An *In Silico* Design of the Humanized Single-Chain Variable Fragment (scFv) Specific to Epithelial Cell Adhesion Molecule (EpCAM)

การออกแบบแอนติบอดีสายเดี่ยวของมนุษย์ซึ่งจับจำเพาะกับโมเลกุลเอปแคมที่ผิวเซลล์มะเร็งด้วยวิธีทางคอมพิวเตอร์

Suthee Wipasa (สุธี วิภาษา)* Dr.Tararaj Dharakul (ดร.ธารารัชต์ ธารากุล)**
Dr.Watip Tangjittipokin (ดร.วทิพย์ ตั้งจิตติโภคิน)*** Dr.Prapat Suriyaphol (ดร.ประพัฒน์ สุริยผล)****

ABSTRACT

The aim of this study was to computationally design and retain binding affinity of the humanized anti-Epithelial cell adhesion molecule (EpCAM) single-chain variable fragment (scFv). The 3-D structures of EpCAM and scFvs were obtained from homology/comparative modelling. The anti-EpCAM mouse scFv was 3-D docked with EpCAM and then humanized by CDR grafting. After 3-D docking and contact-based affinity prediction, anti-EpCAM humanized scFv lost binding affinity. Prediction of affinity change upon amino acid mutation guided affinity optimization by back mutation of the humanized scFv and the affinity was restored. In conclusion, the EpCAM-binding affinity of the designed humanized scFv can be retained after optimization of the affinity by back mutation.

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อออกแบบแอนติบอดีสายเดี่ยวที่มีความคล้ายคลึงของมนุษย์ที่จับจำเพาะกับโมเลกุลเอปแคมด้วยวิธีทางคอมพิวเตอร์ โดยคงไว้ซึ่งสัมพรรคภาพหรือความแรงในการจับจำเพาะ โมเลกุลแอนติบอดีสายเดี่ยวและเอปแคมที่ใช้ในการศึกษาได้มาจากวิธีเทียบเคียงโมเดลที่มีความเหมือน จากนั้นทดสอบการจับจำเพาะสามมิติแอนติบอดีสายเดี่ยวของหนูถูกเปลี่ยนให้มีความคล้ายในมนุษย์โดยคงไว้ซึ่งบริเวณจับจำเพาะหรือซีดีอาร์ แอนติบอดีสายเดี่ยวของมนุษย์สูญเสียสัมพรรคภาพในการจับจำเพาะเอปแคม เพื่อรักษาสัมพรรคภาพในการจับจำเพาะ หลังจากการพยากรณ์สัมพรรคภาพโดยการเปลี่ยนแปลงกรดอะมิโนนำไปสู่การเปลี่ยนแปลงย้อนกลับของตำแหน่งใดๆ ที่พบว่าสัมพรรคภาพลดลง พบว่าแอนติบอดีสายเดี่ยวของมนุษย์ได้สัมพรรคภาพในการจับเอปแคมคืนมา ทั้งหมดจึงสรุปได้ว่าแอนติบอดีของมนุษย์สามารถคงไว้ซึ่งสัมพรรคภาพการจับจำเพาะต่อเอปแคมด้วยการเปลี่ยนแปลงย้อนกลับ

Keywords: Epithelial cell adhesion molecule (EpCAM), In silico, Single-chain variable fragment (scFv)
คำสำคัญ: วิธีทางคอมพิวเตอร์ เอปแคม แอนติบอดีสายเดี่ยว

**Student, Master of Science Program in Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University*
***Professor, Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University*
****Assistant Professor, Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University*
*****Assistant Professor, Department of Medical Bioinformatics, Faculty of Medicine Siriraj Hospital, Mahidol University*

Introduction

Epithelial cell adhesion molecule (EpCAM) is 40-kDa glycoprotein overexpressed on the surface of epithelial cancer cells. EpCAM has its cell surface-exposed part of high antigenicity, called 6-kDa, and this elicits the majority of anti-EpCAM antibodies specific to this region (Pavšič et al., 2014). Whereas the other part, called 32-kDa, determined proteolytically separated part was known to have less antigenicity and only a few of anti-EpCAM antibodies recognized it. Due to these features, EpCAM has become the target of interest for the development of several anti-EpCAM antibodies and antibody-based fragments in cancer therapeutic and diagnostic applications (Schnell et al., 2013).

