the crystallization. After optimization, thin

Figure III-14 M12-VH domain crystals (150 x 50µm; after 5-7 days) grown at 17°C in
0.04% (w/v) NaN₃ (II.2.8). The protease was added to the scFvM12 solution (20mM Tris, pH7.5) in 1:100 molar ratio (protein:protease) just before setting the hanging drops. The protein to well buffer dilution was 1:2.
plates like crystals appeared at 17°C by using polyethylene glycol as precipitating agent (Figure III-14). The collected crystals were washed several times in the reservoir solution and dissolved overnight at 4°C in 50mM Tris, pH8.0.

The dissolved crystals were tested to analyse the integrity of the marker tags. Theoretically, the tags are a flexible portion of the antibody and therefore, most susceptible towards the proteolytic digestion. Furthermore, they are present at the C-terminal end of the scFvM12 antibody, therefore, prone to digestion. The western blot (Figure III-15) shows the absence of the tags upon proteolytic digestion.

![Figure III-15](image)

Figure III-15 Western-blot (II.2.4.4, 15% SDS-PAGE) for solubilized V\text{H} fragment crystals developed with anti-9E10 antibody (II.1.6, 1:5000 dilution, A) and anti-His antibody (II.1.6, 1:5000 dilution, B) respectively. M: 10\muL of pre-stained molecular weight standard. Lane 1: 2\muL of scFvM12 (8mg/ml) used as positive control, Lane 2: 5 \muL of dissolved V\text{H} fragment crystal. The crystals were collected and dissolved in 50mM Tris pH8.0 and the solution kept overnight at 4°C before western blotting. The difference in the marker position is due to the difference in gel running.

![Figure III-16](image)

Figure III-16 MALDI-TOF mass spectrum (II.2.6) of the solubilized crystals obtained after in-drop digestion of scFvM12 antibody fragment with subtilisin Carlsberg at 4°C and 17°C. The crystals were washed in reservoir solution and dissolved in deionized water. 2\muL of this solution (~2mg/ml) was diluted to 10\muL with deionized water. 2\muL of diluted solution mixed with MALDI matrix solution containing 0.1% (v/v) TFA (trifluoroacetic acid) in 1:1 ratio and loaded on the MALDI target.
Mass spectrometry was used to assess the identity of the crystals generated as result of subtilisin digestion of scFvM12. The mass spectrum of the solubilized crystals (Figure III-16) indicated the clear peak of 12.346kDa, which is in close correspondence with the theoretical mass (12.348kDa) of the M12-VH domain. Therefore, the mass analysis suggests that the subtilisin protease cleavage of scFvM12 produced the isolated M12-VH domain crystals.

The sequence of the solubilized M12-VH domain crystals was analyzed by using electrospray ionization tandem mass spectrometry (ESI-MS/MS; Q-tof-2™, Micromass UK Limited, UK) (Figure III-17). The analysed sequence belongs to M12-VH domain.

**Figure III-17 ESI-MS/MS spectrum (II.2.6) of the N-terminal peptide generated after tryptic digestion of the solubilized crystals.** bMax indicates the b-ions, representing the N-terminal peptide sequence of the VH domain of the scFvM12 antibody fragment.

**III.3.3.2 By thermolysin treatment of scFvM12**

Trials were setup using the mixture of the scFvM12 and thermolysin (50:1 molar ratio) produced microneedles after 10 days. To improve the quality of the crystals, the variables listed in II.2.9.1.1 were altered. The change of the protein to protease molar ratio from 50:1 to 200:1 substantially increased the duration of crystal appearance to 5-6 months and resulted in compact but disordered crystals (Figure III-18). These crystals underwent dissolution and complete cleavage after storing for some time (~1month). These solubilized crystals were subjected to ESI-MS/MS mass spectrometry to find out which scFvM12 domain has been crystallized. After sequence analysis (data not shown), it was observed that the obtained crystals belong to the M12-VH domain.
III.3.4 Crystallization of M12-V\textsubscript{H} in complex with MUC1 antigen

The crystallization of the M12-V\textsubscript{H} domain in complex with MUC1 peptide was performed by the soaking method (II.2.9.1.3). Firstly, M12-V\textsubscript{H} domain crystals were generated through protease treatment at 4°C by adding the subtilisin Carlsberg to scFvM12 in 1:250 molar ratio. The compact and nicely looking crystals appeared in approximately 12 months (Figure III-19). The MUC1 peptide was added to the drops containing the M12-V\textsubscript{H} crystals and the crystals remained soaked for another 2 months. During the incubation of the crystals with peptide, the crystals were frequently examined for any morphological change. No morphological changes could be observed. After 2 months, the crystals were flash cooled using a nitrogen cryostream.

Figure III-19 Crystal of M12-V\textsubscript{H} domain (~180x60x60\textmu m) soaked with the MUC1 peptide antigen (II.2.9.1.3).
III.3.5 Co-crystallization of scFvM12 in complex with MUC1 peptide

The scFvM12 has affinity towards the hydrophilic epitope ‘PDTRP’ of the MUC1 mucin (Wong et al. 2001). Until now, several structures of either murine or humanized Fab antibody fragments directed against MUC1 have been solved (Dokurno et al., 1998 and Banfield et al., 1997). Using co-crystallization, crystallization of scFvM12 antibody in complex with MUC1 peptide antigen was achieved.

During co-crystallization, protein of interest is incubated with its ligand and the resulting complex subjected to crystallization either directly or pre-purified through size-exclusion chromatography. The scFvM12 was incubated with the MUC1 peptide overnight at 4°C and the mixture directly subjected to crystallization. Microcrystals (II.2.9.1.4) was observed in few preliminary screens performed by using the ‘Crystal Screen I & II’ (Hampton Research). Tiny but good looking crystals were collected and dissolved in non-denaturating loading dye buffer to confirm the complex crystal formation by using the native gel electrophoresis (Figure III-20).

![Native gel electrophoresis of scFvM12 and MUC1 peptide complex crystals](image)

**Figure III-20** Native gel electrophoresis (II.2.4.6) of scFvM12 and MUC1 peptide complex crystals (II.2.9.1.4). M: 3µL of scFvM12 (8mg/ml) used as marker. Lanes 1, 2 & 3 refer to solubilized crystals taken from the conditions respectively. The collected crystals were centrifuged at 500xg for 4min and the crystal pellet dissolved in 15µL of 50mM Tris pH 8.5. 8µL of solubilized protein were mixed with 5µL of 2x non-denaturing loading dye for native gel electrophoresis.

![Tiny and thin plate like crystals of the scFvM12 in complex with MUC1 peptide](image)

**Figure III-21** Tiny and thin plate like crystals of the scFvM12 in complex with MUC1 peptide (II.2.9.1.4). The scFvM12 and the MUC1 peptide were used in a 1:5 molar ratio in 3.5M sodium formate at 17°C.
The variables of the initial condition listed in II.2.9.1.4, II.2.9.1, II.2.9.3 and II.2.9.4 were varied either alone or in combination. The effect of these alterations did not prove successful in increasing the size of the complex crystals.

The pyramidal tiny thin plate-like crystals were observed in 3.5 M sodium formate buffer. Initially, the crystals start appearing after 40-60 days (Figure III-21) but after the optimization of the initial condition, the duration of crystal appearance has been reduced to 10 days.

### III.4 Data Collection

#### III.4.1 scFvM12

The rod shaped crystals (Figure III-9A) were flash cooled and tested for diffraction on a Rigaku X-ray generator (Center for Protein Engineering, University of Liege). The observed diffraction pattern of the crystal was too poor to collect the dataset. Therefore, to collect a good quality dataset, the crystal growing conditions and the cryo-protectant type for the crystals had to be optimized. The probable reason for the poor diffraction of the scFvM12 was the conformational heterogeneity (Figure III-13) and therefore the resulting disorder inside the crystal. The disordered crystals usually have either poor diffraction or they do not diffract at all.

#### III.4.2 scFvM12 in complex with MUC1 peptide

The tiny crystals (Figure III-21) of the complex were too small for data collection on a laboratory scale X-ray generator; therefore, these crystals were tested on synchrotron beamline X13 (EMBL outstation, DESY, Hamburg, Germany). These frozen crystals failed to diffract even using the intense X-ray beam. However, few crystals showed very poor diffraction, which was not sufficient to record. The most probable reason could be their small size or the internal disorder.

#### III.4.3 M12-VH domain

The M12-VH domain dataset was first collected on an in-house X-ray generator and the structure was solved at 1.8Å resolution with unit cell parameters $a = 72.01$ Å, $b = 38.35$ Å, $c = 37.57$ Å, $\alpha = \gamma = 90$, $\beta = 109.746$ and R-factors as $R_{\text{cryst}} 23.54$ (%) & $R_{\text{free}} 28.69$ (%). The refined model lacks the residues 40-43, 52-56, 93-107 and 112-113. The structure has 80 water molecules. This structure lacks some of the CDR region residues, which are responsible for antigen binding.
Meanwhile, improved and very compact M12-V$_{H}$ crystals were grown (II.2.10.1). The frozen crystal stored in liquid nitrogen was mounted on a synchrotron beamline X13 (EMBL outstation Hamburg, DESY, Germany). In total three frozen crystals were tested. The first crystal was discarded due to the presence of ice-rings in the diffraction image. Ice-rings indicate the formation of ice during flash cooling of the protein crystal. Twinning was observed for the second crystal probably as a result of the simultaneous growth of the two crystals from the same nucleus (data not shown). The third crystal was well-ordered and diffracted to 1.5Å resolution (II.2.11) (Figure III-22). Overloading was tolerated for a first

![Figure III-22 Single diffraction image for the M12-V$_{H}$ domain collected (II.2.11) at synchrotron beamline X13 (DESY, Hamburg, Germany).](image)

**Table III-1 Data collection statistics for M12-V$_{H}$ domain at 1.5Å resolution.** $R_{\text{merge}}$ is defined in legend of Table III-2.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit-cell length (Å)</td>
<td>$a = 71.336, b = 37.965, c = 37.187$</td>
</tr>
<tr>
<td>Unit-cell angles (°)</td>
<td>$\alpha = \gamma = 90, \beta = 109.674$</td>
</tr>
<tr>
<td>Mathew’s coefficients (Å$^3$ Da$^{-1}$)</td>
<td>1.92</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>35.94</td>
</tr>
<tr>
<td>Unit cell volume (Å$^3$)</td>
<td>94832.64</td>
</tr>
<tr>
<td>Number of molecules in the asymmetric unit</td>
<td>1</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>34592</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>13739</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.802</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
<td>1.5</td>
</tr>
<tr>
<td>Highest resolution shell (Å)</td>
<td>1.53 - 1.50</td>
</tr>
<tr>
<td>Observed completeness</td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>(90.7%)</td>
</tr>
<tr>
<td>high-resolution shell</td>
<td>(92.9%)</td>
</tr>
<tr>
<td>$R_{\text{merge}}$</td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>(0.047)</td>
</tr>
<tr>
<td>high-resolution shell</td>
<td>(0.158)</td>
</tr>
</tbody>
</table>
dataset to collect the maximum number of unique reflections. In order to increase the completeness of the data at low resolution, a second set of dataset was collected at longer exposure times resulting in no overloads. This data was collected using a rotation angle of 2° in order to speedup data collection. No decay of the crystal was observed during data collection as a result of intense x-ray beam exposure.

The processing of the data with DENZO/SCALEPACK indicated that the obtained data belonged to the space group C2. The space group was confirmed by the intensity pattern of the data and the absence of systematic absences for this space group. After processing, the data was found to exhibit good data statistics (Table III-1). The data was found to have a mosaicity of 0.8.

### III.5 Quality of M12-V_{H} domain dataset

The overall quality of the high-resolution dataset was evaluated by using the standard data quality indicators calculated per resolution shell. The overall data quality was excellent.

#### Table III-2 Data statistics breakdown with resolution for M12-V_{H} domain dataset (1.5Å).
The values correspond to the X-ray wavelength (0.802Å). *R_mer* is defined as $100\sum_{hkl}|<I> - I_{n}|/\sum_{hkl}I_{n}$, which is summed over all reflections, where $<I>$ is the mean intensity of the reflection hkl and $I_{n}$ is the intensity of the nth observation of a reflection hkl. *I/σ* is defined as $<I>$ divided by the average error on I.

<table>
<thead>
<tr>
<th>Resolution shells (Å)</th>
<th>Reflections in each shell</th>
<th>Chi²</th>
<th>$R_{\text{merge}}$ (%)</th>
<th>I/σ*</th>
<th>Completeness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.00 - 4.07</td>
<td>688</td>
<td>1.185</td>
<td>0.036</td>
<td>26.01</td>
<td>86.5</td>
</tr>
<tr>
<td>4.07 - 3.23</td>
<td>755</td>
<td>1.202</td>
<td>0.034</td>
<td>33.00</td>
<td>97.2</td>
</tr>
<tr>
<td>3.23 - 2.82</td>
<td>759</td>
<td>1.067</td>
<td>0.034</td>
<td>36.79</td>
<td>99.9</td>
</tr>
<tr>
<td>2.82 - 2.56</td>
<td>770</td>
<td>1.141</td>
<td>0.037</td>
<td>36.60</td>
<td>100.0</td>
</tr>
<tr>
<td>2.56 - 2.38</td>
<td>749</td>
<td>0.944</td>
<td>0.043</td>
<td>31.05</td>
<td>100.0</td>
</tr>
<tr>
<td>2.38 - 2.24</td>
<td>758</td>
<td>1.025</td>
<td>0.046</td>
<td>30.35</td>
<td>100.0</td>
</tr>
<tr>
<td>2.24 - 2.13</td>
<td>761</td>
<td>0.986</td>
<td>0.046</td>
<td>29.72</td>
<td>100.0</td>
</tr>
<tr>
<td>2.13 - 2.04</td>
<td>758</td>
<td>1.154</td>
<td>0.053</td>
<td>28.13</td>
<td>100.0</td>
</tr>
<tr>
<td>2.04 - 1.96</td>
<td>764</td>
<td>0.952</td>
<td>0.055</td>
<td>24.59</td>
<td>100.0</td>
</tr>
<tr>
<td>1.96 - 1.89</td>
<td>753</td>
<td>1.092</td>
<td>0.067</td>
<td>21.78</td>
<td>100.0</td>
</tr>
<tr>
<td>1.89 - 1.83</td>
<td>748</td>
<td>0.972</td>
<td>0.076</td>
<td>20.01</td>
<td>100.0</td>
</tr>
<tr>
<td>1.83 - 1.78</td>
<td>749</td>
<td>0.831</td>
<td>0.084</td>
<td>17.34</td>
<td>100.0</td>
</tr>
<tr>
<td>1.78 - 1.73</td>
<td>756</td>
<td>0.792</td>
<td>0.094</td>
<td>15.98</td>
<td>100.0</td>
</tr>
<tr>
<td>1.73 - 1.69</td>
<td>746</td>
<td>0.735</td>
<td>0.113</td>
<td>14.08</td>
<td>100.0</td>
</tr>
<tr>
<td>1.69 - 1.65</td>
<td>761</td>
<td>0.767</td>
<td>0.127</td>
<td>13.15</td>
<td>100.0</td>
</tr>
<tr>
<td>1.65 - 1.62</td>
<td>740</td>
<td>0.773</td>
<td>0.150</td>
<td>11.31</td>
<td>100.0</td>
</tr>
<tr>
<td>1.62 - 1.58</td>
<td>767</td>
<td>0.766</td>
<td>0.158</td>
<td>10.72</td>
<td>100.0</td>
</tr>
<tr>
<td>1.58 - 1.55</td>
<td>736</td>
<td>0.745</td>
<td>0.180</td>
<td>9.71</td>
<td>100.0</td>
</tr>
<tr>
<td>1.55 - 1.53</td>
<td>754</td>
<td>0.656</td>
<td>0.188</td>
<td>8.61</td>
<td>100.0</td>
</tr>
<tr>
<td>1.53 - 1.50</td>
<td>745</td>
<td>0.651</td>
<td>0.214</td>
<td>7.26</td>
<td>99.5</td>
</tr>
<tr>
<td>All reflections</td>
<td>15017</td>
<td>0.917</td>
<td>0.047</td>
<td>27.70</td>
<td>99.1</td>
</tr>
</tbody>
</table>
III-RESULTS

In most of the resolution shells the data completeness was 100% or very close to it. The data collected separately for high and low resolution terms (II.2.11) were merged and the Wilson plot for the merged data was observed with no regions of irregularity.

III.6 Model building and refinement for the M12-VH domain

Initial phase determination was performed through molecular replacement by using AMoRe (Navaza, J., 1994). The M12-VH domain structure resolved at 1.8Å was used as a search model (see appendix).

The Matthew’s coefficient (VM) for M12-VH domain crystal having the space group C2 found to be 1.92 suggesting the presence of the single molecule in the asymmetric unit of the unit cell. The solvent content of the crystal calculated as 35.94%.

After obtaining the phase information through molecular replacement, the model building for the M12-VH domain was performed through wARP/ARP package (Perrakis et al., 1999). After the whole refinement, the refined model provided the statistics listed in Table III-3. The side chain of the amino acid is manually fitted to the electron density by using the program ‘O’ in between refinement cycles. During refinement the geometry of the model was examined by using the program ‘PROCHECK’. After the R-factor ceases to

<table>
<thead>
<tr>
<th>Resolution range</th>
<th>20.00-1.50Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of unique reflections</td>
<td>14358</td>
</tr>
<tr>
<td>Number of atoms refined</td>
<td>1024</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td></td>
</tr>
<tr>
<td>main-chain</td>
<td>16.67</td>
</tr>
<tr>
<td>side-chain</td>
<td>21.60</td>
</tr>
<tr>
<td>Observation to parameter ratio</td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>3.72</td>
</tr>
<tr>
<td>final</td>
<td>2.99</td>
</tr>
<tr>
<td>R-factor (%)</td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>33.7</td>
</tr>
<tr>
<td>final</td>
<td>15.8</td>
</tr>
<tr>
<td>$R_{free}$ factor (%)</td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>32.7</td>
</tr>
<tr>
<td>final</td>
<td>19.7</td>
</tr>
<tr>
<td>Connectivity index</td>
<td>0.96</td>
</tr>
<tr>
<td>Overall figure of merit</td>
<td>0.8810</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>0.973</td>
</tr>
<tr>
<td>free</td>
<td>0.947</td>
</tr>
<tr>
<td>r.m.s. deviation</td>
<td></td>
</tr>
<tr>
<td>bond distances (Å)</td>
<td>0.018</td>
</tr>
<tr>
<td>bond angles (°)</td>
<td>1.735°</td>
</tr>
<tr>
<td>torsion angles (°)</td>
<td>6.651°</td>
</tr>
</tbody>
</table>

Table III-3 Summary of structure refinement statistics for M12-VH domain (1.5Å).
improve further, the refinement cycle stopped. 141 water molecules were added to the refined model along with a single polyethylene glycol molecule.

