Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Bioinformatics in molecular immunology laboratories demonstrated: Modeling an anti-CMV scFv antibody

Chua Kek Heng[1*] and Rofina Yasmin Othman[2]

[1]Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Malaysia; [2]Institute of Biological Sciences, Faculty of science, University of Malaya, Malaysia; Chua Kek Heng* - E-mail: khchua@um.edu.my; * Corresponding author

received February 4, 2006; accepted March 20, 2006; published online March 27, 2006

Abstract:

A scFv (single chain variable fragment) antibody clone from anti-CMV (anti-cucumber mosaic virus) was successfully constructed from immunized mouse and the DNA sequence was submitted to GenBank (AY337618 and AY337619). The expression of a 32 kDa recombinant antibody in bacteria was verified using ELISA (enzyme-linked immunoassay) and western blot. However, elucidation of specific anti-CMV scFv function requires detailed and time consuming immuno-assays. Alternatively, useful functional information on anti-CMV scFV antibody can be obtained using available Bioinformatics tools and techniques without performing tedious assays. Here, we use the commonly used Bioinformatics tools and databases such as BLAST (basic local alignment search tool), GenBank, PDB (protein databank), KABAT numbering, SWISS-MODEL and Insight II to gain specific functional insights into anti-CMV scFv.

Keywords: anti-CMV scFv; sequencing; GenBank; homology modeling; CDR; epitopes

Background:

Functional antibodies have been extensively used in pharmaceutical and clinical applications. Antibodies have a typical structure consisting of two identical heavy and light chains joined together by disulfide and non-covalent bonds. Fv (variable fragment) plays a role in the antigen-binding activities of an immunoglobulin molecule. It is the smallest unit of immunoglobulin and is easily manipulated for immunological application. **[1, 2, 3]** ScFv (single-chain variable fragment) refers to the antibody fragments consisting of V_H (heavy) and V_L (light) chains connected by a peptide linker. **[4, 5**, **6]**

ScFvs are commonly constructed from hybridoma, mouse immunoglobulin, sheep immunoglobulin **[7]**, chicken immunoglobulin **[8, 9]** and human antibody repertoire. **[10**, **11]** They are generally produced at large scales using genetically engineered cloning vectors in bacterial hosts. **[12]** The most widely used peptide linker in scFv construction consists of a 15 residues sequence with repeats $(Gly_4Ser)_3$. The linker provides the molecule with a flexibility to move approximately 35 to 40 Å between the carboxy terminal of the V_H and the amino terminus of the V_L chains. **[5**, **6]** It should be noted that the affinity and stability of the scFv antibodies containing the $(Gly_4Ser)_3$ residues are generally comparable to those of the native antibody. **[6]** However, in some other antibody classes, the linker is made of residues GLU and LYS and they have a role in the solubility enhancement of scFv. **[13]** The stability and affinity is mainly contributed by disulfide bond linkers in these molecules. **[14**, **15]** Despite availability of several scFv structures at the PDB, the anti-CMV scFv structure is not known. In this report, we describe sequencing, GenBank data submission, modelling of an anti-CMV scFv antibody.

Methodology:

Plasmid Extraction and sequencing

Plasmid DNA of the anti-CMV ScFv antibody clone was prepared from bacterial culture using QIAGEN plasmid mini kit and subjected to sequencing using an ABI 3770 automated sequencer. The obtained sequences were then submitted to GenBank (AY337618 and AY337619).

Anti-CMV ScFv antibody sequence

The nucleotide sequences were then translated into protein sequences using the TRANSLATE program at Expasy. **[24]**

Sequence database search

Sequence database search for V_H (variable heavy chain) and V_L (variable light chain) was performed using the BLAST program at NCBI. **[16, 25]**

Expression analysis

The expression of a 32 kDa recombinant antibody in bacteria was verified using ELISA (enzyme-linked immunoassay) and western blot.

CDR (complementarity determining regions)

The CDR regions in the anti-CMV ScFv were determined using KABAT numbering **[26]** and their canonical structures were determined using Chothia description. **[17, 22]**

Homology modeling and energy minimization

SWISS-MODEL was used to model V_H and V_L chains separately. **[18, 19, 20, 26]** The V_H and V_L models were then connected by a synthetic peptide $[(Gly_4Ser)_3]$ using BUILDER/Insight II **[27]** followed by energy minimization in a CFF91 force field.