Nowadays, many computer-aided approaches have been developed for the rational designs of antibodies and other biotherapeutics (Roy et al., 2017; Sormanni et al., 2018) for enhanced effectiveness in clinical and diagnostic uses. Antibody-based fragment like single-chain variable fragment (scFv) is not exactly an antibody but a fusion protein of the antibody variable domain that contains complementarity determining regions (CDRs) for binding the antigen. Because of smaller size compared to intact antibody, scFv exhibits better tissue penetration and elicits lower immunogenicity when used in patients (Ahmad et al., 2012). ScFv has been used in many applications. Some of those has been commercially approved to be remarkably used in treatment of disease (Yannuzzi et al., 2019) and has been under computerized improvement of structural properties (Wang et al., 2011).

Anti-EpCAM 19D12 mAb generated in the previous study was found specific to 32-kDa part of EpCAM (Chantima et al., 2017). However, this was not exactly known whether where on EpCAM 32-kDa the 19D12 mAb did bound to. Another clone of anti-EpCAM mAb was previously studied in generation of anti-EpCAM humanized scFv (huscFv). EpCAM-binding affinity of the mAb was optimized by reducing root-mean-square deviation (RMSD) values between anti-EpCAM mouse scFv (mscFv) and the huscFv in order to increase structural similarity between two scFvs (Khantasup et al., 2015). Although RMSD values between the mscFv and the huscFv reduced when superimposed, huscFv represented 27-fold and 60-fold reduction of EpCAM-binding affinity compared to mscFv and intact parental mAb, respectively. It was hypothesized that if anti-EpCAM 19D12 mscFv was humanized, the 19D12 huscFv might also result in decreased affinity. Prediction of affinity *in silico* might help address and resolve this phenomena.

In this study, computational approaches were used to assess the design of anti-EpCAM 32-kDa humanized scFv (huscFv) clone 19D12. The attempt to retain binding affinity was to use prediction tools for screening of amino acids upon sequence of 19D12 huscFv that represented reduced affinity and the 19D12 huscFv was affinity-optimized by back mutation of those residues, in expectation for regaining of EpCAM-binding affinity of the original 19D12 mouse scFv (mscFv).

Objective of the study

The objective of this study was to design and retain binding affinity of the anti-Epithelial cell adhesion molecule (EpCAM) humanized single-chain variable fragment (scFv) by *in silico* approaches.

Materials and methods

Homology modelling of 3-D structures

Both anti-EpCAM 19D12 scFv and EpCAM were modelled from homologous template structures by homology modelling webserver SWISS-MODEL (https://swissmodel.expasy.org/interactive). The amino acid sequences of 19D12 scFv and EpCAM were only required as input data and directly uploaded onto webserver interface. The automated mode was used for modelling of 3-D structures. Once the modelling process was completed, protein data bank (PDB) files were obtained by exporting the model. Resultant 3-D models of 19D12 scFv and EpCAM were opened in molecule visualizer program PyMOL in order to see 3-D structures and thus re-edited protein chain letter. In this study, the input command for chain alteration of the scFv was 'alter (chain ?),chain='X'' and 'A' was set for EpCAM chain. This was essential for further steps of contact-based affinity prediction and affinity change upon mutation which required chain identification.

Protein-protein docking

Anti-EpCAM 19D12 scFv and EpCAM were computationally docked by protein-protein docking webserver ClusPro 2.0 (https://cluspro.bu.edu). The antibody mode was used for docking. PDB files of 19D12 scFv and EpCAM, derived from homology modelling, were requisite for the process and input to 'receptor' and 'ligand', respectively. Docking then proceeded by clicking 'dock'. ClusPro 2.0 resulted in 10-30 most populated clusters. The first docking cluster which was mostly populated cluster with lowest energy score and, viewed through PyMOL, offered rational shape complementarity between 19D12 scFv and EpCAM was selected.

