COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 6, Issue: 7, March 2013, e201303009, http://dx.doi.org/10.5936/csbj.201303009

CSBJ

THE ALBUMIN-BINDING DOMAIN AS A SCAFFOLD FOR PROTEIN ENGINEERING

Johan Nilvebrant [a,*], Sophia Hober [a]

Abstract: The albumin-binding domain is a small, three-helical protein domain found in various surface proteins expressed by gram-positive bacteria. Albumin binding is important in bacterial pathogenesis and several homologous domains have been identified. Such albumin-binding regions have been used for protein purification or immobilization. Moreover, improvement of the pharmacokinetics, through the non-covalent association to albumin, by fusing such domains to therapeutic proteins has been shown to be successful. Domains derived from streptococcal protein G and protein PAB from *Finegoldia magna*, which share a common origin and therefore represent an interesting evolutionary system, have been thoroughly studied structurally and functionally. Their albumin-binding sites have been mapped and these domains form the basis for a wide range of protein engineering approaches. By substitution-mutagenesis they have been engineered to achieve a broader specificity, an increased stability or an improved binding affinity, respectively. Furthermore, novel binding sites have been incorporated either by replacing the original albumin-binding surface, or by complementing it with a novel interaction interface. Combinatorial protein libraries, where several residues have been randomized simultaneously, have generated a large number of new variants with desired binding characteristics. The albumin-binding domain has also been utilized to explore the relationship between three-dimensional structure and amino acid sequence. Proteins with latent structural information built into their sequence, where a single amino acid substitution shifts the equilibrium in favor of a different fold with a new function, have been designed. Altogether, these examples illustrate the versatility of the albumin-binding domain as a scaffold for protein engineering.

5TH ANNUAL INTERNATIONAL CONGRESS OF ANTIBODIES 2013

Introduction

Many gram-positive bacteria express surface proteins with ability to bind serum proteins [1]. The surface proteins typically contain tandemly repeated serum protein-binding domains with one or several specificities, which often include albumin binding [2,3]. The bacteria can thereby camouflage themselves with bound host-proteins to evade the immune system and potentially also scavenge protein-bound nutrients [4,5]. Albumin is the most abundant protein in plasma and expression of albumin-binding proteins has been shown to promote bacterial growth and virulence [5,6]. The bacterial species that express albumin-binding domains are usually part of the normal human flora and they are opportunistic pathogens. There are many different types of albumin-binding proteins with different size and function. For example, more than 40 albumin-binding domains have been found in one protein, forming a rod-like structure in a giant cell wall-associated fibronectin-binding molecule. This protein was found on the surface of *Staphylococcus aureus* and is called Ebh (ECM-binding protein homologue, Uniprot Q2FYJ6) [7,8]. These huge proteins, which have also been found on streptococci (i.e. extracellular matrix-binding protein (Embp), Uniprot Q8KQ73) [9], are in addition able to bind fibronectin. They mediate adhesion and have been shown to be required for biofilm formation *in vivo*. An additional mechanism of albumin binding was recently identified when it was shown that human serum albumin (HSA) adsorbed to bacteria could bind to and inactivate the antibacterial chemokine MIG/CXCL9 (monokine-induced by gamma-interferon/CXC ligand), which is released by activated epithelium [10]. This albumin-dependent event protects from the antibacterial activity and promotes bacterial survival at the epithelium. Even though all functions of bacterial surface proteins are not yet fully elucidated, they clearly provide the bacteria expressing them with an evolutionary advantage.