### III.7 Overall structure of the M12-V<sub>H</sub> domain

The M12-V<sub>H</sub> gene belongs to the human gene family subgroup 3 (DP47) and locus 3-11 (Tomlinson et al. 1992). The V<sub>H</sub>3 subgroup is the germline family, which is most frequently observed in phage-display libraries based on human genes (Sheets et al., 1998). In overall appearance, the M12-V<sub>H</sub> has a main-chain resemblance to other human V<sub>H</sub> fragments. The residue numbering scheme and CDR definitions follow the standard Kabat conventions (Kabat et al., 1991). The final refined model of M12-V<sub>H</sub> fragment has R<sub>cryst</sub> of 15.8% and R<sub>free</sub> of 19.7% within the resolution range of 19.43-1.50Å. The structure has good stereogeometry, as expected at this resolution, none of the residues is outside of the allowed regions in the Ramachandran plot (Figure III-36) and it meets or exceeds all main chain and side chain tests of PROCHECK (Laskowski et al., 1993). The r.m.s. deviation for bond angles, and torsion angles is with in the acceptable range (Table III-3). The Luzzati mean coordinate error for the structure is 0.20Å (Luzzati, P. V., 1952). The σ<sub>A</sub> error estimate is 0.293 Å (Read, R. J., 1986). The r.m.s. deviation is 2.54778°. The average B-factor for the structure is 19.39Å<sup>2</sup>. The structure had poor electron density for the side chains of the residues 40-43 due to an apparent conformational disorder. The electron density for the rest of the structure is clearly visible and well modeled. In M12-VH domain, the electron density for three atoms Arg19 (NH1), Gln39 (NE2), Lys75 (NZ) was absent and distinct conformations were observed for residues Ser7, 12, 17, 21, 49, 50, 59, 77, 112 (OG) and Val12, 37(CG1, CG2) in the final structure model. The structure has 1024 atoms in total and 141 water molecules and a single molecule of PEG500. The structure has been deposited to the Protein Data Base (PDB) with an accession code of ‘1T2J’.

The M12-V<sub>H</sub> domain displays the well-characterized immunoglobulin fold, in which V<sub>H</sub> consists of two twisted, antiparallel β-sheets packed tightly against each other. Like many other V<sub>H</sub> structures, this domain also has the polypeptide backbone segments of the CDR loops of the heavy chains swept away and flattened against the globular body of the V<sub>H</sub> domain. The net effect is the production of a broad surface suitable for the antigen binding. The CDR-H3 loop contains series of hydrophobic side chains (TrpH96, MetH99, ValH102), projecting directly into the solvent. Leu-H11 and Met-H99 are the most accessible amino acids of the M12-V<sub>H</sub> domain with on accessibility surface area greater than 30%.

The major portion of the hydropathic plot (Figure III-23) has negative value, which indicates that most of the residues are hydrophilic in nature, while some of the residues
especially in the CDR’s are strongly hydrophobic, as indicated by their positive values in the plot. Depending upon the overall theoretical hydropathic values, it can be predicted that the isolated protein will be more water soluble with reduced chances of undergoing precipitation or aggregation as observed in most of the isolated V_{H} domains.

![Hydrophobicity plot](image)

**Figure III-23** Average hydrophobicity of the M12-V_{H} domain residues calculated by using the ‘Hydropathy’ program.

The overall molecular surface area for M12-V_{H} is 6443.78Å^{2}. The figure III-30 indicates the surface charge of the M12-V_{H} domain. The three CDR regions indicate the occurrence of non-polar residues, which are required for the maximum affinity of the antibody towards binding of the PDTRP epitope.

![Surface charge potential](image)

**Figure III-24** Surface charge potential of the M12-V_{H} (1.5Å). The numbers indicate residues involved in the V_{H}/V_{L} interface. The red color shows the acidic amino acids (negative) while the blue color represents the basic amino acids (positive) and white region represent a neutral surface charge. The figure was prepared by using the ‘GRASP’ program (Nicholls et al., 1991).
III.8 Structure description of the M12-V$_H$ domain

The M12-V$_H$ fragment structure displayed the typical immunoglobulin fold. The immunoglobulin fold (Ig fold) constitutes a class of $\beta$-proteins presenting a sandwich architecture where the strands, distributed in two sheets, are connected in a typical way.

![Figure III-25 M12-V$_H$ domain structure (1.5Å).](image)

The domain possesses 10 strands, which are labeled according to Halaby et al., (1998). The N and C represent the N and c-terminus of the protein. CDR regions are labeled according to Kabat et al., (1979). The domain exhibits the typical beta-sandwich structure, which is characteristic for immunoglobulins.

The connectivity of the strands is used to differentiate between the Ig fold and other similar Greek Key $\beta$-architecture. The M12-V$_H$ domain has a V-domain type $\beta$–strand topology with 10-strands constituting the two beta sheets (Figure III-25).

The nomenclature for the topologically identical V-type immunoglobulins is used (Williams et al., 1988). The $\beta$-sandwich is subdivided into a four-stranded $\beta$-sheet consisting of strands D, B, E, A and a six-stranded $\beta$-sheets formed by strands C, C’, C’’, F, G and A’.

```
QVQLQESGGG  LVQPGGSLRL  SCAASGFTFS  NSAMSWV  RQA  PGKGLEWVSS
|     |     |     | CDR-H1     |     |
| 10  | 20   | 30   | 40         | 50  |

ISGSGGNTYS  ADSVKGRFT1  SRDNKNSLY  LQMNSLRAED  TAVYYCARDW
| CDR-H2     |     |     |     | CDR-H3   |
| 52a        | 60  | 70   | 82abc | 90

YGMDVWGQGT  TVTVSS
|     |     |     |     |
| 99  | 101 | 113 |
```

*Figure III-26 Amino acid sequence of the M12-V$_H$ domain.* The numbering and CDR’s (underlined) were defined according to Kabat et al. (1991).
The secondary structure was assigned by using the ‘Insight-II’ program (II.1.8.II). The stereogeometry of the model was verified with PROCHECK program (Laskowski et al. 1993).

The V-domain is characterized firstly by the presence of an extra loop between strand C and D and secondly the occurrence of the conserved disulphide bonds. In V-related domains, there are about 65-75 residues between the two cysteines (Halaby et al., 1998). The amino acid sequence of M12-VH used for tracing the electron density map is shown in Figure III-26. The amino-acid sequence of the scFvM12 is given in the appendix.

The conserved disulfide bridge (Cys-H22 and Cys-H92) in the M12-VH domain is right-handed spiral type (Figure III-27) and has the following dihedral angle geometry:

\[
\begin{align*}
22\beta & \quad \alpha & \quad S & \quad S & \quad \alpha & \quad 96 \\
112.43^\circ & & 114.34^\circ & & 67.2^\circ & & 110.95^\circ & & 31.84^\circ
\end{align*}
\]

The conserved disulphide bridge (Cys22-Cys92) in the M12-VH domain is right-handed spiral type (Figure III-27) and has the following dihedral angle geometry:

The quality of the 2F0-FC map, contoured at 1σ was very fine. Even upon refinement, the observed extra electron density did not allow to interpret the structure in terms of amino acids.
for residues 40-43. The coordinates of the PEG500 molecule was downloaded from the HICUPP database for heterogeneous residues (Uppsala Software factory, Sweden) and manually modeled into the electron density by using the ‘O’ program (Figure III-28). The obtained electron density indicated the presence of a single molecule of PEG500. Ordered water molecules are generally appearing in well-defined spheres of electron density. The refined model of the M12-V\textsubscript{H} fragment contained 141 water molecules.

The majority of the ion-pairs found in proteins consist of two charges of opposite sign, the arginine was observed in unusual arrangement (Magalhaes et al. 1994). The M12-V\textsubscript{H} domain has the ArgH66-ArgH83 interaction at the surface of the protein and is well hydrated. This interaction is not bridged through water molecules rather the interaction is surrounded by water molecules (Figure III-29).

III.8.1 Antigen binding loops of the M12-V\textsubscript{H} domain

Mostly, the CDR region of antibodies contains non-polar and hydrophobic residues to interact with the antigen. Glycine residues are frequently present to allow the sharp turns because of the flexibility permitted to this residue due to the lack of the side chain. The specificity and affinity of antibodies towards antigen is a function of the canonical structure present in the hypervariable region, the size, shape and chemical character of its surface residues and their positions relative to each other.

Water molecules within the M12-V\textsubscript{H} domain structure have been observed in hydrogen bonding with only CDR-H2 residues. The canonical structure classification was performed using the software implemented on the antibodies-structure and sequence server (www.bioinf.org.uk/abs/chothia.html, Martin et al., 1996). A variation exists in the groove
for antigen binding created by the CDR's in different antibodies. In the M12-V₃ domain, the groove is not very deep (data not shown) and the surface is more or less flat. Usually, flat surfaces provide reduced surface area for the antigen binding in antibodies.

### III.8.1.1 CDR-H1 loop

The CDR-H1 loop packs across the top of the M12-V₃ domain, bridging two beta sheets. On the basis of sequence variation, residues 31 to 35 were defined by Kabat et al. (1979) as the first CDR. Structural work has shown that the region outside the beta sheet frameworks are able to adopt different conformation and comprises residues 26 to 32. It was shown that the size variations in this region involved insertions at position 31 (Chothia et al., 1992). The CDR-H1 size in M12-V₃ domain is 10 residues (H26-H35), while structurally it consists of only five residues (H31-H35). Three canonical structures have been observed for this region (Chothia et al., 1992). Of these, the canonical structure 1 is the most commonly observed conformation. The CDR-H1 loop in M12-V₃ fragment (Figure III-30) corresponds to known canonical structures, predictable from its sequence (Al-Lazikani et al., 1997) and is similar to the canonical class 1/10A. The CDR-H1 loop structure is determined by a sharp turn around Gly-H26 and the conformation of the residues Ala-H24, Phe-H27, Phe-H29 and Met-H34. The unique feature of this CDR-H1 loop is the presence of Ser-H32 (Chothia et al., 1992) residue. Normally, the allowed residues at position H32 are Ile, His, Tyr, Phe, Thr, Asn, Cys, Glu and Asp. The Phe-H29 side chain is buried between the side chains of residues Thr-H28 and Ser-H30. In M12-V₃ domain, additional hydrogen bonds are present between the side chains of Ser-H32 and Ala-H33. The additional hydrogen bond stabilizes the CDR-H1 loop and leads to the torsion angle deviation of residues H27 and H28 from the

![Figure III-30 CDR-H1 loop of the M12-V₃ domain at 1.5 Å. The dashed red lines indicate the hydrogen bonds.](image)
average torsion values (Table III-4). The inaccessibility of the Ser-H32 indicates the potential loss of its contribution on the antigen binding.

### III.8.1.2 CDR-H2 loop

On the basis of sequence variation, residues H50 to H65 were defined by Kabat et al. (1979) as the second CDR region in V_H domains. Subsequently, it was shown that residues 56 to 58 form the short C’’ strand and that the region showing conformational variation is limited to residues 52 to 56. Four canonical structures have been observed for this region. The CDR-H2 loop in M12-V_H domain has 16 residue (H51-H65) depending upon the sequence variability and belongs to the canonical class similar to class 3/10B having unobserved residues at positions H50 and H59. Structurally, the CDR-H2 loop is made up of six amino acids (H52-H56).

![Figure III-31 CDR-H2 loop (main-chain) of the M12-V_H domain at 1.5Å. The dashed red lines indicate the hydrogen bonds. The figure was prepared by using Swiss pdb viewer (Guex et al., 1997).](image)

The residue 52a and 55 determine the H2 hairpin loop conformation in the V_H 3 family. The M12-V_H domain, has glycine at H52a and H55 positions, therefore, they can adopt any conformation (Figure III-31). The M12-V_H domain has serine at H50 and H59 while the most allowed residues (Chothia numbering) at these positions are Gly, Thr, Tyr, Phe, Ile, Gln, and Val at H50 and tyrosine at H59 respectively. The main chain of the CDR-H2 loop adopts a Type I beta turn (Ser52 –Gly54).

The phi and psi angles for the CDR-H2 loop show a significant deviation from the average values (Table III-4). The reason for this deviation is the high sequence similarity of the CDR-H2 loop to subclass 3C of the CDR-H2 canonical structure 3 (Al-Lazikani et al.,
The torsion angle at H54 position (82.8/6.7) is almost exactly similar to the camel \text{V}_{H}H\text{ antibody (83/7, PDB code-1MEL).}

<table>
<thead>
<tr>
<th>Residues</th>
<th>Reference values</th>
<th>H1 &amp; H2 loop of M12-\text{V}_{H}\text{ Domain}</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>-158 157</td>
<td>-155.4 170.1</td>
</tr>
<tr>
<td>28</td>
<td>-88 111</td>
<td>-85.9 93.2</td>
</tr>
<tr>
<td>52a</td>
<td>-71 -177</td>
<td>-51.8 -41.3</td>
</tr>
<tr>
<td>53</td>
<td>80 -110</td>
<td>-85.3 -19.5</td>
</tr>
<tr>
<td>54</td>
<td>-156 32</td>
<td>95.1 3.7</td>
</tr>
<tr>
<td>55</td>
<td>-68 180</td>
<td>-69.2 -20.3</td>
</tr>
<tr>
<td>56</td>
<td>-57 160</td>
<td>-138.6 134.8</td>
</tr>
</tbody>
</table>

The structural comparison of the CDR-H2 region from different antibodies with M12-\text{V}_{H}\text{ domain indicated that the CDR-H2 region of antibody does not have much similarity regarding their torsion angle. The torsion angles are not only a function of the sequence variability but also determined by the hydrogen bonding between the different residues of the CDR region. In general, the insertion and deletion of one residue in the CDR regions does not have a profound influence on the binding affinity of the antibody towards the antigen (Lantto et al., 2002).

The comparison of dihedral angles for CDR-H2 loop of high-resolution antibodies and M12-\text{V}_{H}\text{ (Table III-5) indicated a deviation from all antibodies except 1VHP, which is a camelized human \text{V}_{H}\text{ domain.}

<table>
<thead>
<tr>
<th>Residue</th>
<th>1AQK Phi</th>
<th>1AQK Psi</th>
<th>1IGM Phi</th>
<th>1IGM Psi</th>
<th>1VHP Phi</th>
<th>1VHP Psi</th>
<th>M12-\text{V}_{H} Phi</th>
<th>M12-\text{V}_{H} Psi</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>-66.9</td>
<td>167.9</td>
<td>-106.1</td>
<td>88.5</td>
<td>-83.0</td>
<td>147.3</td>
<td>-70.2</td>
<td>160.5</td>
</tr>
<tr>
<td>52a</td>
<td>-55.6</td>
<td>-38.1</td>
<td>-</td>
<td>-</td>
<td>68.9</td>
<td>-165.8</td>
<td>-51.8</td>
<td>-41.3</td>
</tr>
<tr>
<td>52b</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-46.9</td>
<td>-25.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>52c</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>143.2</td>
<td>-10.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>53</td>
<td>-100.7</td>
<td>3.6</td>
<td>49.3</td>
<td>-75.3</td>
<td>-44.0</td>
<td>-37.0</td>
<td>-85.3</td>
<td>-19.5</td>
</tr>
<tr>
<td>54</td>
<td>78.3</td>
<td>7.1</td>
<td>-72.1</td>
<td>-7.8</td>
<td>-90.6</td>
<td>138.8</td>
<td>95.1</td>
<td>3.7</td>
</tr>
<tr>
<td>55</td>
<td>-78.9</td>
<td>-18.0</td>
<td>74.7</td>
<td>37.8</td>
<td>-147.1</td>
<td>149.4</td>
<td>-69.2</td>
<td>-20.3</td>
</tr>
<tr>
<td>56</td>
<td>-127.7</td>
<td>150.3</td>
<td>-83.0</td>
<td>-116.9</td>
<td>-125.2</td>
<td>156.8</td>
<td>-138.6</td>
<td>134.8</td>
</tr>
</tbody>
</table>

camelized human \text{V}_{H}\text{ fragment. It is also interesting to note that in both isolated domains the torsion angles of residues 53-56 are deviated significantly from the average value which

1997).
suggests the role of these residues either participate in some kind of bonding with the framework residues or interacts with the residues of the $V_L$ partner. This shows that the isolation of the $V_H$ from its $V_L$ partner does have some effect on the overall constraint of the molecule, which is more compact due to the exposure of the hydrophobic surface to the hydrophilic environment. The camelization of the $V_H$ domain was also documented to influence the geometry of the $V_H$ domain (Riechmann, L. 1996).

### III.8.1.3 CDR-H3 loop

On the basis of sequence variation, variable heavy chain residues 95 to 102 were defined by Kabat et al. (1979) as the third CDR region. The relations between the sequences and the structures of the CDR-H3 loop are not well understood as for the other CDR’s. The insertions in this region can be placed after residue 100 and numbered as 100a, 100b etc by Kabat et al. (1979). The CDR-H3 loop of M12-V$H$ domain structurally has seven residues (H95-H102) but its formation is neither of the kinked type or of the irregular type but has more an open U-shaped like structure, which is normally only present in 10-residues long CDR-H3. This relatively short CDR-H3 has been observed in other antibodies (Dokurno et al., 1998 and Kala et al., 2002). The M12-V$H$ has one deletion at H100 position according to the Kabat numbering system. The M12-V$H$ domain did not have accessible surfaces for Asp-H95, Gly-H98, Asp-H101 and Val-H102 residues. Therefore, hydrophobic residues are probably the major contributor for antigen binding. The structure of the CDR-H3 loop in M12-V$H$ domain was further investigated.

#### III.8.1.3.1 CDR-H3 torso

Until now, 49 CDR-H3 loops have been classified (Morea et al., 1998). The CDR-H3 torso region is of the bulged type in M12-V$H$ fragment, which is characterized by the presence of a $\beta$-bulge at residue H101 (Figure III-32). The bulge is stabilized by hydrogen bond interactions between the side-chain of the conserved Trp-H103 and the side chain carbonyl oxygen of Asp-H101. In addition, a salt bridge commonly occurs between Arg (or Lys) H94 and Asp-H101 leading to a C-terminal $\beta$-bulge in 41 of the 49 CDR-H3 loops in what would otherwise be a $\beta$-hairpin (Morea et al., 1998). The conserved salt bridge is also present in the M12-V$H$ CDR-H3 torso region. Smith and Xue (1997) pointed out that the first residue of the CDR-H3 is predominantly Gly, while the amino acid at this position is inherently variable based on the junctional diversity that arises during the V-D-J rearrangement (Smith et al., 1997). They also stated that the amino acid at position 95, second to Gly, is Asp. In M12-V$H$ domain, first and second amino acids are Asp-H95 and Trp-H96. Aburatani et al., (2002)
stated that the presence of Asp at position 95 also strengthens the interaction between \( V_H \) and \( V_L \) domains.

![Figure III-32 CDR-H3 loop of the M12-V\(H\) domain (1.5Å).](image)
The dashed red lines indicate the hydrogen bonds. Beta-bulge is stabilized by the hydrogen bond interaction between Asp-H101 and Trp-H103. Figure was prepared by using Swiss pdb viewer (Guex et al., 1997).

### III.8.1.3.2 CDR-H3 apex

Until now, the apices of 62 CDR-H3 loops have been classified (Morea et al., 1998) and it is reported that the head of the shorter loops (less than 14 residues long) has the form of a short hairpin. The conformation of the short hairpin depends primarily on the sequence of the residues within the loop and in particular on the position of the \( a \) Gly, Asn or Asp which are more frequently observed to have conformations with positive values of phi and psi or a Pro which can more easily accommodate a cis peptide (Chothia et al., 1987). In M12-\( V_H \) fragment Gly-H98, Tyr-H97 and Asp-H101 main-chain amino groups are involved in single hydrogen bonding with the side-chain carbonyl oxygen of Asp-H95 while the main chain of Met-H99 is involved in the formation of two hydrogen bonds with the side-chain carbonyl group of Asp-H95 (Figure III-32). An additional electrostatic interaction was observed between the main-chain Gly-H98 and side-chain carbonyl oxygen of Asp-H95. As a result of these extensive hydrogen bonding within the apex of CDR-H3 loop, the loop is directed towards the \( V_H-V_L \) interface.

### III.8.2 Hydrogen bonding

The M12-\( V_H \) domain has antiparallel and mixed type of \( \beta \)-sheets. The antiparallel sheet consists of 4 \( \beta \)-strands having the topology (-1-2x 1) and the mixed sheet consists of 6 \( \beta \)-strands having the topology (3x 1 1 –3 –1). The intramolecular hydrogen-bonding profile for the whole structure is documented the appendix. The structure mainly consists of anti-parallel
beta strands and three $3_{10}$-helices (Table III-6). One inverse gamma turn is also present in the structure.

