4.1 DESIGN OF TWO-DOMAIN M-CRYSTALLIN AND S80R MUTANT

Background:
Single domain ancestral crystallins are speculated to be the precursors for the evolution of crystallins of higher organisms, including those of eye lenses. Gene duplication is presumed to play pivotal role in escalating the functional diversity of crystallin group of proteins. Linker sequences determine the nature of interactions between partner domains, domain stability and solubility of tow-domain and multi-domain crystallins. In this backdrop, domain duplicates of M-crystallin, an archear member of βγ-crystallin super family of proteins, were designed by incorporating different linker sequences between partner domains. In the following sections of thesis work, modified proteins were aptly used to study various structural and biophysical traits of M-crystallin.

4.1.1 Rational design of the proposed constructs

Following different schemes, the gene sequence of M-crystallin was duplicated by incorporation of linker sequences of different nature: (i) Flexible linker- which codes for eight Glycine residues between partner domains (ii) Semi-flexible linker- which codes for ten residues [(Gly)$_4$-Arg-Ser-(Gly)$_4$] and (iii) Rigid linker- which codes for a seven residue [Phe-His-Ser-Gly-Ser-His-Arg] natural linker (of human γD crystallin), between adjacent domains.

In each case, the gene sequence encoding M-crystallin single domain was first PCR amplified with suitable overhangs to facilitate tethering of the two ends and finally ligated to produce the amplified target which codes for two-domain protein. The strategies employed for the amplification of the above mentioned domain duplicates are described below:

Scheme 1: Incorporation of flexible linker

The gene sequence was PCR amplified in two stages as shown in Figure 4.1.
Step 1: Six glycine residues are incorporated at 3’ end of PCR-1 and 5’ end of PCR-2
Step 2: Forward primer of PCR-1 and reverse primer of PCR-2 are used to ligate the amplified products of PCR-1 and PCR-2. The resulting sequence, PCR-3 encodes M-crystallin two domain protein with eight residue glycine linker between individual domains.
Figure 4.1: Schematic illustration of M-crystallin gene duplication using flexible linker. Single domain gene sequence is shown in grey, restriction sites flanking the target sequence as highlighted in pink. Shorter blue bars represent the linker sequence incorporated at respective end of the gene sequence. Longer blue bar represents the final linker sequence tethering the individual gene sequences.

Scheme 2: Incorporation of semi-flexible linker

The gene sequence was PCR amplified in two stages as shown in Figure 4.2.

Step 1: Four Gly, followed by Arg and Ser residues are incorporated at 3’ end of PCR-4 and Arg and Ser followed by four glycines at 5’ end of PCR-5.

Step 2: Amplified products of PCR-3 and PCR-4 were digested by BglII and BamH1 respectively to produce 5′-3′ cohesive ends, and ligated to produce the final amplified gene. The resulting sequence encodes M-crystallin two domain protein with (Gly)$_4$-Arg-Ser-(Gly)$_4$ linker.

Figure 4.2: Schematic illustration of M-crystallin gene duplication using semi-flexible linker. Colour codes are used same as in scheme 1.
Cloning and over-expression of modified M-crystallin

Scheme 3: Incorporation of rigid linker

The gene sequence was PCR amplified in two stages as shown in Figure 4.3.
Step 1: Sequence encoding Phe-His-Ser-Gly-Ser-His was incorporated at 3′ end of PCR-6 and Gly-Ser-His 5′ end of PCR-7. BamH1 site GGATCC codes for the Gly-Ser stretch.
Step 2: Amplified products were digested by BamH1 to produce 5′-3′ cohesive ends, and ligated to produce the final amplified gene. The resulting sequence encodes M-crystallin two domain protein with Phe-His-Ser-Gly-Ser-His-Arg linker peptide.

Figure 4.3: Schematic illustration of M-crystallin gene duplication using rigid linker. Colour codes are used same as in scheme 1.

RESULTS

All the PCR reactions were set up under identical conditions, using appropriate primers for amplification of each gene as described in the above schemes. The unused reactants and enzyme were removed from the reaction mixture by PCR clean-up protocol (Qiagen kit). Purified products were eluted in sterile water and quantitated by UV absorbance at 260 nm. A260/280 was taken as an estimate to assess the purity of the products. The size of the fragments was checked on 1.2% Agarose gel, using gene ladder for molecular size reference (Figure 4.4A). Gene fusion was carried out by a second round of amplification for the domain duplicate with flexible/Gly8 linker. A mixture of amplified products, single gene and gene duplicate were obtained due to redundant forward and reverse primers. The target product (600 bp) was isolated by gel extraction and purified according to the above protocol (Figure 4.4B). Final products were verified by PCR amplification.
Chapter 4

of the individual two-domain sequences by using the appropriate forward and reverse primers (Figure 4.4C). The overhangs corresponding to individual PCR products are listed in Table 3 and final gene products are described in Table 4.

Figure 4.4: Agarose gel electrophoresis of PCR amplification products. (A) Purified PCR products corresponding to single domain M-crystallin with the designated overhangs. (B) PCR amplification of gene products from PCR-1 and PCR-2 to generate gene duplicate with flexible linker (PCR-3). Target band (~600 bp) corresponding to two-domain fragment was gel excised and purified. (C) Final amplification for the verification of the target gene sequences using appropriate forward and reverse primers.