Contact-based affinity prediction

Anti-EpCAM 19D12 scFv was predicted for binding affinity *in silico*. PRODIGY (PROtein bindDIng energGY prediction; https://bianca.science.uu.nl/prodigy) was the computer-aided webserver used for contact-based prediction. The PRODIGY protein-protein mode was used for affinity prediction. Through this, PDB file of the 3-D docking complex of 19D12-EpCAM, provided by ClusPro 2.0, was submitted onto webserver working interface. Importantly, chain identification was needed. Chain 'X' for 19D12 scFv and 'A' for EpCAM were input to interactor box 1 and 2, respectively. Temperature was already set as 25ºC by the program. Then, prediction initiated by clicking 'submit PRODIGY'.

Prediction of affinity change upon mutation

Anti-EpCAM 19D12 mscFv-EpCAM docking complex was submitted to the webserver mCSM-AB (biosig.unimelb.edu.au/mcsm_ab/prediction). In addition to requirement of PDB file of the complex, the detail of amino acid mutation and chain identification were needed in order to observe affinity change upon mutation of the 19D12 huscFv after humanization. Amino acid change was annotated as follows; X E5V, for example, meant glutamic acid (E) in the position 5 of mscFv chain X was mutated into valine (V). Text (.txt) file containing the list of mutation was uploaded onto the webserver along with PDB file. Prediction began by clicking 'run regression'.

Humanization by CDR grafting

By principle, CDR grafting is the method that the original CDRs are transferred onto the human immunoglobulin germline sequence mostly identical to original antibody sequence, called framework. The most identical human germline sequence, selected from IMGT database, was aligned with the anti-EpCAM 19D12 mscFv. In each 19D12 mscFv framework, amino acids different to those in human germline were changed (mutated) in order to make the 19D12 huscFv sequence more human-like.

Results

Anti-EpCAM 19D12 mscFv bound EpCAM at membrane-proximal region

To investigate original binding site on EpCAM, anti-EpCAM 19D12 mouse monoclonal antibody (mAb) was computationally studied in format of anti-EpCAM 19D12 mouse scFv (mscFv). Both 19D12 mscFv and EpCAM had their most identical homologous templates and were successfully modelled. From protein docking result, with selection of the first docking cluster (mostly populated and lowest structure energy), 19D12 mscFv was found binding to EpCAM 32-kDa at membrane-proximal region and not hindered by glycans. CDRs of 19D12 mscFv exposed and conformed good shape complementarity to their recognition site (Figure 1).

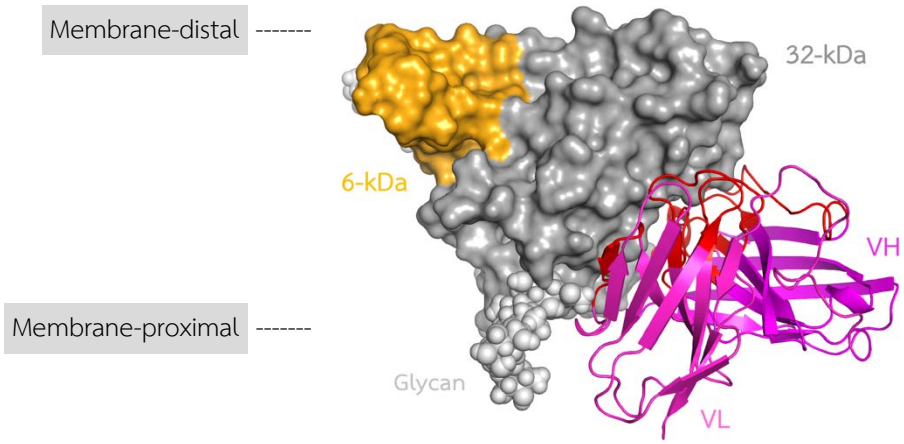


Figure 1 3-D representation of anti-EpCAM 19D12 mscFv (magenta ribbon) in docking with EpCAM. CDRs of 19D12 mscFv are colored in red. EpCAM 32-kDa and 6-kDa parts are represented in dark-grey and orange surface, respectively. Glycans on EpCAM are shown by light-grey sphere.

Anti-EpCAM 19D12 huscFv lost binding affinity after humanization

To make 19D12 mscFv more human-like, a simple approach like CDR grafting resulted in 19D12 humanized scFv (huscFv). IGHV3-21*04 (similarity 75.5%) and IGHJ4*01 (85.7% similarity) were the human germline sequences selected as framework acceptor for 19D12 mscFv VH. For VL, germline sequences IGKV1-16*01 (62.1% similarity) and IGKJ2*01 (90.9% similarity) were chosen. From sequence alignment, there were 32 unmatched residues to be changed (mutated) into their human counterparts (Figure 2).