**Table III-6 β-Strand boundaries and turns.** Boundaries of the beta strands are defined by the hydrogen-bonding patterns of the structural model and the classification of turns was performed on the basis of their phi/psi angles. The Table was prepared using the program PROMOTIF (Hutchinson et al., 1996).

<table>
<thead>
<tr>
<th>Element</th>
<th>Type</th>
<th>Boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Strand A</td>
<td></td>
<td>Gln3-Val7</td>
</tr>
<tr>
<td>β-turn</td>
<td>II</td>
<td>Glu13-Gly16</td>
</tr>
<tr>
<td>β-Strand A'</td>
<td></td>
<td>Gly10-Val12</td>
</tr>
<tr>
<td>β-Strand B</td>
<td></td>
<td>Leu18-Ser25</td>
</tr>
<tr>
<td>γ-turn</td>
<td>Inverse</td>
<td>Phe27-Phe29</td>
</tr>
<tr>
<td>Helix</td>
<td>$3_{10}$</td>
<td>Ala28-Gly32</td>
</tr>
<tr>
<td>β-Strand C</td>
<td></td>
<td>Met34-Gln39</td>
</tr>
<tr>
<td>β-turn</td>
<td>II</td>
<td>Ala40-Lys43</td>
</tr>
<tr>
<td>β-Strand C'</td>
<td></td>
<td>Leu45-Ile51</td>
</tr>
<tr>
<td>β-turn</td>
<td>I</td>
<td>Ser52-Gly54</td>
</tr>
<tr>
<td>β-Strand C''</td>
<td></td>
<td>Thr57-Ser59</td>
</tr>
<tr>
<td>β-turn</td>
<td></td>
<td>Ala60-Val63</td>
</tr>
<tr>
<td>β-turn</td>
<td>I</td>
<td>Asp61-Lys64</td>
</tr>
<tr>
<td>β-turn</td>
<td>II</td>
<td>Val63-Arg66</td>
</tr>
<tr>
<td>β-turn</td>
<td>IV</td>
<td>Lys64-Phe67</td>
</tr>
<tr>
<td>β-Strand D</td>
<td></td>
<td>Phe67-Arg72</td>
</tr>
<tr>
<td>Helix</td>
<td>$3_{10}$</td>
<td>Asn73-Lys75</td>
</tr>
<tr>
<td>β-Strand E</td>
<td></td>
<td>Ser77-Met82</td>
</tr>
<tr>
<td>β-turn</td>
<td>IV</td>
<td>Met82-Leu82c</td>
</tr>
<tr>
<td>Helix</td>
<td>$3_{10}$</td>
<td>Arg83-Thr81</td>
</tr>
<tr>
<td>β-Strand F</td>
<td></td>
<td>Ala88-Ala93</td>
</tr>
<tr>
<td>β-turn</td>
<td>I</td>
<td>Asp95-Gly98</td>
</tr>
<tr>
<td>β-turn</td>
<td>IV</td>
<td>Gly98-Val102</td>
</tr>
<tr>
<td>β-Strand G</td>
<td></td>
<td>Thr107-Val111</td>
</tr>
</tbody>
</table>

**Table III-7 Hydrogen bonding of CDR residues to the remaining M12-VH domain residues.** Antibody residues having atoms within 4.0Å distance to the loop residues were included.

<table>
<thead>
<tr>
<th>Antibody residues contacting the loop region</th>
<th>Interacting Loop residues</th>
</tr>
</thead>
</table>
III-RESULTS

Figure III-33 Secondary structure elements, relative accessibility and the hydropathic profile for the M12-V_

H domain. TT: Turn/Loops, \eta: Helix, \beta: Beta sheets, Gray star: Residues with alternate conformations, acc: Relative accessibility, hyd: Hydropathy, bottom line characters: intermolecular contacts, Green digit: disulphide bridge. The figure was prepared using the ENDscript program (Gouet et al., 2002).

Figure III-34 Two-dimensional pictorial representation of intra-molecular hydrogen bonding profile of V_

H domain. Blue colored line indicates the sequence of beta strands connection to each other. Brown colored arrowed lines indicate hydrogen bonds with the arrow pointing from the (NH) donor residue to the (CO) acceptor group. Amino-acids are represented by one letter code. Oval represents the helix while square indicates the beta strands. Amino acids are numbered sequentially not by Kabat System. The V-domain strand type was defined according to Halaby et al. 1998.
All buried polar sidechain atoms are involved in hydrogen bonds. The buried polar side chain atoms are all associated with some distinct conformational feature of the domain. The hydrogen bonds (Figure III-34) were calculated for H-O distance less than 2.5Å, with an N-H-O angle between 120° and 180° and H-O-C angle between 90° and 150°. Ser-H25 (O'γ) bond to the carboxyl oxygen of Glu-H3 stabilizing the first turn. Carboxyl oxygen of Thr-H28 bound to main-chain nitrogen of Gln-H31 and carboxyl oxygen of Phe-H29 bound to carboxyl oxygen of Ser-H32 to stabilize the 3_10-helix turn.

M12-V_H domain exhibits all the conserved hydrogen bonds except such as side chain to side-chain (Figure III-40): Asp-H86 and Tyr-H90, Arg-H94 and Asp-H101, side chain to main-chain CO: Arg-H66 and His-H82a, Thr-H87 and X84. Main chain NH to side chain: X69 and His-H59, X75 and Asp-H72. M12-V_H has Gln, while usually histidine is observed at position 82a. In this list, X refers to positions without dominant residue preference.

### III.8.3 V_H-V_L interface

The V_L interface of the V_H domains, when part of Fv fragments, is formed mainly by β-strands C’ and G, which both include a β-bulge and the CDR-H3 loop. Within the C’ strand residues H44, H45 and H47 contributing most of the V_L interface (Chothia et al., 1985). In M12-V_H fragment, residue H44 is a glycine, the most frequently occurring residue at this position (Kabat et al., 1991) and the residue H45 is a Leu and H47 is a Trp. These conserved hydrophobic residues contribute significantly to the interface interaction of the V_H with the V_L domain. In most of the V_L-associated V_H domain, the side chain of Val-H37 and Trp-H103

<table>
<thead>
<tr>
<th>Kabat residue Numbe</th>
<th>Conserved residues</th>
<th>Residues in M12-V_H domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>H^{1001}N^{636}S^{402}E^{184}</td>
<td>S</td>
</tr>
<tr>
<td>37</td>
<td>V^{2336}I^{200}</td>
<td>V</td>
</tr>
<tr>
<td>39</td>
<td>Q^{2518}K^{67}</td>
<td>Q</td>
</tr>
<tr>
<td>45</td>
<td>L^{2636}P^{16}</td>
<td>L</td>
</tr>
<tr>
<td>47</td>
<td>W^{2518}L^{64}Y^{50}</td>
<td>W</td>
</tr>
<tr>
<td>91</td>
<td>Y^{2149}F^{79}</td>
<td>Y</td>
</tr>
<tr>
<td>93</td>
<td>A^{2202}T^{1222}V^{102}</td>
<td>A</td>
</tr>
<tr>
<td>95</td>
<td>Y^{399}G^{375}S^{340}D^{226}</td>
<td>D</td>
</tr>
<tr>
<td>100</td>
<td>F^{1285}M^{550}</td>
<td>deletion</td>
</tr>
<tr>
<td>103</td>
<td>W^{469}</td>
<td>W</td>
</tr>
</tbody>
</table>
are in contact with the $V_L$ while the Arg-H38 side-chain is buried within the $V_H$. It has also been observed that the H39 makes a hydrogen bond contact with L38 residue of the at $V_H$-$V_L$ interface. The M12-$V_H$ domain has all the conserved residues at the $V_H$-$V_L$ interface except at position 100, which has undergone a deletion in M12-$V_H$.

![Antigen binding groove defined by the three CDR's of the M12-$V_H$ domain.](image)

The surface curvature of the whole molecule indicates the presence of a groove at the $V_H$-$V_L$ interface (Figure III-35). The pocket at the interface indicates the burial of the hydrophobic amino acids to facilitate the $V_H$-$V_L$ interaction. The conformations of the interacting residues 39, 44, 45 and 47 have not undergone any change when compared to the other $V_H$-$V_L$ structures.

### III.9 M12-$V_H$ structure quality assessment

The M12-$V_H$ domain structure (1.5Å) has 91.8% residues in the most favored regions of the Ramachandran plot (Figure III-36). The rest of the residues are in the additional allowed regions with no outliers. The majority of the residues are found in the anti-parallel $\beta$–strands ($\Phi = -60^\circ$ to $150^\circ$ and $\Psi = 100^\circ$ to $170^\circ$; Richardson, 1981) region. The Ramachandran plot shows the good stereogeometry of the M12-$V_H$ domain.

The M12-$V_H$ domain revealed an average B-factor of 16.67Å$^2$. In M12-$V_H$, most of the residues having the individual temperature (B) factors values less than 20.0Å$^2$ as expected for the structure at this resolution (Figure III-37). In accordance with antibody structures, the B-factors were lowest in the regions corresponding to the $\beta$–strands. The variation of the temperature factors is in accordance with the quality of the electron density. For example, the polypeptide backbone and the side chains at the N- and C-terminus exhibited significantly higher B values than those for the rest of the $V_H$ molecule. Main chain
parameters like peptide bond planarity, alpha carbon tetrahedral distortion etc. were average with regard to Ramachandran plot quality assessment. A combined plot (Figure III-43) of the average main-chain B-factors and the real space correlation coefficients (Jones et al., 1991) confirmed that all CDR residues are structurally well ordered except the few residues in the framework regions e.g. H40, H70 and H103 with an average correlation coefficient of 0.85 and an average B-factor of 16.67 Å².

In this structure, the side chain of few residues was absent within the electron density such as ArgH67, ArgH87 and LysH76. The quality of the model was checked (Table III-9) against the structure factors by using the ‘SFCHECK’ (Vaguine et al., 1999) program incorporated into the CCP4 suite of programs.
III-RESULTS

Table III-9 Statistics of the refined model against the structure factors calculated by using the ‘SFHECK’ program. \( R_{\text{stand}}(F) \) is defined as \( \langle F \rangle = \langle \sigma (F) \rangle / \langle F \rangle \), \( B \): B-factor, \( B_{\text{overall}} \): overall B-factor, \( \sigma (B) \): standard deviation of the B-factor.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>M12-V_H domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume not occupied by the model</td>
<td>12.2%</td>
</tr>
<tr>
<td>( \langle B \rangle ) for atomic model (Å²)</td>
<td>21.9</td>
</tr>
<tr>
<td>( \sigma (B) ) (Å²)</td>
<td>9.07</td>
</tr>
<tr>
<td>( R_{\text{stand}}(F) )</td>
<td>0.029</td>
</tr>
<tr>
<td>( B_{\text{overall}} ) (by Patterson) (Å²)</td>
<td>23.5</td>
</tr>
<tr>
<td>Estimated minimal error (Å)</td>
<td>0.029</td>
</tr>
<tr>
<td>( \langle u \rangle ) Luzzati plot error (Å)</td>
<td>0.148</td>
</tr>
</tbody>
</table>

III.10 Crystal packing

In the C2 space group, the asymmetric unit constitutes only 1/4 of the total unit cell volume, and in the case of \( V_H \) fragment, a single monomer is present in the asymmetric unit. Figure III-38 shows the packing of molecules within the unit-cell. The overall packing is tight with small solvent channels. The solvent content of the crystal was found to be 35.4% by assuming the protein density as 1.34. The dense packing of the protein in the crystal lattice was favorable for both high quality and high-resolution diffraction. The intermolecular crystal contacts formed due to the result of packing of the molecules are listed in the appendix. The crystal contacts are calculated for <4Å distance.

![Figure III-38 Crystal packing of Cα trace of M12-V_H domain (1.5Å) in the unit cell. The packing is drawn in Z (0,0,0) axis.](image)

The major crystal contacts between the molecules related by symmetry involved the CDR residues Thr-H28, Asn-H31, Ser-H53, Gly-H55, Asn-H56, Trp-H96, Asp-H101 and Val-H102. A contact with these symmetry-related molecules was also formed at the rim of the aromatic rings of the CDR residues Gly-H26, Gly-H54 and Met-H99.

The following crystal contacts are observed commonly in the molecular packing of \( V_H \)
domain (1.8Å and 1.5Å):

III.11 M12-VH domain crystallization in complex with MUC1 peptide

The attempt to crystallize the M12-VH domain in complex with MUC1 peptide did not proved successful. Since the MUC1 peptide was added to a drop containing the subtilisin enzyme it is possible that the peptide might have been cleaved prior to binding. This

Figure III-39 ESI-MS/MS mass spectrum (II.2.6) for the reservoir solution [100mM MES, pH6.5, 10mM ZnSo₄, 25% (v/v) PEG 550MME, 4mM DTT and 0.04% (w/v) NaN₃] from the wells used to grow the crystals of M12-VH domain in complex with MUC1 antigen (II.2.9.1.3).

Figure III-40 ESI-MS/MS mass spectrum (II.2.6) for the hanging drops used to grow the crystals of M12-VH domain in complex with MUC1 antigen (II.2.9.1.3). MUC1 peptide was added to the hanging drops after the formation of M12-VH crystals and incubated for approximately 2 months.
possibility was explored by using the ESI-MS/MS of the mother liquor (in which the crystals were allowed to grow) in comparison to the well buffer solution. The ESI-MS/MS mass spectrum indicated the presence of peptide in the mother liquor (Figure III-40); therefore, the disintegration of the peptide is not responsible for the absence of binding of the peptide to the CDR residues of the M12-V<sub>H</sub> domain. The lack of MUC1 peptide binding to the M12-V<sub>H</sub> domain is explained in the next chapter (IV.2.4) on the basis of the M12-V<sub>H</sub> structure.

### III.12 Stability analysis of scFvM12

The stability and the conformational homogeneity of a protein are essential for its crystallization. Therefore, it is a prerequisite to perform the stability analysis, to find out possible source(s) of heterogeneity (both macro or micro) generated in the crystallization drops as a result of change in temperature, pH or ionic strength. Silver staining and MALDI-TOF mass spectrometry were used to check the protein stability at different time intervals.

After 120 days of incubation at 4°C, the scFvM12 (~13mg/ml) showed a minor degradation at all the pH values (7.5, 8.0, 8.5) but interestingly, the protein aggregated strongly at pH 7.5 (Figure III-41).

![Silver stained SDS-PAGE](image)

**Figure III-41 Silver stained (II.2.4.5) SDS-PAGE (18%, 2.4.3) analysis to study the effect of pH and ionic strength on scFvM12 stability at 4°C after zero (A) and 120 days (B) of incubation.** 2µL of 2.8 mg/ml of protein was loaded in each lane. M: Molecular weight standard (10µL of Mark12). Lanes 1, 2, 3 corresponds to scFvM12 at pH 7.5, pH 8.0, and pH 8.5 respectively, while 4, 5, 6 are the lanes for the scFvM12 at pH 7.5, 8.0, 8.5 with 100mM NaCl. Low ionic strength was chosen to mimic the crystallization condition.

Although the degradation product bands (Figure III-42, Lane 4, 5, 6) were slightly clear in the presence of NaCl, the ionic strength below the physiological ionic strength is not sufficient to disturb the hydrophobic interaction between the V<sub>H</sub>-V<sub>L</sub> fragments of the scFvM12.

The protein is shows significant degradation (Figure III-42) irrespective of the pH and ionic strength after 60 days of incubation at room temperature, therefore, it can be concluded that the protein stability has direct correlation with temperature. Therefore, higher the temperature the stronger the degradation of the protein.
Figure III-42 Silver stained (II.2.4.5) SDS-PAGE (18%, II.2.4.3) analysis to study the effect of temperature on the scFvM12 stability after incubating at 4°C (40 days, A) and room temperature (60 days, B). 2µL of scFvM12 (2.8mg/ml) was loaded in each well. M: Molecular weight standard (10µL). Lanes 1, 2, 3, 4, 5, and 6 corresponds to scFvM12 incubation at 4°C while the lanes 7, 8, 9, 10, 11 and 12 corresponds to scFvM12 incubation at room temperature (RT). The samples loaded from left to right in order of scFvM12 at pH 7.5, 8.0, 8.5 and pH7.5+100mM NaCl, pH8.0+100mM NaCl, pH8.5+100mM NaCl respectively for each temperature. pH was maintained by using 20mM of Tris buffer.

Figure III-43 MALDI-TOF (II.2.6) mass spectrums for the stability analysis of scFvM12 (2.8mg/ml) antibody after four months storage at 4°C. A – 20mM Tris pH 7.5; B – 20mM Tris pH7.5, 100mM NaCl; C – 20mM Tris pH8.0, 100mM NaCl; D – 20mM Tris pH8.5, 100mM NaCl.

In all the mass spectrums (Figure III-43), the main peak emerged was around 25kDa, which indicated that the protein degraded during storage. The most susceptible portion is the linker.
and the marker tag. After linker cleavage, the major fraction of the protein peak should appear around 12kDa, corresponding to the isolated $V_H$ and $V_L$ domains of the protein. Since, the 12kDa peak does not correspond to the parent peak in the mass spectrum, therefore, this peak should be considered the outcome of protein protonation during the mass measurement. The appearance of 25kDa peak indicates either the complete or incomplete cleavage of the marker tags.

The theoretical calculation of the mass corresponding to 25.18kDa and 25.16kDa indicates the complete removal of the His$_6$ and the c-myc tag, while the 25.674kDa mass indicates the complete removal of the His$_6$ tag along with removal of ten residues of the c-myc tag.

The removal or inactivity of the marker tags was confirmed through the western blotting by using the anti-c-myc and anti-His antibody (Figure III-50). This confirmed the complete cleavage of the marker tags during storage without the addition of the exogenous proteases over four months.

![Figure III-44](image)

**Figure III-44 Western blot (II.2.4.4) analysis of scFvM12 antibody stored at 4°C for four months with and without 100mM NaCl at different pH values.** The blot developed with anti-9E10 antibody (1:4000 dilution, A) and anti-His antibody (1:4000 dilution, B). M: Prestained molecular weight marker (10µL), Lane: 1, 2, and 3 indicates the scFvM12 antibody in 20mM Tris at pH 7.5, pH8.0 and pH8.5; Lanes 4, 5, and 6 indicates the scFvM12 antibody in 20mM Tris at pH 7.5, pH8.0 and pH8.5 with 100mM NaCl.

The scFvM12 antibody in Tris at pH 7.5 revealed a single band, when developed with anti-9E10 antibody this shows that only the histidine tag is cleaved at this particular pH. The occurrence of cleavage due to the presence of the exogenous proteases in the samples can be ruled out because there is only loss of purification tags although the linker is also highly susceptible to proteolytic cleavage. This phenomenon is difficult to explain.

### III.13 M12-\(V_H\) domain engineering

The cloning and the expression of the native and camelized M12-\(V_H\) domains were supported by Melanie König (Intitute for Biologie VII, RWTH- Aachen, Germany).
III.13.1 Cloning strategy

For the camelization, the $V_H$ domain of scFvM12 antibody was cloned as reference antibody fragment for the comparison of the functionality and solubility properties of the native and camelized antibody $V_H$ domain. The cloning strategy is depicted in figure III-45.

The $V_H$ domain of scFvM12 was cloned from the nucleotides sequence listed in appendix. As it is evident from the nucleotide sequence the restriction site for the NdeI and XhoI enzymes are not present. Therefore, they had to be introduced with PCR primers. The $V_H$ domain gene of scFvM12 was amplified directly from the pSynI vector containing the gene segment for the full size antibody fragment scFvM12. After amplification, the $V_H$ domain gene was cloned in the pK3C vector for cytoplasmic expression. The camelized $V_H$ fragment gene was amplified from the $V_H$ gene of the scFvM12 by using the appropriate primers (see in appendix) and cloned in the pK3C bacterial cytoplasmic expression vector.