Table 3: Sequences encoded by PCR amplified products and respective overhangs

<table>
<thead>
<tr>
<th>PCR product</th>
<th>Sequence encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-1</td>
<td>Nde1-M-crystallin-(Gly)6</td>
</tr>
<tr>
<td>PCR-2</td>
<td>(Gly)6-M-crystallin-BamH1</td>
</tr>
<tr>
<td>PCR-3</td>
<td>Nde1-M-crystallin-(Gly)8-M-crystallin-BamH1</td>
</tr>
<tr>
<td>PCR-4</td>
<td>BamH1-M-crystallin-(Gly)4-BglII</td>
</tr>
<tr>
<td>PCR-5</td>
<td>BamH1-(Gly)4-M-crystallin-Xho1</td>
</tr>
<tr>
<td>PCR-6</td>
<td>Nde1-M-crystallin-BamH1</td>
</tr>
<tr>
<td>PCR-7</td>
<td>BamH1-M-crystallin-Xho1</td>
</tr>
</tbody>
</table>

Table 4: Final gene products and the linker sequences between the domains

<table>
<thead>
<tr>
<th>Linker type</th>
<th>Sequence encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible</td>
<td>Nde1-M-crystallin-(Gly)8-M-crystallin-BamH1</td>
</tr>
<tr>
<td>Semi-flexible</td>
<td>BamH1-M-crystallin-(Gly)4-Arg-Ser-(Gly)4-M-crystallin-Xho1</td>
</tr>
<tr>
<td>Rigid</td>
<td>Nde1-M-crystallin-Phe-His-Ser-Gly-Ser-His-Arg-M-crystallin-Xho1</td>
</tr>
</tbody>
</table>
The target gene sequences were restriction digested with respective enzymes and incorporated into pET21a vector. Ligated products were transformed into *E.coli* DH5α cells and positive clones were verified by ampicillin resistance. DNA sequences were confirmed by gene sequencing and restriction digestion of the fragment.

### 4.1.2 Site-directed mutagenesis-Synthesis of S80R mutant

M-crystallin is structurally characterized by paired Greek-key motif and possesses two canonical Ca²⁺-binding sites on the flexible loops. The universal signature sequence for Ca²⁺-binding was reported as N/D-N/D-X₁-X₂-S/T-S and mutations in this motif have been shown to affect the structure, stability and solubility of the βγ-crystallins. Selective mutation of (S/T) in the 5th position of the motif to R (Arg) results in loss of Ca²⁺-binding but enhances the innate domain stability of the protein. This point mutation is a characteristic feature of the vertebrate lens βγ-crystallins. In this connection, the corresponding Ser residue in the second Ca²⁺-binding motif of M-crystallin (S80) was mutated to Arg, (henceforth named S80R), by site directed mutagenesis.

Gene sequence encoding S80R mutant was produced by custom gene synthesis (Biotech Desk, India). The target sequence was flanked by Nde1 and BamH1 restriction sites at 5’ and 3’ ends respectively. The modified gene sequence was incorporated into pET21a vector within the same restriction sites. Positive clones were verified from gene sequencing (Figure 4.5) and restriction digestion of the insert.

![Figure 4.5: Multiple sequence alignment of target sequence and obtained sequence from the positive clones](http://www.ebi.ac.uk/Tools/msa/clustalw2/).
Chapter 4

4.1.3 Over-expression of unlabelled and isotopically labelled proteins

Positive clones were transformed into *E.coli* BL21 (DE3)-RIL expression host cells and protein over expression and purification was performed as described in chapter 3. The eluted fractions from size exclusion chromatography were checked for purity on 15% SDS-PAGE and concentrated using amicon centrifugal filter with a molecular cut-off of 3 kDa (Figure 4.6). The protein was quantitated by UV-vis absorption at 280 nm to estimate the yield of the protein.

![SDS-PAGE analysis of purified proteins](image)

**Figure 4.6: SDS-PAGE analysis of purified proteins.** (A) Two-domain M-crystallin tethered by various linker sequences as marked in lane 1: Flexible/(Gly)$_8$, lane 2: Semi-flexible/(Gly)$_4$-Arg-Ser-(Gly)$_4$, lanes 3: rigid linker/Phe-His-Ser-Gly-Ser-His-Arg (B) Purified S80R mutant.

4.1.4 Conclusion

The effect of modifications on the expression and solubility of M-crystallin is drastic. The yield of protein in case of flexible, semi-flexible and rigid linkers was found to be 18, 26 and 42 mg/l of culture respectively. The concentrated solutions of two-domain M-crystallin with semi-flexible and rigid linkers were found to be transparent and stable over a period of 2 months. However, concentrated solution of two-domain protein with flexible linker showed increased turbidity beyond ~80 μM upon long standing. This readily shows the effect of linker peptides on the stability and longevity of βγ-crystallin domains.
The yield of S80R mutant was ~66 mg/l of culture and is almost the same as that of wild type protein. Concentrated solution of the mutant remained stable over a period of 2 months, however, a fraction of the protein (~1/5th) precipitated out of solution, forming a clear phase separation between soluble and insoluble fraction. This indicates that the mutant affects the solubility of the protein and lowers it from that of wild type protein.

During the course of the work presented in the thesis, these modified proteins were used in defined contexts to compare and understand the structural and functional aspects of M-crystallin.