Streptococcal protein G (SPG), which binds to immunoglobulins and albumins of several species, is expressed on the surface of certain streptococcal strains [11-13] and is one of the best-characterized bacterial surface proteins. As indicated in figure 1, SPG from the opportunistic streptococcal strain G148 has two functional regions containing three immunoglobulin-binding (C1-C3) and three albumin-binding domains (ABD1-3), respectively [12,14]. The immunoglobulin-binding domains share a common four-stranded beta-sheet fold with a single alpha helix packed onto the sheet (4β+α) [15]. Of the three homologous albumin-binding domains, the C-terminal ABD3 has been most extensively studied; it is referred to as G148-ABD in the text and G148-ABD3 in figure 2A. Nuclear magnetic resonance (NMR) spectroscopy has established that this 46 amino acid domain folds into a left-handed anti-parallel three-helix bundle (3α) [4,16], similar to the structure of the immunoglobulin-binding domains of the well-studied staphylococcal protein A [17,18]. This structural element is found in several other proteins, which indicates that the 3α-fold is energetically and functionally favorable since it has been utilized broadly [19]. Interestingly, a structural evaluation of the repeating units in the giant albumin-binding protein Ebh showed that its domains, one of which is responsible for albumin binding, are connected by a long helix that

[a]*Division of Protein Technology, School of Biotechnology, KTH Royal Institute of Technology, AlbaNova University Center, SE-106 91 Stockholm, Sweden*

* Corresponding author.
E-mail address: johan.nilvebrant@biotech.kth.se (Johan Nilvebrant)

participates in two three helix bundles in two adjacent repeating units [8]. This helix is responsible for the global rod-like structure of the protein.

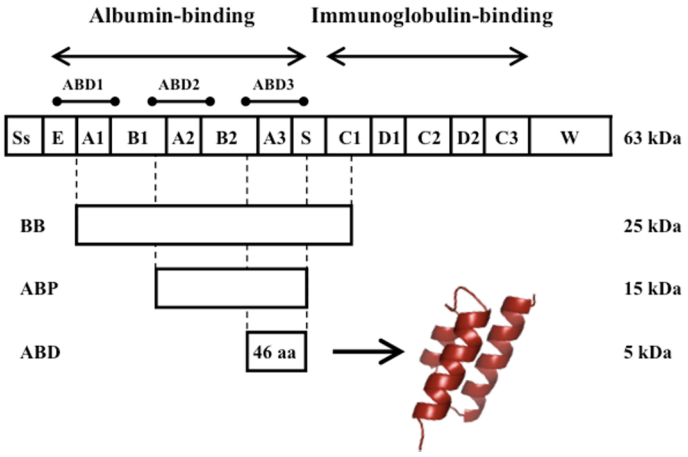

Figure 1. Schematic representation of streptococcal protein G. Protein G consists of an N-terminal signal sequence (Ss), an albumin-binding region containing three albumin-binding domains and a C-terminal immunoglobulin-binding region. A spacer (S) separates the binding regions and a C-terminal sequence (W) anchors the protein to the cell wall. Various albumin-binding parts, BB, ABP and the smallest albumin-binding unit, the 46 amino acid albumin-binding domain (G148-ABD), are indicated. ABD folds into a stable three-helix bundle structure (the picture was generated from PDB-file 1GJT).

Historically, the most widespread use of SPG has been as a biotechnological tool mainly used for affinity purification of immunoglobulins exploiting the broad species- and subclass specificity of its immunoglobulin-binding domains [20,21]. Albumin-binding regions spanning one or several albumin-binding domains, for example BB and ABP [22] that are indicated in figure 1, have been used for affinity purification or depletion of albumin [21]. Moreover, the use of an albumin-binding region as a fusion tag can facilitate affinity purification of a target protein, improve its solubility or be used for directed immobilization [22-24].

Several homologous albumin-binding domains have been identified in surface proteins from different bacterial species. The sequence diversity among these is illustrated by the 16 homologues included in Figure 2A. Alongside G148-ABD, the so-called protein G-related albumin-binding (GA) module from protein PAB (peptostreptococcal albumin-binding) of the anaerobic bacterium *Finegoldia magna* (*F. magna*) has been thoroughly investigated both structurally and functionally [19,25,26]. Analysis of the gene encoding PAB suggested that its albumin-binding domain (ALB8-GA representing the best characterized variant, see Figure 2A) originates from protein G and that it was introduced as a result of an interspecies module-shuffling event [25]. Available data on various albumin-binding domains suggest a correlation between the species specificity of the surface proteins and the host specificity of the bacteria that express them [4]. G148-ABD and ALB8-GA exhibit 59 % amino acid sequence identity, but the species specificity of G148-ABD is much broader than for ALB8-GA whereas the binding affinity of ALB8-GA for HSA is roughly twofold higher. ALB8-GA has only been found in human isolates of *F. magna* and, consequently, it is believed to have evolved to bind HSA with higher affinity than its predecessor. In contrast, streptococci expressing G148-ABD have much broader host specificity and this domain binds albumin from several non-primates better than ALB8-GA [4].