III.13.2 Amplification of the M12-$V_H$ domain

The gene for cloning the native $V_H$ domain of scFvM12 was amplified through PCR by using the different concentration of MgCl2. The sense and anti-sense primers (appendix) used for the gene amplification were same in all the PCR amplification reactions.
amplified gene fragments were labelled as M12-V\textsubscript{H} (2/4/5/7) respectively corresponding to the different concentrations of MgCl\textsubscript{2} (1.0mM, 2.5mM, 3.0mM, and 4.0mM).

![PCR amplification image]

**Figure III-46 PCR amplification (II.2.1.3.2) of the camelized (Trp47Gly & Trp47His) M12-V\textsubscript{H} domain.** The camelized M12-V\textsubscript{H} domain was amplified by PCR and separated on 1.2% agarose gel (II.2.1.1). 5\mu L of amplified product was loaded in each lane. Lane 1: Marker digested with Pst I, Lane 2: Primer combination 1a for M12-V\textsubscript{H}camG, Lane 3: PCR mix with out DNA; negative control, Lane 4: Primer combination 1b for M12-V\textsubscript{H}camG, Lanes 5, 7, 9 & 12 are PCR mix with out DNA; negative control, Lane 6: Primer combination 2a for M12-V\textsubscript{H}camH, Lane 8: Primer combination 2b for M12-V\textsubscript{H}camH, Lane 10: M12-V\textsubscript{H}camG (Trp47Gly), Lane 11: M12-V\textsubscript{H}camH (Trp47His). The primer combinations are 1a (M12-V\textsubscript{H} forward primer + M12-V\textsubscript{H} backward primer, Trp47Gly), 1b (M12-V\textsubscript{H} backward primer + M12-V\textsubscript{H} forward primer, Trp47Gly), 2a (M12-V\textsubscript{H} forward primer + M12-V\textsubscript{H} backward primer, Trp47His) and 2b (M12-V\textsubscript{H} backward primer + M12-V\textsubscript{H} forward primer, Trp47His).

The gene for the camelization of M12-V\textsubscript{H} fragment was amplified using the original M12-V\textsubscript{H} gene (clone number 4) as DNA template. The forward and backward primer combination (see appendix) was used for the PCR amplification to introduce the appropriate mutated codons for amino acid mutations Gly44Glu, Leu45Arg and Trp47Gly/His. The resulting amplified gene products for position 47 were labeled as V\textsubscript{H}camG (Trp47Gly) and V\textsubscript{H}camH (Trp47His) respectively (Figure III-46).

All the primers used for the amplification of the M12-V\textsubscript{H} genes possessed overhangs to generate the restriction sites for N\textit{del} and X\textit{ho}I restriction enzymes.

### III.13.3 Restriction and ligation

The PCR amplified gene products (Figure III-46) for the original and the camelized M12-V\textsubscript{H} domains were pooled and purified from the agarose gel. The gel purified M12-V\textsubscript{H} genes were digested with N\textit{del} and X\textit{ho}I restriction enzymes. The sticky end gene fragments, generated after restriction digestion, were ligated into the pK3C vector and the recombinant vector was transformed into \textit{E. coli} DH5\textalpha cells by the heat shock transformation. Single randomly selected colonies of M12-V\textsubscript{H} (2/4/5/7) clones were checked through PCR in order to confirm the presence of inserts. All recombinant colonies contained the inserts (data not shown).
The recombinant plasmid containing the camelized genes was isolated from the five randomly selected colonies to check the presence of the inserts through PCR (data not shown). All the clones tested for the presence of gene of interest were found positive with an expected gene fragment size of ~350bp.

The colonies found to contain the desired native and camelized M12-VH genes were selected and the isolated DNA was subjected to sequencing. Only clones carries the correct M12-VH gene were used for further analysis.

The recombinant plasmid harboring the correct sequence of the M12-VH genes were isolated from the E. coli DH5α bacterial cells and transformed in to the E. coli BL21 (DE3). The expression of the protein was checked through SDS-PAGE (III.2.4.3).

III.13.3.1 M12-VH domain

The pK3C plasmid (appendix) lacks the pelB leader sequence necessary to express the protein in the bacterial periplasm. The vector contains the His6 tag at the c-terminus necessary to purify the protein by Ni-NTA affinity chromatography. For the cytoplasmic expression of M12-VH fragments, the recombinant plasmids were transformed into the E. coli BL21 (DE3) strain and the expression was carried out under the control of the strong phage T7 promotor. When protein expression was carried out overnight with 1mM of IPTG at 37°C in LB medium, the observed expression level of the protein was very poor (data not shown). The reduction of temperature from 37°C to 18°C and the change of IPTG concentration did not affect the level of expression. The change of LB medium to more nutrient rich medium i.e. 2xTY in improved protein expression but the protein undergone misfolding and only expressed in inclusion bodies which can be solubilized by the addition of 1% (v/v) Triton X-100 in the purification buffer.
III-RESULTS

To overcome solubility problem of the protein, the recombinant plasmid containing the gene of interest was subcloned in to *E. coli* BL21 (DE3) codon plus RIL strain (Novagen). The expression of the protein in the codon plus strain resulted in to increased in expression levels (Figure III-48).

![Western blot analysis](Image)

**Figure III-48 Western blot analysis (II.2.4.4) to observe the effect of *E. coli* bacterial strains on the cytoplasmic expression of the native and the camelized M12-V<sub>H</sub> domains (II.2.2.1).** The blot was developed with anti-His antibody (1:4000 dilution). A: *E. coli* BL21, B: *E. coli* BL21 codon plus RIL, M: Prestained molecular weight marker (10µL). Lanes 1, 2, 3, 4 are M12-V<sub>H</sub> fragment (4), M12-V<sub>H</sub> fragment (7), M12-V<sub>H</sub> fragment camG (4) and M12-V<sub>H</sub> fragment camH (6) supernatant. Lanes 5, 6, 7, and 8 are pellets of M12-V<sub>H</sub> fragment (4), M12-V<sub>H</sub> fragment (7), M12-V<sub>H</sub> fragment camG (4) and M12-V<sub>H</sub> fragment camH (6) pellet. The pellet was lysed in 15ml of lysis buffer for 3hrs and the supernatant kept separately. The pellet was washed in binding buffer containing 1% v/v TritonX-100. The pellet was boiled in 20µL of 5x sample loading dye buffer before loading onto the Gel. 10µL of each sample loaded in each lane of the 12% SDS-PAGE Gel. The numbers included in parenthesis represents the selected clone number.

The temperature and the IPTG concentration were optimized but no significant effect on protein expression level was observed by changing these variables. The best parameters for the obtaining the large yield of the soluble cytoplasmic expression of the V<sub>H</sub> fragments were obtained by using the nutrient rich medium i.e. 2xTY and the change of bacterial strain from *E. coli* BL21 (DE3) to *E. coli* BL21 (DE3) codon plus. The induction was performed overnight at 28°C with 0.5mM IPTG (final concentration). The expression level of the proteins was checked through western blotting (Figure III-48).

**III.13.4 M12-V<sub>H</sub> domains**

After lysis of the bacterial pellet through sonication and lysozyme treatment, the His<sub>6</sub> tag containing protein from inclusion bodies was purified by Ni-NTA chromatography (II.2.3.2). The inclusion bodies are subjected to refolding by using the strong denaturant guanidine-HCl (6M). It seems that the M12-V<sub>H</sub> did not refold correctly because it was not binding to the Ni-NTA column under the selected purification conditions (II.2.3.2.1). The expression level was slightly improved by expressing the protein in nutrient rich 2xTy medium (data not shown). The protein expression was performed at 37°C in 2xTy medium and the culture was induced for 4hrs with 1mM of IPTG (II.2.3.2.2).
After purification, the protein was concentrated and loaded on 12% SDS-PAGE. Since, the gel was showing the inclusion of large number of impurities, it is difficult to conclude the high level of expression of the desired proteins.

Western blot analysis (Figure III-50) of the Ni-NTA purified M12-V$_H$ domain revealed that the expression is poor but the protein remained intact during the harsh purification procedure from bacterial cytoplasm. The M12-V$_H$ domain expression was found to be variable (Riechmann, L. 1996). To improve the cytosolic expression, the parameters such as temperature, length of incubation with IPTG etc needs to be optimized. Once the soluble protein is available in sufficient amount, the functionality of the M12-V$_H$ domains needs to be checked by ELISA.
IV DISCUSSION

The display of antibody on the surface of phages has tremendously increased the potential to apply them for the specific targeting of ligands and pathogens presented to them (Winter et al. 1994). Antibody domains with specificity for virtually any antigen can be isolated in-vitro (Griffiths et al., 1994) and afterwards can be produced in eukaryotic cells or bacteria for biotechnological and pharmaceutical applications. Bacterial folding of the entire immunoglobulin molecules is less efficient than the folding of their Fv domains or Fab domains. For some applications like the targeting of solid tumors, smaller antibody domains (e.g. scFv, V\textsubscript{H} or V\textsubscript{L} domains) are advantageous over full-size antibodies in respect of their biodistribution (Mayer et al., 1999) due to their faster diffusion and clearance. It was observed that the variable heavy chain domain of antibodies could sometimes retain a significant portion of antigen affinity in the absence of light chains (Utsumi & Karush, 1964). Recently, antibodies naturally devoid of light chain and retaining full functional activity against the antigen were reported (Hamers-Casterman et al., 1993). Since then, minimizing the size of antigen-binding proteins to a single immunoglobulin domain has been one of the goals of antibody engineering which involves the modification of the human V\textsubscript{H} domain by mimicking camelids heavy chains antibodies (V\textsubscript{HH}) for their use as a small recognition unit (Davies & Riechmann, 1994).

MUC1 is a polymorphic, type I transmembrane glycoprotein containing a variable number of 20 amino acid tandem repeats. The exposure and the extra cellular accessibility of the aberrantly glycosylated MUC1 in cancer makes this marker a suitable immunotarget. So far, structures of either murine or humanized Fab antibody domains directed against MUC1 have been solved (Dokurno et al., 1998; Banfield et al., 1997). But there is no report for the availability of structural details of a single domain antibody directed against MUC1. This thesis work reports the crystallization of human single domain antibodies V\textsubscript{H} generated through limited in-vitro proteolysis of the scFvM12 antibody domain. The structure of the M12-V\textsubscript{H} domain has been determined at high resolution (1.5Å).

In antibodies, the V\textsubscript{H}-V\textsubscript{L} interface is hydrophobic in nature resulting in unstable and insoluble the expression of isolated V\textsubscript{H} domains. The hydrophobic residues mainly involved in the V\textsubscript{H}/V\textsubscript{L} interaction are H44, H45 and H47. These residues were mutated in the M12-V\textsubscript{H} domain to mimic the camelized V\textsubscript{H}H domain. Simultaneously, both the native and camelized M12-V\textsubscript{H} domains were expressed in the bacterial cytoplasm. Although the reducing environment of the cytoplasm inhibits the disulfide bridge formation in antibodies, certain
antibodies were found to be stable in the absence of the conserved disulfide bridge. The M12-V_{H} domain inherently possesses the Arg-H66 and Ser-H52 residues, which were reported to have stabilizing effect and allowed the antibody expression in the cytoplasm (Proba et al., 1998). Therefore, the bacterial cytoplasmic expression of the single domain antibodies should provide further evidence of the stabilizing effect of the Arg-H66 and Ser-H52 on one hand and the effect of camelization on the cytoplasmic expression level of these small recognition units on the other hand.

The crystallization of the scFvM12 antibody alone was performed to observe at the molecular level how the two domains, V_{H} and V_{L}, coming from two different sources behave in terms of their interface interactions.

The binding of any antibody depends on its CDR regions and it is possible that antibody domain affinity towards the antigen can alter in the absence of all natural CDR’s. To study the structure-function relationship and the contribution of the V_{H} and the V_{L} domain of scFvM12 in antigen binding, the crystallization of both the M12-V_{H} and the scFvM12 with and without the MUC1 peptide antigen was investigated.

**IV.1 Expression and purification of scFvM12**

**IV.1.1 Bacterial expression**

Initially, the scFvM12 was expressed in the *E. coli* BL21 (DE3) periplasm with a yield of 0.2-0.3 mg/L of bacterial culture. The yield was optimised using the different bacterial strains (data not shown). *E. coli* TG1 was identified to express the highest yield (~0.6mg) per litre of the bacterial culture in shake flasks. The reason for this variable expression levels appeared to be strain specific. The primary sequence of the antibody affects the folding pattern and misfolding of the protein results into the formation of inclusion bodies, which can be detrimental to the host cell. Varying the host cells can alter the protein expression levels and the folding state of the protein (Kenealy et al., 1987).

Crystallization trials needed to be set-up using at least 5mg/ml of highly pure protein. Obviously, the yield of the scFvM12 in shake flask was not sufficient to meet this demand. Therefore, it was essential to scale up the protein expression level. Protein production in *E. coli* could be increased significantly by the use of high-cell-density fermentation (HCDF) in batch, fed batch and continuous mode. These methods can achieve 100g dry cell weight per liter of the medium and provides cost effective production of recombinant proteins. The HCDF was carried out in batch mode (III.1.2) to increase the scFvM12 yield. The yield of the scFvM12 was increased to 24mg/kg of the pellet (wet weight). In HCDF, leaky...
expression of the scFvM12 was observed (Figure III-2) and the bacterial supernatant was found to contain sufficient amount of protein (~7mg per litre).

**IV.1.2 scFvM12 purification**

The affinity (Ni-NTA) purification of scFvM12 yielded apparently homogenous protein. However, ion-exchange purification, multiple peaks were observed. In cation-exchange chromatography, the first broad peak (Figure III-4A, peak A) corresponds to imidazole because no reaction was observed with Bradford reagent and no distinct band was observed on SDS-PAGE. Therefore, it could be the remaining fraction of the imidazole, which is carrying over due to the improper dialysis. Furthermore, the occurrence of multiple peaks during ion-exchange purification of scFvM12 could be either due to presence of aggregates or presence of different conformational forms or the presence of products of the partial degradation of the scFvM12. The multiple peaks were tested through western blotting (Figure III-5A & III-5B) and all of them were found to bear the marker tag except the first observed peak. Western blotting suggests therefore, that the multiple peaks belong to the scFvM12. The iso-electric focusing (IEF) confirmed that the protein was present in different conformational forms (Figure III-5). Since, MALDI-TOF mass spectrometry indicated only presence of a protein population with a unique mass and the formation of aggregates in minor fraction (Figure III-8) and the degradation of the protein also observed upon storage (Figure III-42).

It is advantageous to express the protein in the bacterial supernatant in terms of purification. The main difficulty in the protein purification from the bacterial supernatant is its slimy nature and the presence of a lot of cell debris, which are difficult to separate through centrifugation or filtration. These slimy materials cause frequent obstructing of the column and leads to a rapid increase in system backpressure. Tangential flow filtration was assessed as an additional clarification step but proved to be ineffective, since filtrate flow was unacceptably slow. A possible explanation was the presence of an antifoam additive in the supernatant, since; the used silicon polymers may form a layer on the filtration membrane surface, preventing the passage of liquid through the pores. An additional technical complication is the presence of high concentration of salts in the minimal medium, which interfere with the binding of the protein to the Ni-NTA column, possibly due to the presence of ions, which can be chelated by the His6-tag in the solution, preventing its binding to the Ni-NTA column. During the purification of the protein, the protein of interest remains contaminated with bacterial proteins.
IV.2 Crystallization of antibodies

IV.2.1 Crystallization of scFvM12

Out of the 509 antibody structures had been reported in the PDB database at the time of writing the thesis (Bernstein et al. 1977), only 16 were scFv’s. Most of the antibody structures are Fab domains. Out of the 16 scFv’s structures, 8 were solved in complex with their ligand. The low number of available of scFv structures in total in the PDB database gives an indication about the difficulties encountered in crystallizing these antibody fragments. In general, the concentration of scFv’s for crystallization purposes most often results in a monomer-multimer equilibrium causing their aggregation, which interferes with the crystallization process (Alfthan et al., 1995). The scFvM12 is not an exception to this behavior and aggregate formation is observed in MALDI analysis (Figure III-6) for a minor fraction of the scFvM12.

By crystallizing the gel-purified protein, the protein always yields thousands of microneedles within 3-4 days after setting the hanging drops through the vapor diffusion method. These microcrystals are quite resistant to grow in size. To improve the crystal quality, the microseeding or macroseeding technique was tried but the results were not encouraging. The observed high nucleation rate and the resistance to grow in size even after seeding suggest heterogeneity of the protein in the crystallization drops. The heterogeneity could result from the presence of the multimeric form protein (Figure III-6) or its degradation with in the drop (Stanfield et al., 1993). These conformational heterogeneities can either act as a nucleus for the soluble protein, resulting into thousands of microneedles or interfere with nucleation. The resistant of these needles to grow in size can be explained due to the contamination of the crystal-growing surface with the multimeric species and/or molecules with different conformations.

To reduce the scFvM12 monomer-multimer equilibrium during crystallization, two strategies were applied. Firstly, the most notable is the cleavage of highly flexible regions such as the linker joining the V_L and V_H domains as well as the c-terminal c-myc and His_6-tag with some protease (Essig et al., 1993). We used subtilisin Carlsberg and thermolysin. Secondly, crystallization of the scFvM12 in the presence of its ligand i.e. MUC1 in order to stabilize its CDR loops. Furthermore, the ligand binding might stabilize the V_H/V_L interaction and avoid rearrangements to other monomeric or multimeric states.
IV.2.2 Single antibody domain crystallization

The most flexible portions of the scFv’s are the linker between the $V_H$ and $V_L$ domain and the purification tags. Therefore, theoretically, it is possible to cleave the flexible and susceptible portions of the protein to generate either the single antibody domains, a compact form of the scFv. We used in-drop limited proteolysis of the scFvM12 to generate the isolated $V_H$ domain.

The proteolytic cleavage of the scFvM12 with subtilisin Carlsberg lead to the crystallization of a single protein domain. The washed and solubilized crystals, when examined through MALDI-TOF, indicated the crystallization of the $V_H$ domain (12.346kDa) of the scFvM12 antibody (Figure III-14). The crystallization of the M12-$V_H$ domain after subtilisin treatment indicates that the rest of the scFv portions including the $V_L$ domain are highly susceptible to this protease and degraded. The further degradation of the M12-$V_H$ was circumvented by its crystallization, rendering it inaccessible for further degradation. The morphology of the crystals varied with pH and temperature (III.4.2.C). The difference in morphology at two different temperatures may be the result of the difference:

1. in the kinetics of the subtilisin protease,
2. and/or in the rate of drop supersaturation.

The first explanation is supported by the immuno-blot analysis of scFvM12 digestion by subtilisin under the crystallization condition [100mM MES pH6.5, 25% (v/v) PEG 550MME, 10mM Zinc sulfate heptahydrate, 4mM DTT, 0.04% (w/v) NaN₃] at 4°C. The reduced catalytic activity of the enzyme under crystallization condition could be due to the combined effect of precipitant, high concentration of divalent metal ion and low temperature.

Similarly, the $V_H$ domain crystals were generated through in-drop limited proteolysis of scFvM12 by using thermolysin protease (Figure III-16). It is interesting to observe that the M12-$V_H$ domain crystals, obtained through subtilisin treatment of the scFvM12, remain in crystal form while the M12-$V_H$ crystals obtained through thermolysin treatment undergoes solubilization. This observation suggests that the M12-$V_H$ crystallization is an intermediate step in the complete proteolytic degradation of the protein. The probable explanation is the shift of equilibrium relative to the proportion of the substrate for M12-$V_H$ domain generation and the increase of the M12-$V_H$ domain concentration leads to the drop supersaturation and ultimately, the M12-$V_H$ domain crystallization. Till the equilibrium remains in favor of the supersaturated state, the M12-$V_H$ domain remains in crystal form. After the degradation of the more susceptible species, the equilibrium reverts back and results in to the dissolution of
the crystal. After entering into solution, the M12-VH domain undergoing the complete cleavage.