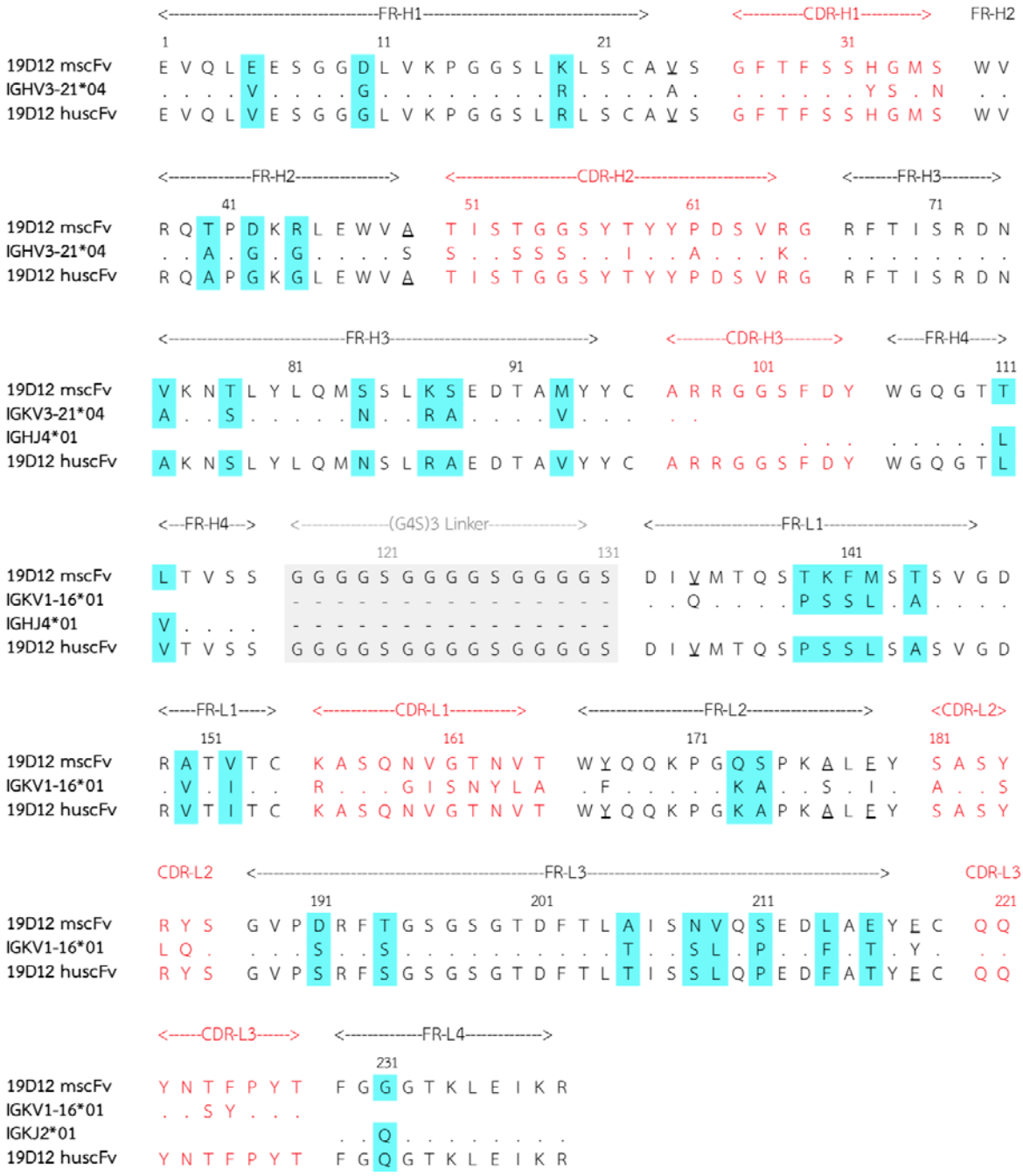
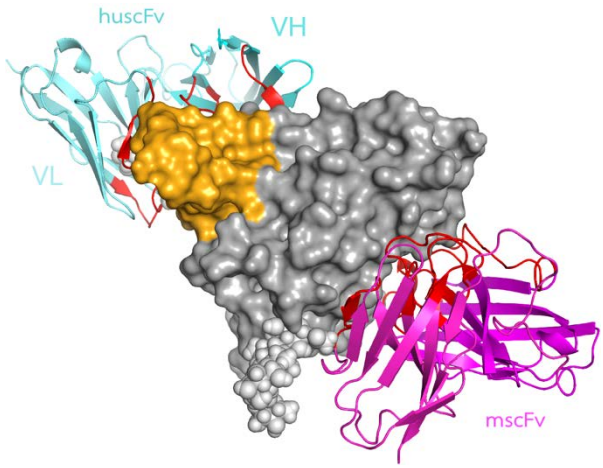