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Results and Discussion:

A phage display technology was successfully carried out to develop an anti-CMV scFv recombinant antibody as an alternative to the tedious and time-consuming hybridoma development. The approximately 700 bp scFv gene was successfully constructed and expressed in bacteria. **[21]** We then sequenced the anti-CMV scFv gene and deposited the sequence at GenBank (AY337618 and AY337619). A number of scFv structures at PDB (www.rcsb.org/pdb) and general information on antigen binding are well document. However, different scFv molecules from varying sources have different antigen binding functional properties in quantitative measures. Therefore, it is our particular interest to probe specifically into the structure of an anti-CMV scFv antibody, whose structure is not known. The GenBank submitted V_H and V_L chains of an anti-CMV scFv antibody with translated protein sequence are shown in Figure 1. A total of six CDRs (three in each chain) are identified using KABAT numbering and are highlighted in BOLD (Figure 1).

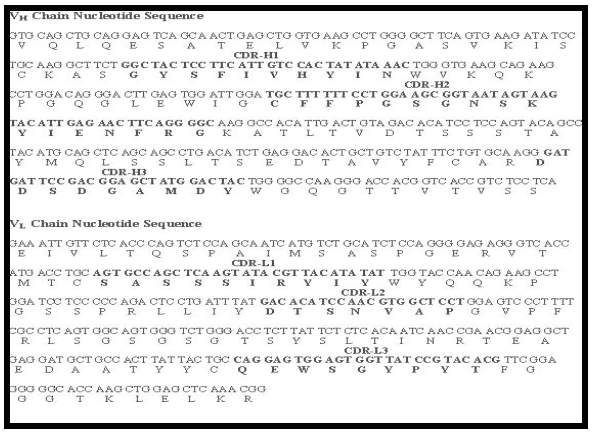

Figure 1: DNA and protein sequences for V_H and V_L chains. The GenBank deposited nucleotide sequences for V_H (GenBank Accession Number AY337618) and V_L (GenBank Accession Number AY337619) chains were obtained from in-house sequencing and the TRANSLATE program is used to translate into protein sequences. The 3 CDRs for both the chains were obtained using KABAT numbering and the CDR sequences are shown in boldface

Figure 2: BLASTP search hit for anti-CMV ScFv against PDB (protein databank). The results show that the anti-CMV scFv is similar to a synthetic scFv at a sequence similarity of 82% [gi|11066686|gb|AAG28706.1|(AF279665)]

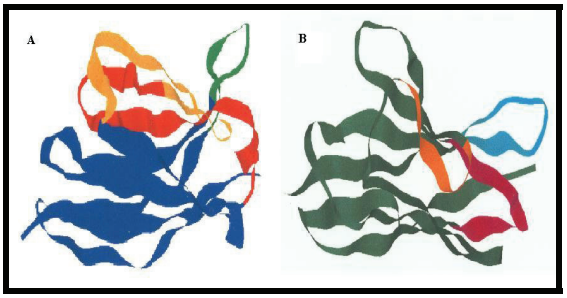

Figure 3: Canonical structures for CDRs in anti-CMV ScFv antibody. **(A)** Heavy chain: CDR-H1 (red), CDR-H2 (yellow) and CDR-H3 (green). **(B)** Light chain: CDR-L1 (purple), CDR-L2 (orange) and CDR-L3 (blue)

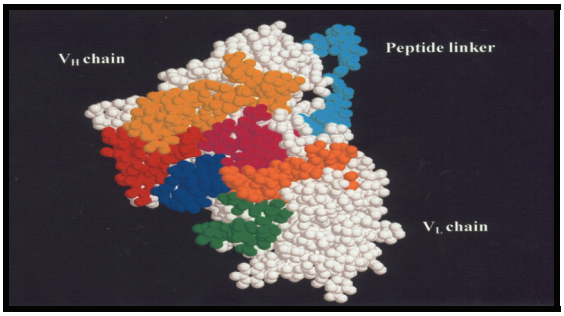

Figure 4: The anti-CMV ScFv antibody model shown in CPK display: The CPK model show all the CDRs on the surface of the molecule. The peptide linker (Turquoise), CDR-H1 (red), CDR-H2 (yellow) and CDR-H3 (green), CDR-L1 (magenta), CDR-L2 (orange) and CDR-L3 (blue). This model was built by joining V_H and V_L chains together by a peptide using BUILDER/Insight II and energy minimized in CFF91 force field