**IV.2.3 Preference for the M12-VH domain crystallization over M12-VL domain**

The scFvM12 antibody was derived by selection of an scFv phage display library with a diversity of $10^9$ (Wong et al., 2001). The VH domain of scFvM12 was derived from the library constructed by pooling total DNA from human spleen cells and two different samples of human peripheral blood lymphocytes while the VL domain was taken from a previously constructed scFv phage antibody library derived from peripheral blood lymphocytes (Sheets et al., 1998).

On sequence comparison with the germline database (V-Base), it was observed that M12-VL gene sequence is nearly identical to germline V-genes (data not shown) while the M12-VH gene had sequence undergone multiple mutations (Figure IV-1).

```
1-3 3-48 EVQLVESGGGLVQPGGSLRLSCAASGFTFS S--YSMN WVRQAPGKGLEWVS YISS--
1-3 3-21 EVQLVESGGGLVKPGGSLRLSCAASGFTFS S--YSMN WVRQAPGKGLEWVS SISS--
1-3 3-11 QVQLVESGGGLVKPGGSLRLSCAASGFTFS D--YYMS WIRQAPGKGLEWVS YISS--
M12VH QVQLQESGGGLVKPGGSLRLSCAASGFTFS N--SAMS WVRQAPGKGLEWVS SISG--
1-1 3-13 EVQLVESGGGLVQPGGSLRLSCAASGFTFS S--YDMH WVRQATGKGLEWVS AIG---
1-3 3-48 SSSTIYYADSVKGRFTISRDNAKNSLYLQMNSLRDEDTAVYYCAR 13
1-3 3-21 SSSITYYADSVKGRFTISRDNAKNSLYLQMNSLRDEDTAVYYCAR 12
1-3 3-11 SGRSTIYYADSVKGRFTISRDNAKNSLYLQMNSLRDEDTAVYYCAR 11
M12-VH SGGNTYSADSVKGRFTISRDNAKNSLYLQMNSLRDEDTAVYYCAR
1-1 3-13 TAGDTYYPGSVKGRFTISRENKNSLYLQMNSLRAGDTAVYYCAR
```

**Figure IV-1** Sequence alignment of M12-VH domain with other VH domains in the V-Base database to determine the germline family for the M12-VH domain.

Since the VH and VL domain came from two different libraries in scFvM12 and the M12-VH domain has undergone multiple sequence mutations, it is possible that the M12-VH domain corresponds to a mature antibody like IgG, while the M12-VL domain belongs to immature antibody. Since, the mature antibodies have stronger affinity towards the antigen in comparison to the immature antibodies, it might be possible that the M12-VH domain is the main contributor towards the binding specificity of the scFvM12 antibody. Furthermore, the formation of multimeric forms of the scFvM12 (Figure III-5) might be due to non-ideal interaction between the VH and VL domains. This might be due to the non-altering rounds of maturation between VH and VL coding genes normally observed in B-lymphocytes during maturation.
The random combinatorial libraries through which scFvM12 had been constructed has potential limitation. The scFvM12 antibody was constructed by randomly pairing the highly complex and MUC1 specificity with general and unspecific populations of V\textsubscript{H} and V\textsubscript{L} cDNA’s. Therefore, the chances for combining the specific V\textsubscript{H} with its original V\textsubscript{L} partner are negligible. Pairing a V\textsubscript{H} domain with unrelated V\textsubscript{L} partner usually results in a less active or even non-functional antibody. Consequently, the V\textsubscript{H} domains that could function as anti-tumour V\textsubscript{H} antibodies might even remain undetected in such single chain Fv or Fab libraries because of an incompatible V\textsubscript{L} partner might prevent binding (Cai \textit{et al.}, 1996). Therefore, it is possible that the V\textsubscript{H} domain in the scFvM12 is seen not showing its full activity in combination with its unnatural V\textsubscript{L} partner. Simultaneously, the existence of fully functional V\textsubscript{H} domains in Camelids (Hammers-Casterman \textit{et al.}, 1993) provides the additional evidence that a functional antibody does not necessarily needs to be associated to a V\textsubscript{L} domain.

IV.2.4 Diffusion over co-crystallization

Since, single antibody domains have vast clinical importance, the structural details of the antibody-antigen complexes could provide the basis for a protein engineering approach for the M12-V\textsubscript{H} domain in terms of its solubility, affinity and avidity towards its antigen MUC1. The crystallization of M12-V\textsubscript{H} in complex with MUC1 antigen was tried by the diffusion method, where the crystal of the M12-V\textsubscript{H} domain was generated first by in-drop proteolytic digestion of the scFvM12 (III.4.4). These crystals were then soaked with peptide to allow binding of the peptide through solvent channels, which are inherent part of the protein crystals. Diffusion of 21mer biotinylated peptide (APDTRPAGSTAPPAHGVTSK-ε-biotin) into the pre-existing crystals of the M12-V\textsubscript{H} domain did not result in any morphological damage or interference in the ability of the crystals to diffract X-rays. There were no resolution losses, widening of the X-rays or decay increases over time. However, after processing the data, the presence of bound peptide could not be detected by difference Fourier analysis.

The structure determination of the M12-V\textsubscript{H} domain showed that the complex formation was inhibited due to tight interdomain crystal contacts at the site of the CDR’s and therefore, not accessible to peptide binding. The peptide was found in the mother liquor after ESI-MS/MS analysis (Figure III-38); therefore, the binding inhibition is due to the inaccessibility of the peptide to the V\textsubscript{H} binding site and not due to a digestion by the presence of the protease. Protein-ligand complex crystallization through soaking of the protein crystals most often suffers from the accessibility problem of the ligand to the binding site of the protein. The low Matthews coefficient (1.98) also suggests the tight packing and
less available space for ligand binding. This phenomenon was also reported in other cases (Faber et al., 1998). The CDR regions of the M12-V\textsubscript{H} involved in the hydrogen bonding between symmetry related molecules and these bonds are so strong that they did not allow the interaction between the peptide antigen and the M12-V\textsubscript{H} domain even if the peptide has the access through the solvent channels. To avoid such problems, co-crystallization would be the only alternative. Co-crystallization needs the complexation of the protein and the ligand prior to crystallization; therefore, the recombinant M12-V\textsubscript{H} domain should be used for studying the structural details of the antibody-antigen complex.

### IV.3 Structural deviations of the M12-V\textsubscript{H} domain

The M12-V\textsubscript{H} domain has the highest sequence (76.11\%) and structural similarity with the recently solved structure of a camelized human single domain (accession code-1VHP, NMR structure).

In the V\textsubscript{H}3 family of germline sequences, there is more variation in CDR-H2, because of the length variation of a two amino acid residue insertion occurring in a group of human sequences (positions 52b and 52c). The CDR-H1 and CDR-H2 loops of the M12-V\textsubscript{H} domain are similar to the type 1 and type 3 canonical structures respectively. Both loops show deviations from the average phi/psi angles (Table III-4). The presence of the disallowed amino acids enables the formation of additional hydrogen bonds and resulted into change of a phi/psi angle. As a result of this additional hydrogen bonding, these CDR's are forming flat binding interface instead of the usually observed cleft-like binding pockets. The product of three genes (V, D and J), the CDR-H3 exhibits extremely high variability. It is therefore considered too hypervariable in structure to assign it to a canonical niche. The CDR-H3 of the M12-V\textsubscript{H} has a deletion at position 100 and is considered to be a rather small loop. The M12-V\textsubscript{H} lacks some of the conserved hydrogen bonds between side chains (Asp-H86 and Tyr-H90, Arg-H94 and Asp-H101), side chain to main-chain CO (Arg-H66 and His-H82a, Thr-H87 and X84), Main chain NH to side chain (X69 and His-H59, X75 and Asp-H72). M12-V\textsubscript{H} has Gln, while usually histidine is observed at position 82a. In this list, X refers to positions without dominant residue preference.

### IV.4 Structural basis of the M12-V\textsubscript{H} domain stability

It is noticeable that the V\textsubscript{H} domain structure has a strongly distinct network of electrostatic interactions. Residue H66 is part of a highly conserved charge cluster within the V\textsubscript{H} domain family. 90.4\% of human V\textsubscript{H} sequences contain Arg at this position and 1.3\% Lys. Residue
Arg-H66 interacts strongly with Asp-H86 by forming a salt bridge. This salt bridge is quite buried; increasing its stabilizing effect compared to that of a solvent-exposed ion pairs (Proba et al., 1998, Figure IV-2). This salt bridge is also present in the M12-VH domain. In addition, M12-VH has a Ser at position H52, which is forming a beta-turn and might therefore have some influence on folding kinetics.

![Figure IV-2 Schematic representation of denaturation midpoints of the light-chain variable domain (VL) and the heavy-chain variable domain (VH) within the scFv domain A48 and its derived mutants (K66R and N52S). Note that VL is chemically identical in all cases, whereas the nature of VH is indicated (Wörn et al., 1998).](image)

Wörne et al., (1998) have shown the stabilizing effect of K66R and N52S mutations on an scFv domain naturally lacking the conserved disulfide bridge (Figure above). The presence of Arg-H66 and Ser-H52 in the VH domain of scFvM12 antibody inherently stabilizes this molecule.

**IV.5 The VH/VL interface of the M12-VH domain**

To ascertain the difference in the VH/VL interface of the M12-VH domain after its separation from the scFvM12 domain, we build a theoretical model of the scFvM12 based on the antibody structure and the structure of the M12-VH domain was superimposed on it. The crystal structure of the M12-VH domain showed a high degree of similarity with the modeled structure for scFvM12 antibody. As predicted, the overall shape of the antigen-binding site is in resemblance between model and M12-VH domain with the hydrophobic residues at the bottom of the VL interface. Comparison with the modeled scFvM12 structure revealed no significant differences for framework residues. The VL interface of the VH domains, when part of Fv domains, is formed mainly by beta-strands C’ and G and the CDR-H3 loop. Within strand C’, residues 44, 45 and 47 are those contributing most to the VL interface (Poljak et al., 1975; Chothia et al., 1985). There is no principle difference in the conformation of these residues when compared to M12-VH structure. The three residues H44, H45 & H47 are highly conversed through the human and mouse VH families. Residues 45 & 47 are more
conserved than residue 44 in which some variability among V\textsubscript{H} families is observed. In most V\textsubscript{H}-domains, residue 44 is a glycine (Kabat et al., 1991). No changes concerning the V\textsubscript{H}-typical beta-structure were observed in strand C’ and strand G. Residues Val-H37 and Arg-H38 have the same orientations in M12-V\textsubscript{H} and the scFvM12 model. The Trp-H103 side-chain is exposed to the solvent. Therefore, it can be anticipated that the M12-V\textsubscript{H} domain does not exhibit any structural difference after its isolation from the V\textsubscript{L} partner.

It is interesting to note that on the V\textsubscript{L}/V\textsubscript{H} interface, no change was observed in the conformation of the interacting residues even after exposure to the solvent. The most probable explanation is that to establish the contact between symmetry related molecules inside the crystal, the V\textsubscript{H}/V\textsubscript{L} interface residues are heavily guarded by intermolecular hydrogen bonding. Therefore, the residues escaped the effects of the solvent exposure.

IV.6 Comparison between the SM3-V\textsubscript{H} domain and the M12-V\textsubscript{H} domain

Murine SM3-Fab is the first tumour-specific antibody crystallized in complex with the MUC1 peptide antigen. The human scFvM12, from which the V\textsubscript{H} domain has been crystallized, is also directed against MUC1. Therefore, the sequence and structure comparison can provide information about the components of the V\textsubscript{H} domain, which might contribute towards antigen binding and how this binding can be altered to improve the affinity of the V\textsubscript{H} domain. Compared to M12-V\textsubscript{H}, SM3-V\textsubscript{H} antibody has two insertions at position at H52 and a deletion at position H99. Interestingly, both antibody domains have deletion at position H100. There is a marked difference in the CDR loop regions between the two antibodies especially the CDRH2 and CDRH3 (Table IV-1). SM3 has a cis-peptide bond between Gly96\textsubscript{H}-Gln97\textsubscript{H}, while there is nothing similar in M12-V\textsubscript{H} domain. The SM3 antibody V\textsubscript{H} domain has Gly44\textsubscript{H}, Leu45\textsubscript{H} and Glu47\textsubscript{H}. Two of the residues are conserved but at position 47 unusually has a hydrophilic residue, while the M12-V\textsubscript{H} domain has all the

![Figure IV-3 Comparison of amino acid of SM3 antibody variable heavy chain with the M12-VH domain.](image-url)

Numbering and CDR's (underlined) are defined according to Kabat et al., (1991).
Table IV-1 Canonical structure assignment according to Al-Lazikani et al. (1997) to the variable heavy chain domain of the SM3 (Dokurno et al., 1998), CTM01 (Banfield et al., 1997) and M12-VH domain. CDR lengths are given in brackets. No canonical forms have been assigned for H3 loop.

<table>
<thead>
<tr>
<th>CDR loop</th>
<th>H1</th>
<th>H2</th>
<th>H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM3</td>
<td>1(5)</td>
<td>4(17)</td>
<td>(6)</td>
</tr>
<tr>
<td>CTM01</td>
<td>1(5)</td>
<td>2(15)</td>
<td>(11)</td>
</tr>
<tr>
<td>M12-VH</td>
<td>1(5)</td>
<td>3(16)</td>
<td>(7)</td>
</tr>
</tbody>
</table>

conserved residues at these positions (Figure IV-3). Therefore, it is clear that the V\textsubscript{L}-interface of M12-V\textsubscript{H} domain has a more hydrophobic surface, which needed to be modified through camelization to reduce the possible insolubility of the single domain antibody.

The SM3-Fab antibody structure was also reported to possess a number of regions with absent electron density (Dokurno et al., 1998). The M12-V\textsubscript{H} domain also lacks interpretable electron density from H40-H43. Depending upon the V\textsubscript{H}/V\textsubscript{L} interface comparison, it was found that residue H39 is directed slightly inward in M12-V\textsubscript{H}. Residue H37 has a different orientation in the M12-V\textsubscript{H} while Trp-H103 is more solvent exposed in the M12-V\textsubscript{H} domain. The main-chain conformation is almost similar in both the V\textsubscript{H}’s (Figure IV-4).

In order to test if the determined M12-V\textsubscript{H} domain could bind the antigen in a similar way as the SM3 antibody, we performed a detailed analysis in respect to the bound peptide in the SM3 structure.
The SM3-V_H was superimposed on the M12-V_H structure and the peptide coordinates extracted (Figure IV-5). In order to relieve initial close contacts, two water molecules (7, 8) were deleted.

The comparison between SM3 and M12-V_H domain revealed that the residues of the SM3-V_H domain, which strongly contributes towards the antigen binding have some conformationally equivalent residues in M12-V_H domain. These conformationally equivalent residues might contribute towards the MUC1 binding in the case of the M12-V_H domain and the SM3-V_H domain. In the SM3-Fab, most of the contacts with MUC1 peptide were provided by the heavy chain (64%). Significant numbers of interactions are contributed by Asn-H31, Tyr-H32 and Trp-H33 of the SM3-VH domain. The conformational equivalent of these residues in M12-V_H domain are Asn-H31 (for Asn-H31), Arg-H94 (for Tyr-H32) and Trp-H96 (for Trp-H33). The conformational equivalency means that all these residues are sequentially distant in the SM3-V_H and the M12-V_H domains but their spatial occupancy is almost the same in both cases. In the SM3-Fab-MUC1 complex structure, 66% of the V_H/V_L interaction surface for SM3 comes from residues in the heavy chain, while 33% comes from residues of the light chain.

In the SM3 structure, the CDR-H2 residues involved in antigen binding either through van-der-Waals interaction or through hydrogen bonding. The hydrogen bonding is either direct or mediated by water molecules. Interestingly, we observed a cluster of water molecules around the CDR-H2 residues in M12-V_H domain, which could play an important role in water mediated antibody-antigen binding. This network is observed only for CDR-H2 and not for the other CDR-regions.
IV.7 Camelization of the M12-V<sub>H</sub> domain

Minimizing the size of antigen-binding proteins to a single immunoglobulin domain has been one of the primary goals of antibody engineering. It was demonstrated previously that small antibody domains perform better *in-vivo* than whole antibodies or Fab's (Yokota *et al*., 1992). More often it was observed that the V<sub>H</sub> domain expression in bacteria is hampered by their insolubility and instability (Boss *et al*., 1984 & Cabilly *et al*., 1984). The discovery of camelid heavy chain antibodies (V<sub>H</sub>H, Hammers-Castermann *et al*., 1993) opened up new opportunities for the development of soluble and single domain antibodies by incorporating the structural features of V<sub>H</sub>H antibodies into human V<sub>H</sub> frameworks. Camelization of

human V<sub>H</sub> is promising for the generation of small antigen-binding domains that should be useful for therapeutic purposes in humans. The three prominent differences between human V<sub>H</sub> and camelids V<sub>H</sub>H domains observed are Gly44Gln, Leu45Arg and Trp47Gly. These residues are involved in V<sub>H</sub>/VL interactions in human antibodies. Camelids V<sub>H</sub>H domains have more hydrophilic residues at these positions to make them more soluble. Sequence and structural homology of camel V<sub>H</sub>H and M12-V<sub>H</sub> shows high degree of similarity (80.65%, Figure IV-6). M12-V<sub>H</sub> belongs to the V<sub>H3</sub> family, which is closest to the camel V<sub>H</sub> gene family among both human and murine V<sub>H</sub> genes family.

The X-ray structures have shown that in addition to residues H44, H 45 and H47, several others at positions scattered throughout the V<sub>H</sub>H surface also contribute to the V<sub>H</sub>H solubility (Spinelli *et al*., 1996). On sequence comparison, the residues like Ala-H14, Gly-
H35, Phe-H37, Ala-H49, Tyr-H59, Thr-H77, Val-H78, Pro-H84, Glu-H108 have been found common in all V_H but none of them is present in the M12-V_H. These residues are not considered for camelization because on one hand the effect of these framework residues on the affinity and solubility of the camelid V_H has not been established and on the other hand some features, inherent in the human V_H, contributes to their solubility (Reiter et al., 1999).

Figure IV-7 Ribbon representation of the structurally aligned M12-V_H domain (blue) over camel V_H domain (bluish green, PDB code – 1HCV). N: N-terminal, C: C-terminal of the protein. H1, H2 & H3 represents the three CDR loop regions. The r.m.s. deviation is 1.863. The M12-V_H domain had 80.65% amino-acid sequence similarity with the camelid V_H domain.

The modification in the V_L interface is required to abolish non-specific binding of the single V_H domains and to reduce the tendency for the formation of dimers because of the hydrophobic nature of the residues that compose the V_H/V_L interface. Therefore, the camelization of the M12-V_H domain is mimicked by the mutations Gly44Gln, Leu45Arg and Trp47Gly/His to reduce the most probable problem of solubility and stability in bacterial expression and purification. The camelization leads to the expression of the soluble M12-V_H domain.