Figure 2 Humanization of 19D12 mscFv through alignment with the selected human immunoglobulin germline sequence. Each dot (.) determined identical amino acid between 19D12 mscFv and the human germline sequences. Highlighting in cyan pinpointed the residue(s) on 19D12 mscFv to be changed (mutated). Grafted CDRs of 19D12 mscFv were labeled in red and unchanged. Dashes (-) meant absence or not presented. The underlined residues were key conformational amino acids.

After humanization, 19D12 huscFv was found losing the recognition site at membrane-proximal region of EpCAM when compared with 19D12 mscFv by complex superimposition and bound to the 6-kDa, instead of 32-kDa (Figure 3A). The huscFv also exhibited reduced EpCAM-binding affinity (Figure 3B).

A



B

ScFv	Predicted K_D (M)
19D12 mscFv	7.2×10^{-11}
19D12 huscFv	1.7×10^{-09}

Figure 3 (A) 3-D representation of anti-EpCAM 19D12 huscFv (cyan ribbon) compared with 19D12 mscFv (magenta ribbon) in docking with EpCAM (orange and dark-grey surface). (B) Affinity in K_D values of both 19D12 scFvs predicted and provided by PRODIGY webserver.

Anti-EpCAM 19D12 huscFv regained binding affinity after optimization by back mutation

To retain the affinity, the amino acid residues causing reduced affinity were noted. Prediction of affinity change upon amino acid mutation of anti-EpCAM 19D12 huscFv showed 10 of 32 amino acids representing reduced change upon mutation (-ΔΔG) as shown in Table 1.

Table 1 Decreased binding affinity change upon mutation (ΔΔG) of amino acids on the sequence of anti-EpCAM 19D12 huscFv.

Mouse residue	Position	Human residue	Predicted ΔΔG
E	5	V	-0.263
D	10	G	-0.251
K	19	R	-0.419
D	42	G	-0.284
T	78	S	-0.304
S	84	N	-0.364
M	93	V	-0.754
L	112	V	-0.227
T	194	S	-0.438
E	216	T	-0.543

These 10 residues were thus back-mutated into their original mouse scaffolds in order to restore EpCAM-binding affinity of 19D12 huscFv (Figure 4). The affinity-optimized 19D12 huscFv sequence was then denoted as b19D12 huscFv ('b' for back-mutated).

