BBA - General Subjects 1861 (2017) 2196–2205

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

Exploiting sequence and stability information for directing nanobody stability engineering

Patrick Kunz[a,*], Tilman Flock[b], Nicolas Soler[c], Moritz Zaiss[d], Cécile Vincke[e], Yann Sterckx[e], Damjana Kastelic[a], Serge Muyldermans[e,1], Jörg D. Hoheisel[a,1]

[a] Division of Functional Genome Analysis, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany
[b] Division of Structural Studies, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom
[c] Crystallographic Methods, Molecular Biology Institute of Barcelona (CSIC), carrer Baldiri Reixac 4-8, 08028 Barcelona, Spain
[d] Department of High-field Magnetic Resonance, Max-Planck-Institute for Biological Cybernetics, Spemannstraße 41, 72076 Tübingen, Germany
[e] Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

ARTICLE INFO

Keywords:
Single-domain antibody (sdAb, nanobody)
Protein engineering
Protein stability
Protein aggregation
Protein design

ABSTRACT

Background: Variable domains of camelid heavy-chain antibodies, commonly named nanobodies, have high biotechnological potential. In view of their broad range of applications in research, diagnostics and therapy, engineering their stability is of particular interest. One important aspect is the improvement of thermostability, because it can have immediate effects on conformational stability, protease resistance and aggregation propensity of the protein.

Methods: We analyzed the sequences and thermostabilities of 78 purified nanobody binders. From this data, potentially stabilizing amino acid variations were identified and studied experimentally.

Results: Some mutations improved the stability of nanobodies by up to 6.1 °C, with an average of 2.3 °C across eight modified nanobodies. The stabilizing mechanism involves an improvement of both conformational stability and aggregation behavior, explaining the variable degree of stabilization in individual molecules. In some instances, variations predicted to be stabilizing actually led to thermal destabilization of the proteins. The reasons for this contradiction between prediction and experiment were investigated.

Conclusions: The results reveal a mutational strategy to improve the biophysical behavior of nanobody binders and indicate a species-specificity of nanobody architecture.

General significance: This study illustrates the potential and limitations of engineering nanobody thermostability by merging sequence information with stability data, an aspect that is becoming increasingly important with the recent development of high-throughput biophysical methods.

1. Introduction

VHH domains derived from heavy-chain antibodies of camelids, commonly named nanobodies, are single-domain antigen-binding fragments with a large potential for numerous applications in research, biotechnology and medicine [1–4]. They exhibit several superior properties compared to conventional antibody scaffolds, such as enhanced solubility, low immunogenicity and the unique ability to bind cryptic epitopes; the last is due to their small size and a third complementarity determining region (CDR3) of unusual length [5]. The most interesting epitopes for which specific nanobodies have been generated are enzyme active sites, constant regions of virus particles, transient states of membrane proteins and non-amyloidogenic regions of human lysozyme [6–9]. Nanobodies are expected to contribute substantially to meeting the enormous demand for versatile, robust and stable binders.

Their structural simplicity and ease of production make recombinant nanobodies an ideal and easily accessible system for protein engineering approaches. Numerous benefits would arise particularly from engineering their thermostability. Nanobody-based therapeutics need to resist proteolytic degradation, a characteristic which was strengthened for various proteins by increasing their conformational

Abbreviations: VHH, variable domain of the heavy chain of heavy-chain antibodies; GSS algorithm, Global Sequence Signature algorithm; CDR, complementarity determining region; MSA, multiple sequence alignment; CD, circular dichroism; DSF, differential scanning fluorimetry; VH, heavy chain variable domain; VL, light chain variable domain; HM, hallmark position; GdmCl, guanidinium chloride

* Corresponding author.
E-mail address: p.kunz@dkfz.de (P. Kunz).
[1] Equally contributing authors.