**IV.8 Comparison between the M12-V_H domain and the camelized human V_H-P8 antibody**

In the PDB database, only 8 structures of single domain heavy chain antibodies are reported. Most of these structures are solved for camelid V_H domains either alone or in complex with ligand like lysozyme. Only one NMR-solved structure of an isolated camelized (Gly44Gln, Leu45Arg and Trp47Gly) human V_H antibody is found in the database (PDB code - 1VHP,
Riechmann, L., 1996). The structure of these two antibodies was compared to find out the introduction of structural modifications as a result of camelization. Both of these antibodies domains belong to the human germline V\textsubscript{H}3 family. Both the V\textsubscript{H}’s have a high sequence similarity (74.33\%). The r.m.s. C\textalpha deviation between the two structures is 3.281. CDR loops have variable length between the two V\textsubscript{H}’s and only the CDR-H1 has sequence similarity with a difference of two residues at position 31 and 32. Since the CDR-H2 and CDR-H3 are entirely different in sequence, therefore, it is not informative to compare them. The M12-V\textsubscript{H} has ten beta-strands, while the V\textsubscript{H}-P8 has only 9 beta-stands constituting two beta sheets (Figure IV-8). The V\textsubscript{H}-P8 strands C, D and G are smaller in size in comparison to the M12-V\textsubscript{H} domain. V\textsubscript{H}-P8 has beta-bulge in strands C’ and G while no beta-bulge is detected in the M12-V\textsubscript{H} structure. Non-aligned regions are terminal residues, residue H5 (framework region I), H31 & H32 (CDR-H1), H47 to H52a, H56 and H59 (base of CDR-H2), H74 & H77 (framework region III), H95 to H100 (CDR-H3) and H110 (framework region IV). The CDR-H3 loop in the M12-V\textsubscript{H} is oriented towards the V\textsubscript{L} interface while in V\textsubscript{H}-P8 it is directed towards the CDR-H1 and CDR-H2 loops. The important difference in CDR-H3 loop of the two V\textsubscript{H}’s is the absence of a conserved salt bridge (Arg-H94 & Asp-H101) in V\textsubscript{H}-P8 at the base of CDR-H3 loop.

The Glu-H44 and Arg-H45 are solvent exposed in the V\textsubscript{H}-P8 while they are directed towards the V\textsubscript{L} in the M12-V\textsubscript{H}. The Val-H37 present at the bottom of a hydrophobic pocket
in the $V_L$ associated $V_H$ domains while in $V_H$-P8; Val-H37 is buried in the interior as a result of distortion in strand C. The NH group of residue H39 in $V_H$-P8 does not form a hydrogen bond with the backbone oxygen atom of residue H89 in strand F, which is normally present in the $V_L$ associated $V_H$ domains. In $V_H$-P8, a commonly observed hydrogen bond between residue H46 and Arg-H38 is missing.

Recently, the structure of a camelized human $V_H$ domain (PDB code- 1OL0) has been described (Dottorini et al., 2004). The structure has a mutation Val37Phe in addition to Gly44Gln, Leu45Arg and Trp47Gly mutations of VH-P8. This mutation has resulted in a return of the flip observed for residues H44, H45 and H47 in VH-P8 to their conventional orientation as hypothesized by Desmyter et al., (1996). Therefore, the four mutations (Val37Phe, Gly44Gln, Leu45Arg and Trp47Gly) are minimally required to reduce the structural constraint during camelization of human $V_H$ domains. Till now, the effect of these mutations on the affinity and avidity of the human $V_H$ domains and their immunological effects has not been established. The functionality and immunological assessment of native and camelized M12-$V_H$ domain can contribute in this direction. The native and camelized M12-$V_H$ domain does not have Val37Phe mutation.

**IV.9 M12-$V_H$ domain cloning**

The native and camelized M12-$V_H$ domain was amplified using the appropriate set of primers (see appendix). The forward primer were designed to bind the first framework region of the heavy chain while the backward primer annealed to the fourth framework region of the heavy chain of the scFvM12 antibody. For the camelization purpose the amplified native $V_H$ gene segment of the scFvM12 was used as a DNA template. Priming takes advantage of sequence conservation of the framework region. The primers had restriction sites for the NdeI and XhoI restriction enzymes in the overhangs that were incorporated in the amplified product for cloning into the pK3C expression vector. Only few degenerate primers were required to clone the majority of the variable region, which is of practical importance (Orlandi et al., 1989). However, the disadvantage of this approach is that it might introduce amino acid substitutions in the framework region, which can affect antibody affinity.

The annealing temperature used for amplification was slightly raised to avoid non-specific amplification of the M12-$V_H$ genes. The amplified products were tested through PCR amplification and automated sequencing. PCR amplification of M12-$V_H$ gene segment and its sequencing was based on the use of specific primers designed from the M12-$V_H$ gene
segment of the scFvM12. To avoid DNA polymerase copying errors, colonies were randomly screened using the gene specific oligonucleotide probes.

For soluble antibody expression, five colonies of each construct were sequenced. The protein expression was induced after transforming the recombinant plasmid into *E. coli* BL21 (DE3). The expression of the protein was tested through western blot analysis (Figure III-49). The protein expression was found to be *E. coli* strain dependent (Figure III-49). Western blot analysis confirmed the presence of M12-V<sub>H</sub> domain (~12kDa) in the cytoplasm of the bacteria was found at variable levels.

The native and camelized M12-V<sub>H</sub> domain was expressed in the cytoplasm of *E. coli* BL21 (DE3) strain. Their cytoplasmic expression leads to the formation of inclusion bodies. The parameters related to the formation of the inclusion bodies are charge average, turn-forming residue fraction, cysteine fraction, proline fraction, hydrophilicity and the total number of residues (Wilkinson *et al*., 1991). The first two parameters are strongly correlated with inclusion body formation, while the last four parameters show a weak correlation. The measures used to circumvent the formation of inclusion bodies include the growth of bacterial cultures at lower temperatures (Shirano *et al*., 1990), the selection of different strains (Kenealy *et al*., 1987) etc. The change of strain to *E. coli* BL21 (DE3) codon plus resulted in the expression of soluble cytoplasmic protein.

**IV.10 Biological implications of single domain antibody**

It has been envisaged that the reduced size of the V<sub>H</sub> domains enables them to penetrate faster and better into tissues. The bio-distribution and renal clearance is faster (Yokota *et al*., 1992). These properties, however, must be tested further by *in-vivo* studies as recent observations indicate a balance between size and tissue (or tumor) localization rate. Using recombinant forms of antibody domains it was demonstrated that there is a significant trade off in which localization rate (% injected dose localized per gram of tumor) typically declines with decreasing size of antibody domain (Adams *et al*., 1998). V<sub>H</sub> domains can be labeled with radionuclides, fluorescent probes, or other detection markers in the same way that antibody domains are being labeled. Fusion proteins can be constructed with V<sub>H</sub> domains (with reporter proteins, fluorescent proteins, toxins, etc.), as well as coupling to various biological agents (Reiter & Pastan 1998 and Haard *et al*., 1998). However, the utility of these molecules is dependent on a high yield, stability and solubility and improved affinity towards the antigen. The solubility can be improved by camelization and affinity improvements can be achieved by the randomization of selected residues followed by
selection rounds via the construction of large libraries. Such large libraries were constructed for scFv domains of antibodies by recombining large $V_H$ and $V_L$ repertoires (Griffiths et al., 1994). The size of the CDR-H3 loop in the library can be enlarged and combined with codon based mutagenic oligonucleotides that can be used to avoid stop codons or cysteine residues within the CDR.

IV.11 Conclusion

Although, $V_H$ domains are well recognised as part of Fv or Fab fragments, the knowledge about the structural aspects of isolated $V_H$ domain alone in the absence of their $V_L$ partner is almost negligible. The probable reason could be their poor solubility as a result of non-specific interaction through the $V_L$ interface (Kortt et al., 1995). Extensive structural information is required to optimise $V_H$ domains for their use as small recognition units in general and in cancer therapy. So far, only the structural details of the isolated camelized human $V_H$ domain is available (Riechmann, L. 1996 and Dottorini et al., 2003). The first detailed three-dimensional structure of non-camelized human $V_H$ domain can contribute in the following directions:

1. The structure function relationship between human single domain antibodies and its antigen is of profound importance especially in cancer therapy where tumor penetration is desirable. The M12-$V_H$ domain provides the opportunity to study this kind of relationship against MUC1, providing the basis for improvement of antibody affinity through display technology like phage or ribosomal display.

2. The structure will be helpful in the design of small peptide antigens, which can be used to raise the cellular immune response against cancerous cells overexpressing MUC1.

3. Although the camelization of $V_H$ domains has been proved to increase the solubility of camelized $V_H$ domains, still there is no consensus regarding the effect of camelization on the affinity and avidity of the antibody. The comparative functionality study of non-camelized and camelized M12-$V_H$ domain can contribute in this direction. Also there is no evidence for the influence of camelization on the immunogenicity of human $V_H$ domains. Camelized M12-$V_H$ can be used as a basis to study the immunogenic effects of camelization of human $V_H$ domains.

4. The structural study of camelized M12-$V_H$ domain in complex with the MUC1 antigen can laid down the foundation for a rational single domain antibody design.
5. By comparing the structure of the scFvM12 in complex with MUC1 antigen and scFvM12 alone could provide the structural basis for the contribution of the two domains generated through random combinatorial libraries in antigen recognition. Furthermore, the $V_H/V_L$ interaction could be improved by rational protein design to facilitate the formation of multimeric entities.
V REFERENCES


123


138


V-REFERENCES


Publication out of this thesis


PDB entry out of this thesis

PDB code – 1T2J
VI APPENDIX

VI.1 scFvM12 sequence

\[
\begin{align*}
\text{QVQLQESGGGLVQPGGSLRLSCAASGFTFS} & \quad \text{NSAMS} & \quad \text{WVRQAPGKGLEWVS} & \quad \text{SIGSGGNTYSADSVK} \\
\text{FR1-VH} & \quad \text{CDR1-VH} & \quad \text{FR2-VH} & \quad \text{CDR2-VH}
\end{align*}
\]

\[
\begin{align*}
\text{RFTISRNASLKLNSLRAEDTAVYCAR} & \quad \text{DWYMDV} & \quad \text{WQQTTVTVSS} & \quad \text{GGGSGGGGGGSSGGS} \\
\text{FR3-VH} & \quad \text{CDR3-VH} & \quad \text{FR4-VH} & \quad \text{LIGNER}
\end{align*}
\]

\[
\begin{align*}
\text{SYVLTQPAVSVALGQTVRTC} & \quad \text{QDDSRGYGYS} & \quad \text{WYQQKPGQAPVLVIY} & \quad \text{AKTNRPB} \\
\text{FR1-VL} & \quad \text{CDR1-VL} & \quad \text{FR2-VL} & \quad \text{CDR2-VL}
\end{align*}
\]

\[
\begin{align*}
\text{GIPDRFSSSGGNTASLITGQAEDADYYC} & \quad \text{NSRDNSTGTHLEV} & \quad \text{FGGTKLTVLGAAAA} \\
\text{FR3-VL} & \quad \text{CDR3-VL} & \quad \text{FR4-VL}
\end{align*}
\]

\[
\begin{align*}
\text{EQKLISEDLNAGAA} & \quad \text{HHHHHH} \\
\text{cMYC-TAG} & \quad \text{HIS-TAG}
\end{align*}
\]

Figure VI-1 scFvM12 antibody amino acid sequence. CDR’s are underlined.

VI.2 DNA sequence of M12-V domain

\[
\text{PstI}
\]

\[
\begin{align*}
\text{cgagtgcaagctcgaggaagccccgaggtttgtacagcctgccccgaggttg} & \quad \text{ctgaagcccaggttt}
\end{align*}
\]

\[
\begin{align*}
\text{caggtgcagctgcaggagtcggggggaggcttggtacagcctggggggtccctgag} & \quad \text{actgctgaagcccaggtt}
\end{align*}
\]

\[
\begin{align*}
\text{gtccacgtcagctctccagcccccccctccgaaccatgtcgagccccccagggact} & \quad \text{ctgagctcagctcag}
\end{align*}
\]

\[
\begin{align*}
\text{QVQLQESGGGLVQPGGSLRLSCAASGFTFS} & \quad \text{NSAMS} & \quad \text{WVRQAPGKGLEWVS} & \quad \text{SIGSGGNTYSADSVK} \\
\text{FR1-VH} & \quad \text{CDR1-VH} & \quad \text{FR2-VH} & \quad \text{CDR2-VH}
\end{align*}
\]

\[
\begin{align*}
\text{RFTISRNASLKLNSLRAEDTAVYCAR} & \quad \text{DWYMDV} & \quad \text{WQQTTVTVSS} & \quad \text{GGGSGGGGGGSSGGS} \\
\text{FR3-VH} & \quad \text{CDR3-VH} & \quad \text{FR4-VH} & \quad \text{LIGNER}
\end{align*}
\]

\[
\begin{align*}
\text{SYVLTQPAVSVALGQTVRTC} & \quad \text{QDDSRGYGYS} & \quad \text{WYQQKPGQAPVLVIY} & \quad \text{AKTNRPB} \\
\text{FR1-VL} & \quad \text{CDR1-VL} & \quad \text{FR2-VL} & \quad \text{CDR2-VL}
\end{align*}
\]

\[
\begin{align*}
\text{GIPDRFSSSGGNTASLITGQAEDADYYC} & \quad \text{NSRDNSTGTHLEV} & \quad \text{FGGTKLTVLGAAAA} \\
\text{FR3-VL} & \quad \text{CDR3-VL} & \quad \text{FR4-VL}
\end{align*}
\]

\[
\begin{align*}
\text{EQKLISEDLNAGAA} & \quad \text{HHHHHH} \\
\text{cMYC-TAG} & \quad \text{HIS-TAG}
\end{align*}
\]
VI.3 Primers

The following primers (MWG) were used for the amplification of the gene of interest. Primers A & B also has mutations Gly44Glu and Lys45Arg. The codon mutated for camelization with respect to wild type is highlighted.

A. Native M12-VH domain
   Forward - 5’- GGG AAT GTC CAT ATG GCC CAG GTG CAG CTG CAG GAG -3’
   Backward - 5’- GGA CTA CCG CTC GAG ACT CGA CAC GGT GAC CGT GGT C -3’

B. Camelization of M12-VH domain (Trp47Gly)
   Forward - 5’- GGT GTC CGC CAG GCT CCA GGG AAG GAG CCG GAG GGG GTC TCA E R E G
   Backward - 5’-CCA CTA CCA CTG ATA CTT GAG ACC CCC TCC CGC TCC TTC CCT GGA GCC TGG CGG ACC- 3’

C. Camelization of M12-VH domain (Trp47His)
   Forward - 5’- GGT GTC CGC CAG GCT CCA GGG AAG GAG CCG GAG CAC GTC TCA E R E H
   Backward - 5’-CCA CTA CCA CTG ATA CTT GAG ACG TGC TCC CGC TCC TTC CCT GGA GCC TGG CGG ACC- 3’
VI.4 M12-V_{H} domain (1.8Å)

The structure of the M12-V_{H} domain at resolution 1.8Å was solved by Hoffmann, K. M. V. (2001, Bionalytics Group, RWTH Aachen, Germany). The data was collected up to 1.8Å on Rigaku X-ray generator (λ=1.542Å) at 100K. The collected dataset contained the following statistics (Table VI-1).

**Table VI-1 Data collection statistics for M12-V_{H} domain (1.8Å).**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit-cell length (Å)</td>
<td>a = 72.011, b = 38.354, c = 37.573</td>
</tr>
<tr>
<td>Unit-cell angles (°)</td>
<td>α = γ = 90, β = 109.746</td>
</tr>
<tr>
<td>Mathew’s coefficients (Å³ Da⁻¹)</td>
<td>1.98</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>37.8</td>
</tr>
<tr>
<td>Mosaicity</td>
<td>0.8</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>103773.240</td>
</tr>
<tr>
<td>Number of molecules in the asymmetric unit</td>
<td>1</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>46010</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>8576</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.542</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
<td>1.8</td>
</tr>
<tr>
<td>Highest resolution shell (Å)</td>
<td>1.86 - 1.80</td>
</tr>
<tr>
<td>Observed completeness overall</td>
<td>(94.4%)</td>
</tr>
<tr>
<td>high-resolution shell</td>
<td>(87.8%)</td>
</tr>
<tr>
<td>( R_{\text{merge}} ) overall</td>
<td>(0.031)</td>
</tr>
<tr>
<td>high-resolution shell</td>
<td>(0.141)</td>
</tr>
</tbody>
</table>

**Table VI-2 Data statistics breakdown for the M12-V_{H} domain (1.8Å) dataset with resolution.** The values correspond to the X-ray wavelength (1.542Å). \*\( R_{\text{merge}} \) is defined as \( \frac{100\sum_{\text{hkl}}|<I|-I_{n}|}{\sum_{\text{hkl}}I_{hkl}} \), which is summed over all reflections, where \(<I>\) is the mean intensity of the reflection hkl and \(I_{n}\) is the intensity of the nth observation of a reflection hkl. \*\( I/\sigma \) is defined as \(<I>\) divided by the average error on I.

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>Chi²</th>
<th>( R_{\text{merge}} ) (%)*</th>
<th>( I/\sigma )*</th>
<th>Completeness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.00 - 3.88</td>
<td>0.887</td>
<td>1.9</td>
<td>37.74</td>
<td>97.9</td>
</tr>
<tr>
<td>3.88 - 3.08</td>
<td>0.997</td>
<td>2.2</td>
<td>37.70</td>
<td>97.6</td>
</tr>
<tr>
<td>3.08 - 2.69</td>
<td>1.002</td>
<td>2.8</td>
<td>31.94</td>
<td>96.9</td>
</tr>
<tr>
<td>2.69 - 2.44</td>
<td>0.915</td>
<td>3.6</td>
<td>23.91</td>
<td>96.1</td>
</tr>
<tr>
<td>2.44 - 2.27</td>
<td>0.936</td>
<td>4.3</td>
<td>21.07</td>
<td>95.1</td>
</tr>
<tr>
<td>2.27 - 2.13</td>
<td>0.840</td>
<td>4.8</td>
<td>18.15</td>
<td>94.2</td>
</tr>
<tr>
<td>2.13 - 2.03</td>
<td>0.900</td>
<td>6.5</td>
<td>14.62</td>
<td>93.5</td>
</tr>
<tr>
<td>2.03 - 1.94</td>
<td>0.983</td>
<td>8.0</td>
<td>12.31</td>
<td>92.4</td>
</tr>
<tr>
<td>1.94 - 1.86</td>
<td>1.020</td>
<td>10.9</td>
<td>9.06</td>
<td>92.0</td>
</tr>
<tr>
<td>1.86 - 1.80</td>
<td>1.020</td>
<td>13.3</td>
<td>6.74</td>
<td>87.8</td>
</tr>
<tr>
<td>All reflections</td>
<td>0.948</td>
<td>3.1</td>
<td>27.38</td>
<td>94.4</td>
</tr>
</tbody>
</table>
The overall quality of the high-resolution dataset was evaluated by using the standard data quality indicators calculated per resolution shell (Table VI-2). The chi^2 is close to 1 for all the resolution shells, therefore, the statistical distribution of the reflections is homogenous in the resolution shells. The intensity and the R_merge value of the data were found to be good in the high-resolution shell.

The phase information for the M12-V_H domain was determined by using the theoretical search model 1H0U (Hougs et al., 1999). The sequence of the theoretical search model 1H0U was appropriately trimmed to increase the homology with the target M12-V_H fragment. The search model has 89.655% sequence identity with the target sequence as shown in figure. The orthogonal co-ordinates of the model were converted into fractional co-ordinates by using the pearl script.