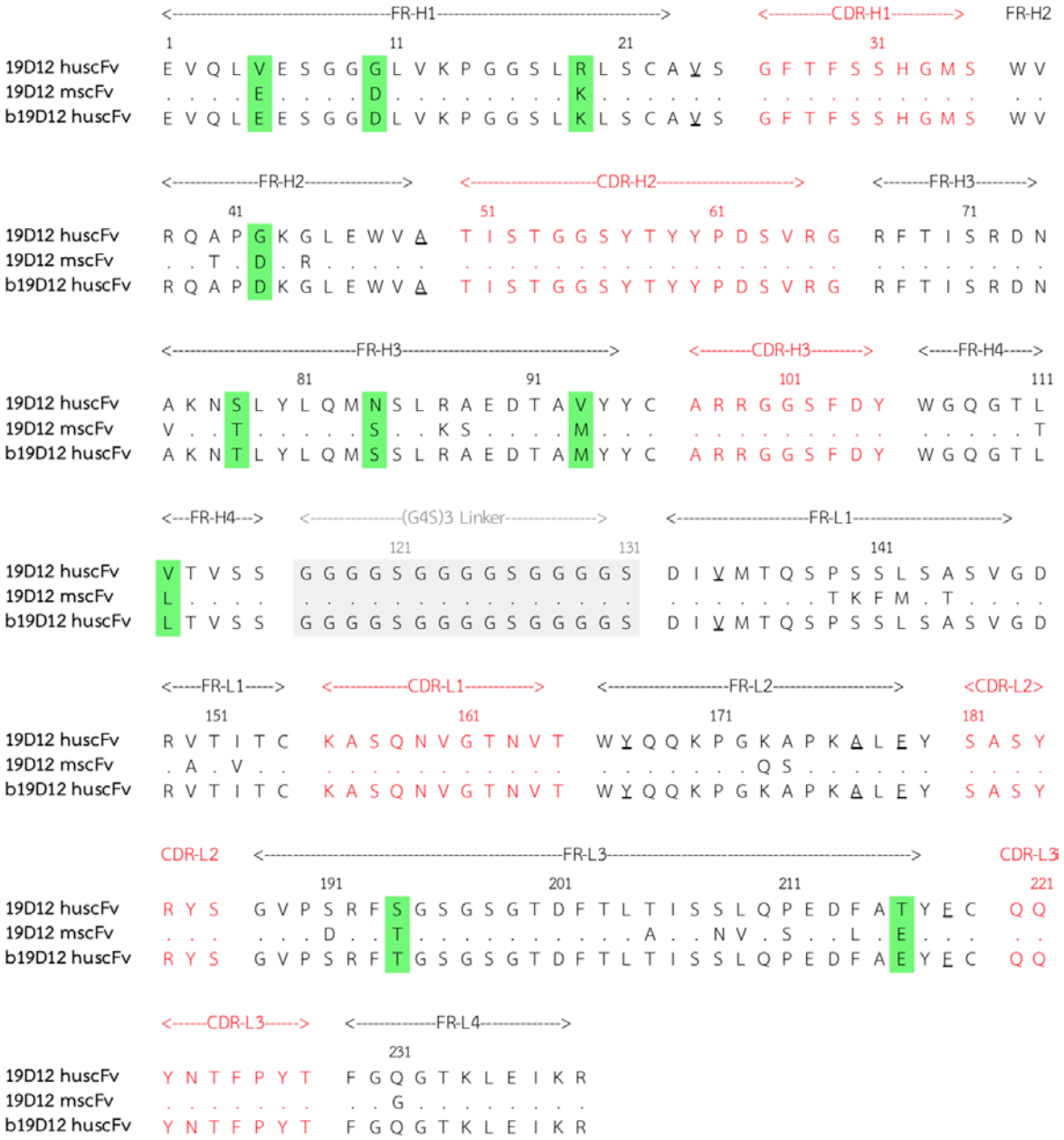
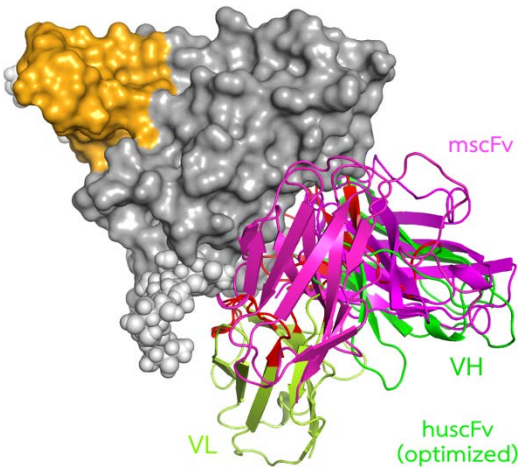


Figure 4 Back mutation of 19D12 huscFv through alignment with 19D12 mscFv sequence. Each dot (.) determined identical amino acid between 19D12 huscFv and 19D12 mscFv. Highlighting in green marked the residue(s) on 19D12 mscFv to be changed (mutated) back into mouse (wild) residues. CDRs were labeled in red and maintained unchanged. The underlined residues were key conformational amino acids.