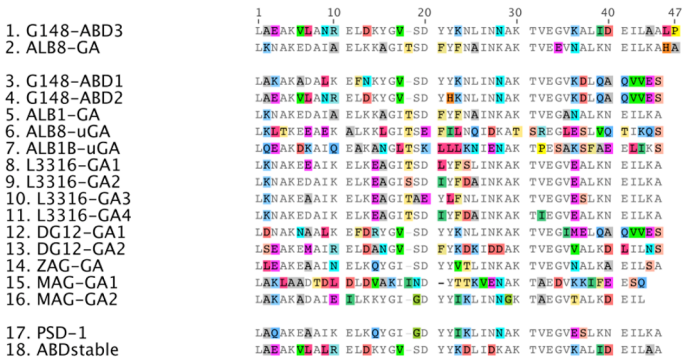

Figure 2A. Sequence alignment of 16 homologous albumin-binding domains and two engineered variants. Conserved amino acids are shown in gray and differences are highlighted in color. G148-ABD3 and ALB8-GA (sequences 1 and 2) represent the best-studied domains. PSD-1 (sequence 17) is an engineered variant with broadened species specificity and ABDstable (sequence 18) is a variant that has been stabilized to alkaline treatment. The picture was generated in Geneious Pro version 5.5.7 created by Biomatters and is based on a similar picture by Johansson et al. [4].

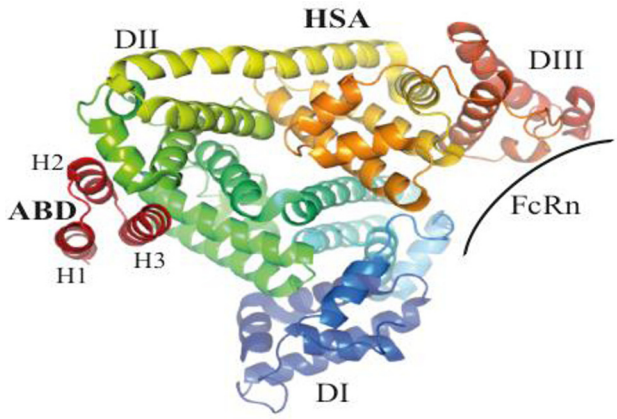

Figure 2B. Structure of the complex formed by ALB8-GA and HSA. The albumin-binding domains recognize a site located in domain II of HSA that does not overlap with the binding site for the neonatal Fc-receptor (FcRn), which plays an important role in albumin homeostasis. The picture was generated from PDB-file 1TF0.

Accumulated structural data on G148-ABD [4,16] and the GA-module [26-29] demonstrate that the domains have very similar tertiary structures. ALB8-GA contains an additional residue in the loop between the first and second helix (Figure 2A) and has a somewhat shorter first helix compared to G148-ABD [4]. The lengths and positions of the second and third helices are almost identical and this region also contains the most highly conserved sequence stretch among the homologues (Figure 2A), which implies that they all share a common overall fold. As would be expected, competitive binding studies have shown that G148-ABD and ALB8-GA have the same binding site on HSA [4]. A crystal structure of ALB8-GA in complex with HSA revealed that this site is located on the exterior of domain II of the albumin molecule [28], figure 2B. The flat binding site consists of a hydrophobic center and two surrounding hydrogen bond networks [28]. A similar structural complex of ALB8-GA and a fatty acid-induced conformational form of HSA demonstrated that both forms could be recognized [29]. Mainly residues in the second helix and the following loop of G148-ABD contribute to albumin binding, as determined by a dedicated mutational study [30].