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>M12-V_H</td>
<td>QVQLQESGGLVQPGGSLRLSCAASGFTPSANSWVRQAPGKLEWVS1GISGGNTYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1H0U-V_H</td>
<td>EVQLLESGGGLVQPGGSLRLSCAASGFTPSYMSWVRQAPGKLEWVAISGGSTYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>M12-V_H</td>
<td>ADSVKGRFTLRENKNSLTVQMSLRAEDTAVYYCACHDWYGMDVGQGTTTVSSGGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1H0U-V_H</td>
<td>ADSVKGRFTLRENKNSLTVQMSLRAEDTAVYYCACHYGMDVGQGTTTVSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure VI-3 Sequence similarity between M12-V_H and 1H0U-V_H domains.**

Molecular replacement was performed with AMoRe (Navaza, J., 1994) using reflections from 20.0-3.0Å. As a result of the rotation search, a single clear solution (α =362.52°, β = 0.00°, γ = 0.00°) having the highest correlation coefficient 21.5% and R_f 52.2% was obtained. The correct solution generated as a result of rotation function is translated and the translation function found to have the coefficients as T_x = 0.3409, T_y =0.0000, T_z =0.2708 with improvement of correlation coefficient of 40.0% and R_f 46.3%. This translated solution was fitted to the search model, which resulted in α =150.0°, β = 49.28°, γ = 57.49° and T_x = 0.3377, T_y =0.0000, T_z = 0.2746 with further improvement of correlation coefficient 54.1% and R_f dropped to 41.1%. Rigid body refinement led to a final R-factor 40.0% and a correlation coefficient of 55.5%.

The model was built manually by using the program ‘O’ (Jones et al., 1991). The initial model has R_cry/R_free factor 47.61(%)/46.65(%). The partial model was refined with in the resolution range 8.0-2.0Å. The residues 40-43, 52-56, 93-107 and 112-113 with poor electron density were removed. 10% of the reflections were kept for calculating the R_free factor. Cycles of manual model building in ‘O’ and refinement through CNS resulted in model having the R_cry and R_free values 23.54 and 28.69. The refinement statistics is given in
Table VI-3. The refined model has 683 atoms and 90 residues. Finally 80 water molecules were added to the structure.

**Table VI-3 Summary of structure refinement statistics for M12-VH domain (1.8Å).**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Refinement resolution (Å)</td>
<td>8.0 – 2.0</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>5609</td>
</tr>
<tr>
<td>After simulated annealing</td>
<td></td>
</tr>
<tr>
<td>$R_{crys}$ (%)</td>
<td>23.54</td>
</tr>
<tr>
<td>$R_{free}$ (%)</td>
<td>28.69</td>
</tr>
<tr>
<td>Number of unique reflections used for $R_{free}$</td>
<td>651</td>
</tr>
<tr>
<td>Final $R_{crys}$ (all data; %)</td>
<td>23.34</td>
</tr>
<tr>
<td>Final $R_{free}$ (%)</td>
<td>28.60</td>
</tr>
<tr>
<td>$I &gt; 2\sigma$</td>
<td></td>
</tr>
<tr>
<td>r.m.s.d. bond length (Å)</td>
<td>0.00548</td>
</tr>
<tr>
<td>r.m.s.d. bond angles (°)</td>
<td>1.30548</td>
</tr>
<tr>
<td>r.m.s.d. torsion angle (°)</td>
<td>1.899</td>
</tr>
</tbody>
</table>

In overall appearance, the M12-VH has main-chain resemblance to other human VH fragments. The residue numbering scheme and CDR definitions follow the standard Kabat conventions (Kabat et al., 1991). A luzzati plot suggests a mean positional error of 0.24Å (Luzzati, P. V., 1952) while the $\sigma_A$ error estimate is 0.293 Å (Read, R. J., 1986). The model exhibits good stereogeometry that is consistent with 1.8Å resolution data and it meets or exceeds all main chain and side chain tests of PROCHECK (Laskowski et al., 1993). The VH molecule displayed the well-characterized immunoglobulin fold, in which VH consists of two twisted, antiparallel β-sheets packed tightly against each other.

**VI.5 Intermolecular crystal contacts for M12-VH domain (1.5Å)**

VI.6 Hydrogen bonding pattern of M12-V<sub>H</sub> domain (1.5Å)

Table VI-4 Intra-molecular hydrogen bonding in M12-V<sub>H</sub> domain (1.5Å). The N and C alphabets indicate the amino Nitrogen and Carbonyl oxygen atoms of amino acids. The OD, OE, OG indicates the delta, epsilon and gamma carbonyl oxygen atom of side chain of amino acids. The numbers in brackets indicate the bond distance. The residues are numbered according to Kabat et al. (1977).

<table>
<thead>
<tr>
<th>Main chain to Main Chain</th>
<th>Main chain to Side Chain</th>
<th>Side chain to Main chain</th>
<th>Side chain to Side Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>25:N 3:O (2.97)</td>
<td>106:N 6:OE1 (2.83)</td>
<td>32:OG 33:O (2.98)</td>
<td>17:OG 2A:OD1(3.00)</td>
</tr>
<tr>
<td>23:N 5:O (2.97)</td>
<td>9:N 107:OG1 (2.97)</td>
<td>71:NH1 32:O (2.89)</td>
<td>38:NH1 86:OD1(2.93)</td>
</tr>
<tr>
<td>7:N 21:O (2.83)</td>
<td>98:N 35:OG (2.79)</td>
<td>71:NH2 32:O (2.95)</td>
<td>38:NH1 90:OH (2.86)</td>
</tr>
<tr>
<td>21:N 7:O (2.85)</td>
<td>74:N 72:OD1 (2.98)</td>
<td>35:OG 95:O (2.75)</td>
<td>38:NH2 46:OE1 (2.97)</td>
</tr>
<tr>
<td>10:N 108:O (2.91)</td>
<td>83:N 86:OD2 (2.83)</td>
<td>43:N 41:O (2.94)</td>
<td>46:OE2 62:OG (2.73)</td>
</tr>
<tr>
<td>12:N 110:O (2.90)</td>
<td>111:N 87:OG1 (2.90)</td>
<td>50:OG 96:O(2.50)</td>
<td>64:NZ 61:OD1(2.94)</td>
</tr>
<tr>
<td>112:N 12:O (2.74)</td>
<td>95:N 95:OD1 (2.72)</td>
<td>71:NH1 52:O (2.88)</td>
<td>66:NH1 86:OD2(2.75)</td>
</tr>
<tr>
<td>16:N 13:O (2.92)</td>
<td>99:N 95: OD2 (2.88)</td>
<td>66:NH1 82B:O (2.87)</td>
<td>71:NE 73:OD1(2.74)</td>
</tr>
<tr>
<td>15:N 82C:O(2.73)</td>
<td>17:OG 2A:OD1(3.00)</td>
<td>66:NH2 62:O (2.91)</td>
<td>103:NE1 101:OD1(2.81)</td>
</tr>
<tr>
<td>82C:N 16:O (2.98)</td>
<td>82:N 82:O (2.90)</td>
<td>73:ND2 52A:O (2.85)</td>
<td>82B:OG 82B:OG (2.92)</td>
</tr>
<tr>
<td>18:N 82:O (2.90)</td>
<td>82:N 18:O (2.75)</td>
<td>77:OG 75:O(2.75)</td>
<td></td>
</tr>
<tr>
<td>20:N 80:O (2.79)</td>
<td>80:N 20:O (2.93)</td>
<td>82B:OG 82B:O (2.96)</td>
<td></td>
</tr>
<tr>
<td>80:N 20:O (2.93)</td>
<td>22:N 78:O (2.74)</td>
<td>90:OH 86:O (2.62)</td>
<td></td>
</tr>
<tr>
<td>22:N 78:O (2.74)</td>
<td>78:N 22:O (2.88)</td>
<td>93:N1 101:O (2.91)</td>
<td></td>
</tr>
<tr>
<td>24:N 76:O (3.00)</td>
<td>76:N 76:O (2.95)</td>
<td>112:OG 112:O (2.94)</td>
<td></td>
</tr>
<tr>
<td>32:N 29:O (2.99)</td>
<td>32:N 29:O (2.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34:N 51:O (2.96)</td>
<td>35:N 93:O (2.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51:N 34:O (2.87)</td>
<td>35:N 93:O (2.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35:N 93:O (2.88)</td>
<td>35:N 93:O (2.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39:N 89:O (2.79)</td>
<td>39:N 89:O (2.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89:N 39:O (2.91)</td>
<td>43:N 40:O (2.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43:N 40:O (2.89)</td>
<td>60:N 48:O (2.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60:N 48:O (2.87)</td>
<td>50:N 58:O (2.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50:N 58:O (2.85)</td>
<td>58:N 50:O (2.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58:N 50:O (2.94)</td>
<td>52:N 56:O (2.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52:N 56:O (2.94)</td>
<td>54:N 52:O(2.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54:N 52:O(2.96)</td>
<td>66:N 63:O (2.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66:N 63:O (2.96)</td>
<td>68:N 81:O (2.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68:N 81:O (2.81)</td>
<td>81:N 68:O (2.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81:N 68:O (2.86)</td>
<td>79:N 70:O (2.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79:N 70:O (2.80)</td>
<td>54:N 52:O (2.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54:N 52:O (2.97)</td>
<td>64:N 60:O (2.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64:N 60:O (2.89)</td>
<td>72:N 77:O (2.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72:N 77:O (2.79)</td>
<td>76:N 72:O (2.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76:N 72:O (2.97)</td>
<td>86:N 83:O (2.86)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VI.7 Structural studies of macromolecules

VI.7.1 Overview

The knowledge of accurate molecular structure is a prerequisite for the study of structure-function relationship, which provides the rational basis for the development of effective therapeutic agents and drugs. There are number of experimental and theoretical methods that can provide information about a macromolecular assembly structure. The experimental methods used for structural characterization of macromolecular assemblies are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron crystallography, electron tomography, immunoelectron microscopy, fluorescence resonance energy transfer (FRET), mass spectroscopy and protein arrays (Sali et al., 2003). Out of the various methods, X-ray crystallography is the most powerful method for structure determination because it is capable of providing an atomic structure of the whole macromolecular assembly (Ban et al., 2000). When suitable crystal and high-resolution crystallographic data are obtained, there is little need for other methods of structure determination.

Recently, the application of NMR spectroscopy has expanded to structure determination of increasingly large subunits and even their complexes (Fiaux et al., 2002). The technique is still restricted for small molecular weight proteins (up to 30kDa) and very time consuming. The main advantage with NMR spectroscopy is its utility for dynamic studies and can provide useful information about the protein in solution. In contrast to NMR spectroscopy, no size limitation exists for the molecule or complex to be studied in X-ray crystallography. Crystallography can reliably provide the answer to many structure related questions, from global folds to atomic details of bonding.

The high accuracy of crystallographic structure requires the growth of good crystals of a macromolecule either alone or in complex with its ligand. Protein crystals differ from the inorganic crystals in terms of size and solvent content. The high solvent content of the protein crystals makes them fragile to handle.
VI.7.2 Methodology

The techniques used for growing the crystals are vapor diffusion, batch, microbatch, microdialysis, free interface diffusion, macro and microseeding. Out of the various techniques, vapor diffusion and microbatch are the most commonly used methods because they are easy to perform, require a small amount of sample, and allow sufficient flexibility during screening and optimization.

The principle involved in the vapor diffusion is that the initial concentration of the precipitating agent in the drop is lower than the well buffer and as an equilibration is reached slowly, the respective concentration of the precipitating agent increases resulting a supersaturation stage. Supersaturation is a metastable state. A supersaturated solution will give eventually rise to nuclei some of which will grow into larger crystals. Supersaturation must be slowly approached to minimize nucleation, so that a small number of larger crystals are obtained. Solutions must be clean to avoid the chance of heterogeneous nucleation. Crystallization is a multi-parameter problem, which is not entirely under rational control. Trial-and-error variation of a multitude of biochemical and biophysical parameters can be used to decrease the solubility of the protein, so that the solution becomes supersaturated. Also the crystallization technique itself may be a parameter.

Vapor diffusion can be achieved either by using the sitting or hanging drop methodology. In sitting drop technique (Figure VI-4A) one places a small (1 to 15µl, in high throughput robotic systems 0.1-0.2µl) droplet of the sample mixed with an equal volume of precipitating agent on a platform in vapor equilibration with the well buffer. The advantage of the sitting drop technique includes speed and simplicity. The disadvantages are that crystals can sometimes adhere to the sitting drop surface making their collection difficult.

![Figure VI-4 Sitting drop (A) and hanging drop (B) crystallization techniques](image)

This disadvantage can turn into an advantage where occasionally the surface of the sitting drop can assist in nucleation. The sitting drop is an excellent method for screening and optimization.
In the hanging drop technique (Figure VI-4B); the small (1 to 10 µl) droplet of the sample mixed with an equal volume of reservoir buffer solution was kept on a siliconized glass cover slide. Then, the slide gently inverted over the reservoir. The advantages of the hanging drop technique include the ability to view the drop through glass without the optical interference from plastic, reduced chance of crystals sticking to the hardware, and easy access to the drop. The disadvantage is that a little extra time is required for setting up the hanging drops.

Once good crystals are grown, they are checked for their optical and biochemical quality. Some protein crystals can be kept for years and will still diffract; others do not last. It may only be possible to keep the crystals in the mother liquor in which they grew. If possible it is preferable to harvest the crystals in a stabilization buffer, as this will allow easier handling when doing heavy metal soaks. It is usually not possible to use fresh mother liquor for harvesting as under the equilibrium conditions the crystals will dissolve or get damaged. Increasing the concentration of precipitant in the mother liquor by several percent can circumvent this problem. The best harvesting solution will be found by trial-and-error.

**VI.7.3 Data collection**

For data collection, the crystals are either mounted in capillary or flash cooled at 100K. Before mounting, the crystal is washed with fresh stabilization buffer to remove any remaining protein solution. Stabilization buffer (or alternatively mother liquor) and a crystal are drawn into a capillary tube by capillary action along with very small amount of reservoir solution to keep the crystal moist. This may be quite difficult in case of highly viscous solutions. The capillary tube will usually be a mounting tube. The data from a crystal mounted in capillary is collected at room temperature and therefore the crystal is susceptible to radiation damage on one side and results in enhanced background due to the presence of tiny amount of mother liquor around the crystal. To avoid the radiation damage of the crystal, the capillary mounting is now less common. In the flash cooling method, ice formation can be largely suppressed by the use of a cryo-protectant (e.g. Glycerol, PEG, MPD etc.), in which the crystal is dipped immediately before mounting. The crystal is held by surface tension in a thin film of containing-containing solution formed within a small loop made from mohair or rayon fibers. If a loop of the correct size is used there will be little mechanical stress and a large surface area will be exposed allowing high cooling rates. Getting the crystal in the loop is tricky and is usually done under the microscope. The crystal is gently pushed to the surface and then scooped up and pulled through the meniscus. The
loop is then rapidly mounted on a goniometer in the beam of a liquid nitrogen gas delivery system. The flash cooled crystal is stored in liquid nitrogen till the time of data collection.

When the crystal exposed to X-rays, most of the x-rays travel straight through it. Some, however, interact with the electron cloud and undergo diffraction as a result of constructive interference. It is often observed that the microscopic goodness of the crystals is not enough rather the crystal should be well ordered internally to get the strong x-ray diffraction. The diffraction of x-ray beam depends upon the molecules arranged in a regular and repeating array. This repeating arrangement of molecules constitutes the unit cell inside the crystal. The unit cell consists of the asymmetric unit defined by the symmetrical arrangement of the protein molecule. The diffracted X-rays produce a characteristic pattern on a detector as a result of the symmetrical arrangement of the protein molecule inside the crystal. Diffraction of X-rays from the crystal can be regarded as the reflection of the beam by sets of parallel planes called lattice planes. If $\theta$ is the angle between the primary X-ray beam and the lattice plane with interplanar spacing $d$ (Figure VI-5), then the path difference between the two waves scattered in $A$ and $C$, respectively, is $BC+CD = 2d \sin \theta$. If this path difference is a multiple

\[ \text{Figure VI-5 Schematic representation of Bragg's law for x-ray diffraction. Blue dotted lines indicate the plane of atoms in the crystal. Arrow indicates the x-ray beam and its diffraction after interacting with the electron cloud of the atom. D: distance between the planes, } \theta: \text{ reflection angle.} \]

‘n’ of the wavelength ($\lambda$), then we get constructive interference between the two waves. Here, $n$ is the order of diffraction. Consequently, the condition for an X-ray reflection is

\[ 2d \sin \theta = n \lambda \]

which is known as Bragg's law. The angle $\theta$ for which the equation holds is called the Bragg angle. For a given $n$ and $\lambda$ and any other angle of incidence, waves emerging from successive planes are out of phase and interfere destructively.

When the X-rays diffract according to Bragg’s law, they are recorded as distinct diffraction spot. A diffraction image contains many reflections, which can easily overlap. The crystal is
constantly rotated by a small angle as well as oscillating simultaneously during data
collection on each image. It is therefore necessary to collect many images to measure all
unique reflections of a protein crystal.

VI.7.4 Phase problem and related methodology

Since the X-rays are considered as waves possess amplitude and phases. For structure
determination, the knowledge of three parameters, namely, the amplitude, the wavelength
and the phases are required. If both the amplitude and the phases are known, the electron
density, which gave rise to the diffracted waves, can be reconstituted by a Fourier synthesis.
However, in a crystallographic diffraction experiment only the amplitude is recorded. The
amplitude can be obtained from the intensity of the diffraction spot but during diffraction
data collection, the phase information is lost. This is called as the phase problem in
crystallography. Therefore, instead of being able to obtain a direct image of the structure in a
crystal, it is necessary to measure the direction and intensity (or amplitude squared) of each
diffracted X-ray beam, and then reconstitute its phase.

The phase problem can be solved either ab-initio or by using the phase information
from a solved structure.

VI.7.5 Ab-initio phase determination

Phases are determined ab-initio in two ways-
A. either the crystals are "derivatised" by introducing heavy metal atoms. If only a few
metal ions are present in the crystal, the "heavy metal structure" can be determined by a
Patterson analysis. If the crystal essentially stays unperturbed ("isomorphous"), the technique
of multiple isomorphous replacement (MIR, Crick et al., 1956) can exploit the heavy
metal structure to determine phases for the protein.
B. Or crystals are labeled with anomalously scattering atoms like heavy metal atoms. If
only a few anomalous scatterers are present in the crystal, the "anomalous scatterer structure"
can be determined by a Patterson analysis. The small variation in the real and imaginary
components of the scattering factor (f' and f") of an anomalously scattering atom, when the
wavelength of the incident beam approaches one of its absorption edges, can be exploited to
determine the phases of the native structure and this approach is called as multiwavelength
anomalous diffraction (MAD, Lye et al., 1980). Naturally a "tunable" X-ray source such as
a synchrotron is needed for this experiment.
VI.7.6 Molecular replacement

The initial estimate of phases can be extracted from the structure factor of a known protein. Therefore, the structure of the new protein can be determined from a single native data set. The known protein is called as the phasing model and the method which entails calculating the phases by placing a model of the known protein in the unit cell of the new protein, is called the molecular replacement (Rossmann, M.G., 1972). In molecular replacement, the rotation and translation functions are executed to fit the target over search model for determining the Patterson map. The Patterson map provides the phase information. The translocated solutions were subjected to rigid body refinement and then reoriented on the model structure to find the correctness of the molecular replacement solution. The resulting solution is converted into a pdb coordinates. The phase information generated as a result of molecular replacement can be improved by using the PHASES program (Furey et al., 1997).

VI.7.7 Model building and refinement

Initially structure factor amplitude data $|F|$ have been measured. The phase problem has been solved either by Molecular Replacement (MR), Isomorphous Replacement (MIR) or Multiwavelength Anomalous Diffraction (MAD), so that initial "observed" phases are available. An electron density map can now be calculated by a Fourier synthesis. The electron density map provides the template for building the three dimensional protein structures. The quality of the map depends on the resolution of the diffraction data. In general, phases derived from MIR produce electron density map, which is not yet of sufficient quality to uniquely identify and position the whole of the amino acid sequence because the phases are not accurate enough, and do not extend to the full resolution of the amplitude data. Therefore, the quality of the electron density map needs to be improved to build the protein model. Density modification consists of alternating steps of electron density map modification proper (real space) and phase combination (reciprocal space), where the fast Fourier transform is used to convert from real space to reciprocal space and vice versa. Once the quality of the phases has become high enough to calculate an interpretable electron density map, it will be possible to build a partial model in the electron density. In the case of isomorphous replacement (MIR) the initial model will usually comprise main chain atoms only, later to include main chain and side chain atoms. In the case of molecular replacement (MR), the initial model usually comprises both main chain and side chain atoms, and the correct sequence can be built into, to replace that of the model. Any model is completed and improved by removing positional errors by alternating cycles of electron density modeling...
with a computer graphics program (manual) and refinement with a refinement program (automatic), while extending the resolution to the full extent of the native amplitude data.