After affinity optimization by back mutation, anti-EpCAM b19D12 huscFv was found binding to EpCAM 32-kDa at membrane-proximal region in the same way as anti-EpCAM 19D12 mscFv (Figure 5A). Moreover, the b19D12 huscFv also regained EpCAM-binding affinity (Figure 5B).

A



B

ScFv	Predicted K_D (M)
19D12 mscFv	7.2×10^{-11}
19D12 huscFv	1.7×10^{-09}
Optimization by back mutation	
b19D12 huscFv	1.4×10^{-10}

Figure 5 (A) 3-D representation of anti-EpCAM b19D12 huscFv, affinity-optimized by back mutation, (green ribbon) compared with 19D12 mscFv (magenta ribbon) in docking with EpCAM (orange and dark-grey surface). (B) Affinity in K_D values of both 19D12 scFvs predicted and provided by PRODIGY webserver.

Discussion

This study pointed out on antibody humanization with the utilization of computer-aided tools in order to retain antigen-binding affinity of humanized antibody. Here, single-chain variable fragment (scFv), a minimalized type of intact antibody that exhibits better tissue penetration, rapid clearance and elicits lower immunogenicity (Ahmad et al., 2012), was used as a format of antibody. The commercially approved Brolucizumab (BEOVU®) is the anti-vascular endothelial growth factor (VEGF) humanized scFv with high binding affinity and used in treatment of neovascular age-related macular degeneration (nAMD) (Yannuzzi et al., 2019). Additionally, other clone of anti-VEGF humanized scFv was studied by molecular dynamics simulation and computationally constructed from homology modelling in order to improve structure of the scFv and resolve its physicochemical problems when used *in vivo* (Wang et al., 2011).

Anti-EpCAM 19D12 mscFv was found binding EpCAM 32-kDa at membrane-proximal region. This described previous *in vitro* study of Chantima et al. (2017) that original 19D12 mouse antibody was able to bind EpCAM after EpCAM-positive cell lines were lysed and thus detected by Western Blot analysis. However, 19D12 mAb missed detection of EpCAM in live cell lines. Pavšič et al. (2014) studied on 3-D structure of EpCAM by using molecular dynamic and found that EpCAM formed into *cis*-dimer when expressed on live cell and the *cis*-dimer have to be structurally stabilized through its transmembrane

domain. EpCAM loses its dimeric state if it exists extracellularly. With recent finding, recognition site for 19D12 antibody at membrane-proximal region of EpCAM monomer was revealed and accessible after EpCAM dimer lost stability and dissociated.

Protein-protein docking by ClusPro 2.0 provided rational result of 19D12 scFv-EpCAM complex. Docking complexes of 19D12 mscFv and huscFv to EpCAM were selected from the first docking cluster. As stated by Kozakov et al. (2005), the mostly populated clusters of lowest energy conformations could contain near-native structure of the complex. Thus, it was speculated in this study that the first docking cluster might be the most near-native complex because it had highest complex members implying high probability of native state. Through structure visualization, good shape complementarity was observed between 19D12 scFv and EpCAM from selection of the first cluster. The example of this was employed by Kozakov et al. (2017) in selection of the first cluster of porcine trypsin (receptor) docked to soybean trypsin inhibitor (ligand) and superimposing the result with available complex acquired from protein data bank. The comparative was a native X-ray structure of enzyme-inhibitor complex and found correctly aligned to the resultant first cluster. However, Comeau et al. (2004) mentioned that the top populated clusters might comprise of about 92% of native structure. Moreover, selection of the first docking cluster seemed to correlate well with the model of enzyme-inhibitor pairs as shown in the study of Vajda et al. (2004). Therefore, to test this finding, *in vitro* experiment of 19D12 mAb binding capacity to EpCAM dimer is suggested. If the result from this docking approach is correct, 19D12 mscFv, for example, should not bind to extracellular EpCAM dimer.