The distribution of CDRs in different regions of the V_H and V_L sequences is insightful, yet limited due to lack of 3D information. Hence, we searched the anti-CMV scFv protein sequence against PDB (protein database) using BLASTP to identify suitable templates for homology modelling (Figure 2). PDB search results show a high sequence similarity (82%) a synthetic peptide. Thus, the availability of a structural homolog at PDB was confirmed. We then submitted the anti-CMV scFv antibody sequence to SWISS-MODEL and the V_H and V_L structures were separately modelled. The models are represented in ribbons generated using RasMol **[29]** in Figure 3. The canonical conformations for CDRs in anti-CMV scFv are mapped in 3D and mapped regions are shown in Figure 3. The individually modelled V_H and V_L structures were linked by a synthetic peptide $[(Gly_4Ser)_3]$ using BUILDER/Insight II **[27]** followed by energy minimization in CFF91 force field. The modeled anti-CMV scFv structure is represented in CPK model and CDRs mapped. Thus, the structure of an anti-CMV scFv was modeled and CDRs mapped to the structure in 3D. However, the regions of CDR involved in antigen-binding are not known for anti-CMV scFv. Therefore, further studies are required to identify the antigen binding CDRs.

Conclusion:

The anti-CMV scFv antibody gene was sequenced and the sequence submitted to GenBank. To gain functional insight, the scFv antibody structure was modelled using SWISS-MODEL and Insight II. The CDRs in the modelled antibody structure were determined by KABAT numbering and mapped to provide insight for further epitope analysis. Thus, the identification and elucidation of CDRs in the anti-CMV scFv antibody is demonstrated using commonly available Bioinformatics tools and techniques.

Acknowledgement:

This study was funded by IRPA R&D grant ID: 09-02-03-0521, and Vote F of the University of Malaya. Special thanks to Li Phing, Suzita, Dr. Jenni, Dr. Amir and Azizah for their help.

References:

[1] J. K. Batra, *et al., Proc. Natl. Acad. Sci.,* 89:5867 (1992) [PMID: 1352878]

[2] P. N. Friedman, *et al., J. Immunol.,* 150:3054 (1993) [PMID: 8454873]

[3] R. J. Kreitman, *et al*., *Blood,* 80:2344 (1992) [PMID: 1421405]

[4] R. Raag & M. Whitlow, *FASEB J.,* 9:73 (1995) [PMID: 7821762]

[5] R. E. Bird, *et al*., *Science,* 242:423 (1988) [PMID: 3140379]

[6] J. S. Huston, *et al*., *Proc. Natl. Acad. Sci.,* 85:5879 (1988) [PMID: 3045807]

[7] G. P. White, *et al., Vet. Immunol. Immunopathol.,* 78:117 (2001) [PMID: 11884451]

[8] W. J. Andris, *et al., J. Immunol. Methods,* 242:159 (2000) [PMID: 10986398]

[9] N. Nakamura, *et al., Cytotechnol.,* 32:191 (2000)

[10] C Rossig, *et al., Med. Pediatr. Onco.,* 35:692 (2000) [PMID: 11107148]

[11] M. Houimel, *et al*., *Tumor Biol.,* 22:36 (2001) [PMID: 11054025]

[12] A. Skerra & A. Plunkthun, *Science,* 240:1038 (1988) [PMID: 3285470]

[13] M. Whitlow, *et al., Protein Eng*., 6:989 (1993) [PMID: 8309948]

[14] U. Brinkmann, *et al., Proc. Natl. Acad. Sci.,* 90:7538 (1993) [PMID: 8356052]

[15] Y. Reiter, *et al*., *J. Biol. Chem*., 269:18327 (1994) [PMID: 7913461]

[16] S. F. Altschul, *et al., Nucl. Acids Res.,* 25:3389 (1997) [PMID: 9254694]

[17] C. Chothia & A. M. Lesk, *J.Mol. Biol.,* 196:901 (1987) [PMID: 3681981]

[18] T. Schwere, *et al., Nucl. Acids Res.,* 31:3381 (2003) [PMID: 12824332]

[19] N. Guex & M. C. Peitsch, *Electrophoresis,* 18:2714 (1997) [PMID: 9504803]

[20] M. C. Peitsch, *BioTechnology,* 13:658 (1995)

[21] K. H. Chua, *et al., Asia Pacific J. Mol. Biol. Biotechnol.,* 11:93 (2003)

[22] R. M. MacCallum, *et al*., *J. Mol. Biol.,* 262:732 (1996) [PMID: 8876650]

[23] M. J. Sippl, *Proteins,* 17:355 (1993) [PMID: 8108378]

[24] http://www.expasy.ch/tools/dna.html

[25] http://www.ncbi.nim.nih.gov/BLAST/

[26] http://www.biochem.ucl.ac.uk/~martin/abs/GeneralInfo.html

[27] http://swissmodel.expasy.org//SWISS-MODEL.html

[28] http://www.msi.com

[29] http://www.umass.edu/microbio.rasmol/

Edited by Y. Z. Chen

Citation: Chua & Othman, Bioinformation 1(4): 118-120 (2006)