Insilico analysis of three different tag polypeptides with dual roles in scFv antibodies

Mozafar Mohammadi a,b, Foroogh Nejatollahi b,c,e, Amirhossein Sakhteman d, Neda Zarei e

a Faculty of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran
b Recombinant antibody laboratory, Department of Immunology, Shiraz University of Medical Sciences, Shiraz, Iran
c Shiraz HIV/AIDS research center, Shiraz University of Medical Sciences, Shiraz, Iran
d School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
e Department of Biotechnology, School of Veterinary Medicine, Shiraz University, Shiraz, Iran

HIGHLIGHTS

• Due to low immunogenicity of traditional linker (GGGGS) of scFv antibodies, it cannot be used as a tag for scFv detection and purification.
• We substituted the traditional linker with His-tag, C-myc or E-tag sequences through molecular modeling.
• Stability and integrity of all models were assessed by molecular dynamic (MD) simulation.

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form
13 April 2016
Accepted 16 April 2016
Available online 22 April 2016

Keywords:
ScFv stability
Polypeptide linker
Molecular modeling
MD simulation

ABSTRACT

Single chain fragment variable (scFv) antibodies are composed of variable heavy (V_H) and variable light (V_L) domains that are joined by a polypeptide linker. Typically, ([Gly4Ser]_n) sequence is used as a linker to retain the integrity of the antigen-binding domain. Due to its low immunogenicity, this sequence cannot be used as a tag for scFv detection and purification. Several evidences have shown that the addition of an N or C-terminal tag for scFv detection and purification will result in the decreased expression and binding capacity of this antibody fragment. In this study, we substituted the traditional linker (GGGGS) with His-tag, C-myc or E-tag sequences through molecular modeling. Stability and integrity of all models were assessed by molecular dynamic (MD) simulation. Based on MD simulation analysis, the model containing E-tag sequence as a linker indicated more stability compared to other molecules. The results suggest that E-tag not only can be substituted for the traditional linker, also eliminates the necessity of using additional tag for scFv detection and purification.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Single chain fragment variable (scFv) is the most common recombinant antibody form, containing the complete antigen-binding site of an antibody (Bird et al., 1988; Nejatollahi et al., 2012b, 2014b, 2013). Generally, scFv molecules are composed of variable heavy (V_H) and variable light (V_L) domains that are linked by a polypeptide linker (Deckert, 2009; Nejatollahi et al., 2014a, 2012a; Mohammadi and Nejatollahi, 2014). A 15-amino acid peptide, (Gly4Ser)_{15}, is typically used as the linker to retain the flexibility and solubility of scFv molecule as well as the integrity of the antigen-binding domain (Huston et al., 1988). However, due to its low immunogenicity this sequence cannot be utilized as a tag for scFv detection and purification (Kuttnner et al., 2004). The introduction of peptide tags which are genetically fused to the desired protein for recombinant protein purification and detection procedures is a revolution in protein engineering (Nygren et al., 1994; Terpe, 2003). Nonetheless, it has been shown that addition of an N or C-terminal tag, resulted in decreased expression as well as binding capacity of scFv (Buhler et al., 2010; Goel et al., 2000). Thus, an additional fusion tag can be ignored if the linker peptide joining the VH and VL domains of the desired recombinant scFv antibody is derived from the linear epitope tags. An important feature of such tags is to be freely accessible in the expressed scFv for both detection and purification (Kuttnner et al., 2004). Moreover, they should not interfere with integrity and stability of scFv variable domains. The most commonly used peptide tag is addition of six histidines (His6) to either N or C terminus of the recombinant protein (Schmidt and Skerra, 1993). The poly histidine
(polyHis) tags which composites ranging from 3 to 10 consecutive histidine residues are the most widely used affinity tags for purifying recombinant proteins (Chaga, 2001; Manjasetty et al., 2008; Li et al., 2010; Ma et al., 2015). Several other short peptides such as the C-myc tag, S-tag and FLAG® tag have also been utilized for detection and purification strategies. All of them are introduced as epitope tags that can be easily detected by specific monoclonal antibodies (Terpe, 2003). The C-myc tag (EQLISEDDL) is an epitope of the human c-myc proto-oncogene which is one of the earliest affinity tags developed (Evan et al., 1985). The C-myc tag is commonly used in protein engineering procedures including ELISA, Western blot and flow cytometry analysis (Kipriyanov et al., 1996). Furthermore, the scFvs expressed on Escherichia coli using the pCANTAB5E expression vector display a 12-amino acid (GAPVPGPVPPLEPR) tag (E-Tag) which can be recognized by anti-E tag monoclonal antibody (Shen et al., 2008) for several purposes such as detection of soluble antibodies in the cellular fractions, purification of soluble antibodies, and immunoassays including ELISA, Western blot and dot blot.