EpCAM-binding affinity of 19D12 huscFv was regained after optimization of the affinity by back mutation. Changing of amino acids into those giving retained affinity might confer EpCAM-binding affinity of 19D12 huscFv although those residues, termed non-interacting surface (NIS) (Kastritis et al., 2014), were not directly contacted to EpCAM. However, light chain (VL) of the affinity-optimized 19D12 huscFv (b19D12) seemed to have its CDR loops not completely contacted to the original recognition site on membrane-proximal region of EpCAM (Figure 5A). This might be resulted from some of amino acids in human germline sequence that affected structural conformation of the original CDRs (Foote et al., 1992) although it did not reduce binding affinity of the huscFv. Zhu et al. (2009) offered antibody 'Resurfacing' or 'Veneering' method. They studied on their original antibody variable domain by homology modelling and humanized it by only changing the exposed 'surface' of the protein while conserving original core protein. Their humanized scFv represented retained antigen binding affinity and specificity of parental antibody and their work also enlightened optimization of 3-D structure in order to make it more similar to the native conformation. This can be manually performed in SWISS-Pdb Viewer (Guex et al., 1997). In this study, there was no experimental data in measurement of EpCAM-binding affinity of 19D12 mAb and scFvs available for comparison and this has never been performed before. Thus, confirmative data, such as cell-binding assays for affinity determination of protein-protein interactions (Hunter et al., 2016), is experimentally suggested in order to compare with the results from this *in silico* study.

Conclusion

With the simple method of scFv humanization and computerized approaches, the anti-EpCAM humanized scFv regained binding affinity after optimization of the affinity by back mutation. To see how reliable the results from *in silico* approaches can be, additional experimental studies were suggested.

References

Ahmad ZA, Yeap SK, Ali AM, Ho WY, Alitheen NB, Hamid M. scFv antibody: principles and clinical application. Clin Dev Immunol 2012; 2012: 980250.

Chantima W, Thepthai C, Cheunsuchon P, Dharakul T. EpCAM expression in squamous cell carcinoma of the uterine cervix detected by monoclonal antibody to the membrane-proximal part of EpCAM. BMC Cancer 2017; 17(1): 811.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004; 20(1): 45-50.

Foote J, Winter G. Antibody framework residues affecting the conformation of the hypervariable loops. J Mol Biol 1992; 224(2): 487-99.

Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997; 18(15): 2714-23.

Hunter SA, Cochran JR. Cell-Binding Assays for Determining the Affinity of Protein-Protein Interactions: Technologies and Considerations. Methods Enzymol 2016; 580: 21-44.

Kastritis PL, Rodrigues JP, Folkers GE, Boelens R, Bonvin AM. Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. J Mol Biol 2014; 426(14): 2632-52.

Khantasup K, Chantima W, Sangma C, Poomputsa K, Dharakul T. Design and Generation of Humanized Single-chain Fv Derived from Mouse Hybridoma for Potential Targeting Application. Monoclon Antib Immunodiagn Immunother 2015; 34(6): 404-17.

Kozakov D, Clodfelter KH, Vajda S, Camacho CJ. Optimal clustering for detecting near-native conformations in protein docking. Biophys J 2005; 89(2): 867-75.

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2017; 12(2): 255-78.

Pavšič M, Gunčar G, Djinović-Carugo K, Lenarčič B. Crystal structure and its bearing towards an understanding of key biological functions of EpCAM. Nat Commun 2014; 5: 4764.

Roy A, Nair S, Sen N, Soni N, Madhusudhan MS. In silico methods for design of biological therapeutics. Methods 2017; 131: 33-65.

Schnell U, Cirulli V, Giepmans BN. EpCAM: structure and function in health and disease. Biochim Biophys Acta 2013; 1828(8): 1989-2001.

Sormanni P, Aprile FA, Vendruscolo M. Third generation antibody discovery methods: in silico rational design. Chem Soc Rev 2018; 47(24): 9137-57.

Vajda S, Camacho CJ. Protein-protein docking: is the glass half-full or half-empty? Trends Biotechnol 2004; 22(3): 110-6.

Wang T, Duan Y. Probing the stability-limiting regions of an antibody single-chain variable fragment: a molecular dynamics simulation study. Protein Eng Des Sel 2011; 24(9): 649-57.

Yannuzzi NA, Freund KB. Brolucizumab: evidence to date in the treatment of neovascular age-related macular degeneration. Clin Ophthalmol 2019; 13: 1323-9.

Zhu H, Yang B, Yang X, Wang L, Xu J, Liao C, et al. A novel antibody fragment targeting HAb18G/CD147 with cytotoxicity and decreased immunogenicity. Cancer Biol Ther 2009; 8(11): 1035-44.