In refinement, the parameters of the model are altered to improve the agreement between amplitudes calculated from the model and the amplitudes obtained by measurement. Refinement is the minimization of a function, which reflects the discrepancy between these two. The final model coordinates will contain random and systematic errors, which represent the history of their generation. The problem is to determine and correct the errors during the process of modeling and refinement. Once the residual errors have become acceptable and the modeling and refinement process has converged, the crystallographic, stereochemical and structural parameters of the model needs to be validated.

### VI.7.8 Structure validation and submission

The crystallographic parameters are checked by the structure factor amplitudes calculated from the model and measured against the observed amplitudes. Although the conventional R-factor \([R-\text{factor} = \Sigma (F_o - F_c)/\Sigma F_o; F_o \text{ and } F_c \text{ are the observed and calculated structure factor amplitudes}]\) will be helpful, the \(R_{\text{free}}\) factor \([R_{\text{free}} \text{ factor} = \Sigma_{\text{free}} (F_o - F_c)/\Sigma_{\text{free}} F_o, \text{ a small subset of reflection (5-10%) kept aside to calculate } R-\text{factor}]\) calculated for the small wedge of data kept out of the refinement, will be a much better statistical criterion to validate the coordinates. Either a Luzzati plot [plot of R-factor against 1/d (where d=resolution)] or the SigmaA [a plot of \(\ln \sigma_A\) against \((\sin \theta/\lambda)^2; \sigma_A\)-coordinate error] method may estimate the overall coordinate precision. Deviations for the main chain torsion angles from Ramachandran plots give a good indication for the stereochemical quality of the model.

Checking structural parameters (e.g. Hydrogen bonds) will aid in the production of as good a coordinate data set as structural knowledge will permit. These parameters test features, which may not yet be obvious from the current electron density (e.g. incorrectly folded structure), or features, which cannot be determined from electron density maps (e.g. the position of the N and O in the side chain of Asn). They may point to incorrect areas in the model. They may aid in getting His/Gln/Asn side chain atoms in the correct orientation. They will help in positioning water molecules in the most appropriate place.

Once the model is thoroughly checked and found acceptable it has to be submitted to the international protein structure repository called Protein Data base (PDB, [www.rcsb.org]).
VI.7.9 Vector maps

**pK3C**

5532 bp

protease cleavage site 3C E11

NcoI (5462)

BamHI (5451)

EcoRI (5445)

SmaI (5442)

XmaI (5440)

SacI (5427)

SalI (5416)

HindIII (5399)

EagI (5382)

NotI (5382)

XhoI (5364)

His*Tag

T7 P

T7 TERM

LACI

ori

lac

MCS

SaiI

NcoI

NotI

c-myc tag

His6 tag

scFvM12 gene

ApaLI

bla

M13 origin

**pSynIscFvhuMucIM12**

4014bp

**VI-APPENDIX**
VI.8 List of abbreviations

%  percentage
\( \lambda \)  wavelength
\( \sigma \)  sigma
Å  angstrom
A  alanine (Ala)
Ab  antibody
Amp'  ampicillin resistance
AP  alkaline phosphatase
asu  asymmetric unit
\( \beta \)  beta
BCA  bicinchoninic acid
BCIP  5-bromo-4-chloro-3-indolyl phosphate
bp  base pair
BSA  bovine serum albumin
C  cysteine (Cys)
CaCl₂  calcium Chloride
cDNA  complementary deoxyribonucleic acid
CTL  cytotoxic T-Lymphocytes
CV  column volume
D  aspartic acid (Asp)
dNTP  deoxynucleotide triphosphate
DTT  dithiothreitol
E  glutamic acid (Glu)
\( E.\ coli \)  escherichia coli
EDTA  ethylenediaminetetraacetic acid
F  phenylalanine (Phe)
FACS  fluorescence-activated Cell Sorter
\( F_c \)  calculated Structure factor
FDA  food and drug administration
\( F_0 \)  observed structure factor
FPLC  fast flow liquid chromatography
Fv  variable fragment of antibody consists of \( V_H \) and \( V_L \) without linker
G glycine (Gly)
GAM goat anti-mouse antibodies
H histidine (His)
H1, H2, H3 variable heavy chain CDR loops
HAMA human anti-murine-antibody immune response
HEPES N- [2-hydroxyethyl]-piperazine-N0- [2-ethanesulphonic acid]
His histidine
I isoleucine (Ile)
IEF iso-electric focussing
ICAM intercellular Adhesion Molecule
IgG immunoglobulin type gamma
IgM immunoglobulin type mu
Ig immunoglobulin
IL-2 interleukin -2
IMAC immobilized metal ion affinity chromatography
IPTG isopropylthiogalactopyranoside
K lysine (Lys)
KCl potassium Chloride
(K) Da (kilo) dalton
Kg kilogram
L leucine (Leu)
LB lauria bertini medium
LiSo4 lithium Sulphate
M molar
M methionine (Met)
mAb monoclonal antibody
MALDI-TOF matrix assisted laser desorption/ionization- time of flight
MES 2-[N-morpholino]-ethanesulphonic acid
MgSo4 magnesium Sulphate
min minute(s)
ml millilitre
mM mili molar
MME monomethylether
MnCl2 manganese Chloride
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOPS</td>
<td>3-morpholino propane</td>
</tr>
<tr>
<td>ESI-MS/MS</td>
<td>electrospray mass spectrometry</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut-off</td>
</tr>
<tr>
<td>µL</td>
<td>microlitre</td>
</tr>
<tr>
<td>N</td>
<td>asparagine (Asn)</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium Chloride</td>
</tr>
<tr>
<td>NBT</td>
<td>nitro blue tetrazolium</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel charged nitriloacetic acid resin</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>O.D.</td>
<td>optical density</td>
</tr>
<tr>
<td>P</td>
<td>proline (Pro)</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PBST</td>
<td>phosphate buffered saline containing 0.05% (v/v) Tween20</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethyleneglycol</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonylfluoride</td>
</tr>
<tr>
<td>Q</td>
<td>glutamine (Gln)</td>
</tr>
<tr>
<td>R</td>
<td>arginine (Arg)</td>
</tr>
<tr>
<td>r.p.m.</td>
<td>rotation per minute</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>root mean square</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>S</td>
<td>serine (Ser)</td>
</tr>
<tr>
<td>scFv</td>
<td>single chain variable fragment consists of $V_H$ and $V_L$ fragment of antibody</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecylsulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecylsulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>$\text{SO}_4^{2-}$</td>
<td>sulphate ions</td>
</tr>
<tr>
<td>T</td>
<td>threonine (Thr)</td>
</tr>
<tr>
<td><em>Taq</em></td>
<td><em>thermus aquaticus</em></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>TEMED</td>
<td>N, N, N’, N’ tetramethylethylenediamine</td>
</tr>
<tr>
<td>TR</td>
<td>tandem repeat</td>
</tr>
<tr>
<td>Tris</td>
<td>tris-[hydroxymethyl]-aminoethane</td>
</tr>
<tr>
<td>U</td>
<td>unit</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>volt</td>
</tr>
<tr>
<td>V</td>
<td>valine (Val)</td>
</tr>
<tr>
<td>V-type</td>
<td>variable domain type</td>
</tr>
<tr>
<td>VH</td>
<td>variable heavy chain fragment</td>
</tr>
<tr>
<td>VH-H</td>
<td>camelid variable heavy chain</td>
</tr>
<tr>
<td>VL</td>
<td>variable light chain fragment</td>
</tr>
<tr>
<td>VNTR</td>
<td>variable number tandem repeat</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>W</td>
<td>tryptophan (Trp)</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>Y</td>
<td>tyrosine (Tyr)</td>
</tr>
<tr>
<td>ZnCl₂</td>
<td>zinc chloride</td>
</tr>
</tbody>
</table>
VI.9 List of tables

Table I-1 FDA approved therapeutic antibodies ......................................................... 9
Table I-2 Therapeutic antibodies undergoing various phases of clinical trials ............... 9
Table I-3 Comparison of the human tandem repeat amino acid sequences .................. 18
Table II-1 Names, suppliers and genotypes of E. coli strains used. ................................. 29
Table III-1 Data collection statistics for M12-VH domain ......................................... 76
Table III-2 Data statistics breakdown with resolution for M12-VH domain ................. 77
Table III-3 Summary of structure refinement statistics for M12-VH domain ................. 78
Table III-4 Deviation (bold) of dihedral angles of CDR-H1 & CDR-H2 loop ............... 86
Table III-5 Comparison of dihedral angle values for CDR-H2 loop ......................... 86
Table III-6 β-Strand boundaries and turns .............................................................. 89
Table III-7 Hydrogen bonding of CDR residues to the remaining M12-VH domain .......... 89
Table III-8 Comparison of the amino acids of the M12-VH domain ......................... 91
Table III-9 Statistics of the refined model against the structure factors ..................... 94
Table IV-1 Canonical structure assignment ......................................................... 114
Table VI-1 Data collection statistics for M12-VH domain (1.8Å) ................................. 144
Table VI-2 Data statistics breakdown for the M12-VH domain (1.8Å) ......................... 144
Table VI-3 Summary of structure refinement statistics for M12-VH domain (1.8Å) .... 146
Table VI-4 Intra-molecular hydrogen bonding in M12-VH domain (1.5Å) ............... 147

VI.10 List of figures

Figure I-1 Schematic representation of an IgG molecule ........................................... 2
Figure I-2 Fragmentation of IgG with papain and pepsin enzyme ............................. 3
Figure I-3 Different forms of antibodies ............................................................... 6
Figure I-4 Schematic representation of IgG heavy chain unarranged gene ................. 11
Figure I-5 Typical bulged (A) and non-bulged (B) CDR-H3 torso region .................. 14
Figure I-6 Illustration of the differences between human VH and camelid VHH ......... 16
Figure I-7 Schematic representation of a full-length human MUC1 ......................... 19
Figure I-8 General MUC1 structure ................................................................. 20
Figure I-9 The single tandem repeat peptide ......................................................... 20
Figure I-10 Two tandem repeat peptides schematically displaying ...................... 22
Figure III-1 HCDF fermentation profile of scFvM12 ............................................. 59
Figure III-2 Western blot analysis (II.2.4.4) for scFvM12 ....................................... 59
Figure III-3A Ni-NTA purification profile of scFvM12 ........................................... 60
Figure III-4A Cation-exchange (II.2.3.1.3) elution profile .................................... 61
Figure III-5A Anion-exchange (II.2.3.1.4) elution profile scFvM12 ......................... 62
Figure III-6A Size-exclusion chromatography .................................................... 63
Figure III-7 Iso-electric focusing (IEF, II.2.4.3) of gel-purified ............................ 63
Figure III-8 MALDI-TOF mass analysis (II.2.6) of gel-purified ............................. 64
Figure III-9A The rod shaped crystals of full size scFvM12 .................................. 65
Figure III-10 Flow cytometric (II.2.7) analysis for assessing the functionality ...... 66
Figure III-11 SDS-PAGE (18%, II.2.4.3) showing the digestion profile of scFvM12 .... 67
Figure III-12 Western-blot (II.2.4.4, 18% SDS-PAGE) ....................................... 67
Figure III-13 SDS-PAGE (18%, II.2.4.3) showing the effect of subtilisin ............... 68
Figure III-14 M12-VH domain crystals (150 x 50µm; after 5-7 days) grown .......... 70
Figure III-15 Western-blot (II.2.4.4, 15% SDS-PAGE) for solubilized VH fragment ... 71
Figure III-16 MALDI-TOF mass spectrum (II.2.6) of the solubilized crystals ......... 71
Figure III-17 ESI-MS/MS spectrum (II.2.6) of the N-terminal .............................. 72
Figure III-18 M12-V_{H} fragment crystals generated after thermolysin------------------- 73
Figure III-19 Crystal of M12-V_{H} domain (~180x60x60µm) soaked ------------------------------- 73
Figure III-20 Native gel electrophoresis (II.2.4.6) of scFvM12 and MUC1 ----------------------- 74
Figure III-21 Tiny and thin plate like crystals of the scFvM12 in complex with -------------- 74
Figure III-22 Single diffraction image for the M12-V_{H} domain collected ------------------- 76
Figure III-23 Average hydrophobicity of the M12-V_{H} domain residues ---------------------- 80
Figure III-24 Surface charge potential of the M12-V_{H} (1.5Å) ------------------------------ 80
Figure III-25 M12-V_{H} domain structure (1.5Å) -------------------------------------------- 81
Figure III-26 Amino acid sequence of the M12-V_{H} domain ------------------------------- 81
Figure III-27 Conserved disulphide bridge (Cys22-Cys92) of the M12-V_{H} domain ------------- 82
Figure III-28 Difference fourier (2F0-FC map, 1.5Å) map ---------------------------------- 82
Figure III-29 Unusual arginine-arginine (H66-H83) interaction in M12-V_{H} domain -------- 83
Figure III-30 CDR-H1 loop of the M12-V_{H} domain at 1.5 Å --------------------------------- 84
Figure III-31 CDR-H2 loop (main-chain) of the M12-V_{H} domain at 1.5Å ---------------------- 85
Figure III-32 CDR-H3 loop of the M12-V_{H} domain (1.5Å) --------------------------------- 88
Figure III-33 Secondary structure elements, relative accessibility ------------------------- 90
Figure III-34 Two-dimensional pictorial representation of intra ---------------------------- 90
Figure III-35 Antigen binding groove defined by the three CDR’s --------------------------- 92
Figure III-36 Ramachandran plot for the M12-V_{H} fragment (1.5Å) ------------------------ 93
Figure III-37 Individual temperature (B) factor analysis and the real space ---------------- 93
Figure III-38 Crystal packing of Cα trace of M12-V_{H} domain (1.5Å) ---------------------- 94
Figure III-39 ESI-MS/MS mass spectrum (II.2.6) for the reservoir -------------------------- 95
Figure III-40 ESI-MS/MS mass spectrum (II.2.6) for the hanging drops --------------------- 95
Figure III-41 Silver stained (II.2.4.5) SDS-PAGE (18%) analysis ----------------------------- 96
Figure III-42 Silver stained (II.2.4.5) SDS-PAGE (18%) analysis ----------------------------- 96
Figure III-43 MALDI-TOF (II.2.6) mass spectrums for the stability analysis ------------------ 97
Figure III-44 Western blot analysis of scFvM12 antibody stored at 4°C --------------------- 98
Figure III-45 Schematic representation of the strategy ------------------------------- 99
Figure III-46 PCR amplification (II.2.1.3.2) of the camelized ----------------------------- 100
Figure III-47 Restriction analysis (II.2.1.4) of the camelized ------------------------------ 101
Figure III-48 Western blot analysis (II.2.4.4) to observe the effect of E. ------------------- 102
Figure III-49 Ni-NTA purification (II.2.3.2) of the native ------------------------------- 103
Figure III-50 Western blot analysis (II.2.4.4) for the ------------------------------------- 103
Figure IV-1 Sequence alignment of M12-V_{H} domain with other VH domains --------------- 109
Figure IV-2 Schematic representation of denaturation midpoints ---------------------------- 112
Figure IV-3 Comparison of amino acid of SM3 -------------------------------------------- 113
Figure IV-4 Ribbon representation of the structurally aligned SM3-V_{H} (purple) --------- 114
Figure IV-5 Superposition of M12-V_{H} domain over SM3-MUC1 complex ------------------- 115
Figure IV-6 Human M12-V_{H} domain sequence comparison ------------------------------- 116
Figure IV-7 Ribbon representation of the structurally aligned M12-V_{H} domain ---------- 117
Figure IV-8 Superposition of the M12-V_{H} structure (yellow β-strands) ------------------ 118
Figure VI-1 scFvM12 antibody amino acid sequence ---------------------------------------- 142
Figure VI-2 Linear nucleotide map of M12-V_{H} gene segment ------------------------------ 143
Figure VI-3 Sequence similarity between M12-V_{H} and 1H0U-V_{H} domains. -------------- 145
Figure VI-4 Sitting drop (A) and hanging drop (B) crystallization techniques ------------- 149
Figure VI-5 Schematic representation of Bragg’s law for x-ray diffraction --------------- 151
I wish to express my gratitude to all those who contributed to this thesis:

Prof. Dr. Rainer Fischer for giving me an opportunity to work in his lab. I am also thankful to him for the critical reading of the manuscript and resulting suggestions.

Prof. Dr. Fritz Kreuzaler for reviewing the thesis and agreeing to be the co-examiner.

Dr. Kurt M. V. Hoffmann for his assistance and continuous scrutiny of my work. Although learning the computational aspects of crystallography is a time consuming and highly demanding job but he encourage me to do it independently.

I am thankful to Jean-Marie Frère and Paulette Charlier (Institut de Chimie, Salt Tilman, University of Liege) for providing access to their X-ray generator facility

I am highly grateful to Dr. Markus Perbandt (Staff scientist, University of Hamburg, Germany) for providing ample amount of time at X13 beamline and Dr. Wojtek Rypniewski (Beamline Scientist, EMBL Hamburg, DESY, Germany) for assisting at synchrotron beamline X13.

I am grateful to Dr. Ricarda Finnern (Fraunhofer IME, RWTH, Aachen, Germany) for providing me the scFvM12 construct.

Dr. Stephan Hellwig (Fraunhofer IME, RWTH, Aachen, Germany) deserves special thanks for his fermentation efforts without which the crystallization of this antibody could have remain a dream.

I am also grateful to the number of other distinguished faculties all around the world and Mr. Kingshuk Karuri who helped me in cracking the complex computational problems.

Thanks to Bhasker Sharma, Richa, Rajan and other friends for their moral encouragement. Mr. Rajeev Kumar deserves special mention for extending his financial support and Anupum Chattoupadhayay for providing the shelter.

Last but not the least, I thank all the peoples of the lab for their cooperation and contribution needed to complete this work.
**LEBENSLAUF**

**Persönliche Daten**

<table>
<thead>
<tr>
<th>Namen</th>
<th>Rajneesh Kumar Gaur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geburtsdatum</td>
<td>30.06.1973</td>
</tr>
<tr>
<td>Geburtsort</td>
<td>Muzaffarnagar, Indien</td>
</tr>
<tr>
<td>Familienstand</td>
<td>Ledig</td>
</tr>
<tr>
<td>Staatsangehörigkeit</td>
<td>Indisch</td>
</tr>
</tbody>
</table>

**Ausbildung**

- **2001-2004** Promotion am Institut für Biologie VII der RWTH Aachen unter Anleitung von Prof. Dr. Rainer Fischer
- **2001-2002** Advanced Certificate Course in Protein Crystallography on web, Birkbeck College, University of London, United Kingdom
- **1997-1999** Master im Biotechnologie, Studium an der Anna Universität, Chennai, Indien
  - Thema der Arbeit: Chemo-enzymatische synthese einer strukturellen Domäne von einer Lipopolysaccharidmolekül von *Leishmania donovani* parasit
- **1993-1997** Bachelor im Pharmacie, Studium an der Jamia-Hamdarad Universität, New Delhi, Indien
- **1989-1992** Bachelor im Naturwissenschaften (Zoologie), Studium an der Delhi Universität, Neu-Delhi, Indien

**Berufserfahrung**

- **1999-2000** Wissenschaftliche assistant (Research Associate), Indianisches Institut der Technologie, Neu-Delhi, Indien
- **2001-2004** Doktorand am Institut für Molekular Biotechnologie, RWTH Aachen, Deutschland