In this study, it was substituted the traditional linker (GGGGS) of scFv (Model 1) with each of three other sequences (His-tag (Model 2), C-myc (Model 3) and E-tag (Model 4)) through molecular modeling. The created structures were then evaluated by several corresponding programs. Finally, stability and other properties of modeled structures were analyzed using molecular dynamic simulation procedure and three proposed linkers were compared with the traditional scFv linker. Molecular dynamics simulation methods have been successfully applied for evaluating the functional and structural aspects of various biological active molecules in a nanosecond timescale (Dror et al., 2012; Sangeetha et al., 2014). Based on the aforementioned applications, molecular dynamics simulation was used in this study.

2. Methods

2.1. Tertiary structure prediction

2.1.1. Homology modeling

The 3D models of all structures were constructed using homology modeling as implemented in SWISS MODEL (Biasini et al., 2014; Kiefer et al., 2009; Arnold et al., 2006). All scFv fragment structures were modeled based on the crystal structure of single chain antibody F5 (resolution 2.5 Å, PDB entry: 4nik1) (Robin et al., 2014). Clustal Omega at http://www.ebi.ac.uk/Tools/msa/clustalo/ was used to align target sequences with corresponding template.

2.1.2. Evaluation of predicted models

To evaluate the tertiary structure quality of each model VADAR 1.8 (Volume Area Dihedral Angle Reporter) (Willard et al., 2003) and SAVES server at http://services.mbi.ucla.edu/SAVES, including the ERRAT2 (MacArthur et al., 1994) and Verify-3D (Bowie et al., 1991; Luthy et al., 1992) programs, were applied.

2.2. Molecular dynamic simulation

All models obtained by homology modeling procedure were subjected to molecular dynamic simulation studies. For this purpose, each PDB structure was embedded in a box with dimensions equal to 1 nm from the edges of the molecule. Subsequently, the system was solvated with spc216 water model. A concentration of 0.15 M NaCl was also inserted to simulate the physiological conditions. Prior to the main run of simulation, energy minimization followed by two short steps of MD was performed in order to equilibrate the whole system. During the minimization step steepest descent integrator with a maximum of 50,000 cycles and 10,000 kJ/mol/nm have been taken. In the two short runs of MD, NVT and NPT ensembles with restrains on protein backbone were taken. The main run of MD was finally performed after removing all restrains from the system. In the case of MD, Leap-frog integrator with LINCS constraint algorithm and particle mesh Ewald coulomb type was used. The most accurate Nose-Hoover thermostat was used along with Parrinello-Rahmen pressure coupling method.

2.2.1. Analysis of MD simulation trajectories

GROMOS force field implemented in Gromacs 4.5.3. Calculation of Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) plots were performed using the commands implemented in Gromacs. VMD software was used to compute the number of hydrogen bonds between VH and VL during 50 ns (Humphrey et al., 1996).

3. Results and discussion

3.1. Tertiary structure prediction

3.1.1. Homology modeling

The knowledge of 3D structures of target proteins is critically important for rational protein engineering. Although X-ray crystallography and NMR are the powerful tool in this regard, they are time-consuming and expensive (Schnell and Chou, 2008; Berardi et al., 2011; Fu et al., 2016). To acquire the structural information in a timely manner, a series of 3D protein structures and their binding sites with ligands were derived by various structural bioinformatics tools such as homology or comparative modeling (Chou, 2004; Wang et al., 2007; Du et al., 2010). In this study, we used homology modeling method, to create the appropriate tertiary structure of the four interested scFv models.

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASY web server at http://swissmodel.expasy.org/. It was used to create all interested tertiary structures based on the crystal structure of the single chain antibody F5 (resolution 2.5 Å, PDB entry: 4nik1) (Robin et al., 2014). The results of clustal Omega program (Sievers et al., 2011) showed 70% identity between the target sequences and template (except linker sequences) (Fig. 1). It was shown that when over 40% of amino acid sequences between target and template protein are identical, 90% of the main-chain atoms could be modeled with a root-mean-square distance (RMSD) error of about 1 Å. In this range of sequence identity, the conformational difference between target and template comes mainly from loops and side-chains (Xiang, 2006).

3.1.2. Evaluation of predicted models

VADAR is a web server for protein structure assessment which represents a compilation of several critical parameters derived from 15 well-known algorithms or previously published techniques for quantitatively evaluating protein structures (Willard et al., 2003). Ramachandran plot, as an output of the VADAR 1.8 program, assessed the final quality of proteins by indicating residues located in appropriate zone as well as the percentage of those in outlier regions. According to φ and ψ torsion angles, over 90% of amino acids related to all predicted structures are located in low free energy folding areas (Core and Allowed phi/psi regions) where there are minimum steric clashes between residue’s atoms (Fig. 3 and Table 1). Moreover, there are only a few amino acids located in generous and outside regions.

The SAVES server use different tools such as Verify_3D that determines the compatibility of an atomic model (3D) with its
own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc) and comparing obtained results with suitable structures. At least 80% of the amino acids should have been scored $\geq 0.2$ in the 3D-1D profile (Bowie et al., 1991). Over 93% of the residues of all models had an average 3D-1D score $\geq 0.2$. ERRAT2 program is another tool in which error values are plotted as a function of the position of a 9-residue sliding window. The error function is based on the statistics of non-bonded atom-atom interactions in the reported structure (compared to a database of reliable high-resolution structures) (MacArthur et al., 1994). All models had an overall quality factor scores near the Glycine-rich traditional linker-containing scFv (Model 1) (Table 2).

According to obtained results from Ramachandran plots, ERRAT and Verify-3D programs all models had proper and acceptable tertiary structures. In the next step, these structures were subjected to molecular dynamic simulation studies to investigate their behaviors in simulated physiological condition.

3.2. Molecular dynamic simulation

Since the pioneer paper entitled ‘The Biological Functions of Low-Frequency Phonons (Chou Kuo-Cheng, 1977) was published in 1977, a series of investigations into biomacromolecules from dynamic point of view have been stimulated. These studies have suggested that low-frequency (or terahertz frequency) collective motions do exist in proteins and DNA (Chou et al., 1989; Chou and Mao, 1988; Chou, 1988). The two main approaches to understanding the low-frequency modes are normal mode dynamics and molecular dynamics simulations (Steele, 1971). Molecular dynamics (MD) simulation, as a function of time, can provide the ultimate details concerning individual atom motion. Understanding the details at the atomic level is important for complicated reactions such as protein folding and cannot be easily
achieved by experimental studies (Caffi and Paci, 2005). To obtain the equilibrium time of each predicted protein models during MD simulation, root mean square deviation (RMSD) plots of the C alpha of atomic coordinates were calculated on the protein backbone. RMSD plot is usually used to evaluate the time needed for a system to reach equilibrium and estimates how long a simulation should be run. As demonstrated in Fig. 2, in the simulations which were performed at 310 °K temperature, all molecules expressed stable dynamics during 50 ns. Simulations were therefore ceased at this time course. Analyses of all proteins were performed on the trajectories within the first 50 ns.

3.3. Analysis of MD simulation trajectories

From 30th ns till the end, during which all conformations reached to a plateau level, Model 3 showed a larger deviation in comparison to other structures, attaining a maximum RMSD of about 0.33 nm. Unlike Model 3, Model 2 containing polyHis showed the least RMSD compared to other three conformations from start point to the end of the simulation. Throughout the plateau state, traditional model (Model 1) and Model 4, which contains an E-tag sequence in its linker, exhibited intermediate variations. A minimum deviation in the average RMSD values of Model 4 against traditional model in comparison with other models, from 15th ns to the end of simulation, was an indication of the possibility that it can be a suitable substitution for traditional scFv, though further analysis was needed (Fig. 4).

In order to determine the linker sequence role in the dynamic behavior of other residues, RMSF values of all modeled structures were analyzed (Fig. 5). The most motions were related to the traditional linker (GGGGS) sequence which fluctuated up to
0.6 nm in the entire simulation period, while other linkers exhibited fluctuations up to 0.4 nm. The high flexibility of traditional linker may be referred to the small size of non-polar Glycine amino acids (Argos, 1990). The Glycine residue is unique among amino acids in that all side chains are hydrogen atoms. Since its conformation has greater freedom it can provide flexibility for adjacent residues (Yan and Sun, 1997). In other areas, most fluctuations were seen in CDRs (complementarity determining region) which interact with antigenic epitopes. It has been shown that on whole antibodies, the antigen binding sites undergo larger fluctuations in both heavy and light chains compared to other regions (Sinha and Smith-Gill, 2005; Peters et al., 2000). Among other three models, Model 4 showed the same fluctuation on CDR regions as the traditional model (Model 1), while Model 2 and 3 exhibited more variation. Antigen binding sites of an antibody consist of inherent flexible and rigid regions. These flexible and rigid areas lie at VH and VL interface for the respective roles in Ab–Ag interactions. The residues with smaller motions are more likely to provide a functional framework for more flexible regions (Sinha and Smith-Gill, 2005). Lower fluctuations seen in Model 4 can be due to the presence of un-successive proline residues in E-tag sequence. Proline is the most preferred amino acid type in both linker and loop regions. It is a cyclic imino acid with no amide hydrogen to donate in hydrogen bonding (Chakrabarti and Chakrabarti, 1998).

Structural studies suggest that the proline-rich sequences form relatively rigid structures and show ‘elbow bending’ dynamics (Radford et al., 1987). It seems that these proline properties are the reason of the same intermediate fluctuations as Model 1. For further evaluation of the conformational stability of desired models, the average number of hydrogen bonds between VH and VL domains of scFv antibody were calculated and compared with each other and also with traditional scFv molecule (Model 1). The association of VH and VL domains in antigen binding sites of an antibody plays a critical role in the mechanism of antigen–antibody interaction (Vargas-Madrazo and Paz-Garcia, 2003). It has been shown that the significant role of hydrogen bonds and salt bridges in this association is undeniable (Sinha and Smith-Gill, 2005).

The average numbers of hydrogen bonds of Model 1, Model 3 and Model 4 scFvs were 9.6, 10 and 10.2, respectively. While the Model 2 showed the least average number of hydrogen bonding, 6.9 between VH and VL domains. Model 4, in comparison with other models, exhibited higher number of hydrogen bonds (Fig. 6) which results in more stability and sufficient interface integrity of variable domains. In other word, insufficient VH-VL interface stability of scFvs has often been suggested as a main cause of their aggregation and non-functionality (Reiter et al., 1994). Transient opening of the interface, allowed by the peptide linker, would lead to hydrophobic exposure of the patches favoring their aggregation.
(Pluckthun, 1992). Consequently, the VH domain of one chain is paired with the VL domain of another chain and vice versa (Whitlow et al., 1993).

4. Conclusion

In the present insilico study, four structures of single chain variable fragment (scFv) with different linkers were created using homology modeling method. Tertiary structure evaluations indicated appropriate results. Molecular dynamic simulation procedure was then applied to investigate the impact of linker sequence on the stability and integrity of scFv antibody. Based on the analyses of MD simulation trajectories, it was proven that Model 4 which contains E-tag sequence can be a suitable substitution for traditional linker in scFv molecules. By substitution of E-tag containing-linker, an additional fusion tag for detection and purification of scFv proteins might be ignored.

Acknowledgement

This research was financially supported by Shiraz University of Medical Sciences, grant no: 93-7365.

References