http://dx.doi.org/10.1016/j.bbagen.2017.06.014
Received 21 March 2017; Received in revised form 2 June 2017; Accepted 18 June 2017
Available online 20 June 2017

stability [10–12]. Protein aggregation represents another challenge [13,14]. Its relevance becomes more and more apparent for therapeutic strategies, during which antibodies are usually administered at high concentrations [15,16]. While not detectable for some nanobodies [17], aggregation has been observed for molecules of this binder class [9,18]. Partial or full protein unfolding is usually a prerequisite for aggregation, making the increase of conformational stability a reasonable way to prevent this deleterious side reaction. Furthermore, application of nanobodies as *in vivo* tools would greatly benefit from stability-engineered binders. For example, *in vivo* expression in mammalian cells [19] requires nanobody folding in absence of a conserved disulfide bond in the nanobody framework, as the bond remains reduced under cytosolic conditions. Folding that is independent of a disulfide bond is expected to be more robust with stronger non-covalent interactions in the nanobody fold. Finally, recent attempts aimed at the engineering of stabilized nanobodies to construct biosensors and drugs for particularly harsh conditions [20–22], potentially increasing the application range of biological reagents to unaccustomed fields.

Protein thermostability is governed by diverse factors. Successful protein stabilization has been achieved through rigidifying flexible sites [23], optimizing surface charge [24], improving hydrophobic packing [25] and introducing disulfide bonds or salt bridges [26,27]. Interestingly, substantial stabilization was achieved both by the sum of several, marginally stabilizing mutations and by selecting single key positions [28,29]. Accordingly, it remains challenging to identify efficiently stabilizing positions in a protein fold, although several strategies have emerged that could meet this task [30]. Directed evolution methods, for example, explore the vast sequence space by high-throughput experiments [31]. Rational design exploits structural knowledge and general stabilizing features like salt bridges and disulfide bonds [26,27]. Sequence-based strategies identify stabilizing residues from a comparison of mesophilic and thermophilic homologues [32], or rely on the consensus-based strategy that attributes a high probability of stabilization to the most frequent residue at a given sequence position of a protein family [33].

The developments in high-throughput biophysical methods open up new avenues for protein engineering as they amend sequence data with quantitative information on conformation and stability [34–36]. This allows a comprehensive survey of the sequence space of a protein class with respect to thermodynamic parameters, as has been applied to antibody variable domains, for example [37,38]. We studied similar data on nanobody thermostabilities. The merger of sequence and quantitative protein stability information revealed sequence features that are characteristic for highly stable nanobodies. At a global level, our analysis identified species-specific subclasses of nanobody architecture, an insight important for nanobody engineering. With respect to particular sites, we identified stabilizing amino acid variations. In an experimental validation, some residue exchanges improved both conformational stability and aggregation behavior of several nanobodies, which were modified accordingly. Since we were working with recombinant nanobodies, such changes could be introduced easily in a systematic manner, yielding nanobody binders of improved thermostability. However, we also found positions in the nanobody framework that did not act as enhancers of stability if modified as predicted, apparently due to species-dependent interactions in the protein framework. A more complex combination of variations is needed in such cases for improving thermostability.

2. Materials and methods

2.1. Nanobody cloning, expression and purification

The nanobody data set comprised 57 dromedary, 4 alpaca and 17 llama nanobodies. Dromedary and alpaca nanobodies were obtained from phage-display screenings, representing high-affinity binders against different protein targets. They are cloned in the pMECS vector with a C-terminal HA- and His6-tag [39]. The llama nanobodies were obtained from a subtractive phage-display library against tumor lysates [40], cloned into the pHEN2 plasmid with a C-terminal Myc- and His6-tag [41]. N-terminal variants of the dromedary nanobodies were obtained by PCR using mutated primers (biomers.net, Ulm, Germany), while all other variants were purchased from Gen-9 (Cambridge, USA). All variants were cloned into the pMECS vector using *NcoI* and *NotI* restriction sites and verified by sequencing. Nanobody constructs present in pMECS and pHEN2 plasmids were expressed in the periplasm of *E. coli* cells WK6 or TG1, respectively, and purified as described [6]. Using sequence-based extinction coefficients [42], the nanobody concentration was assessed at 280 nm using a Nanodrop ND-1000 instrument (Peqlab Biotechnologie, Erlangen, Germany), measuring absorption at least in triplicate. The molecular weight of the N-terminal variants was verified by MALDI mass spectrometry. N- and C-terminal sequences were checked by in-source-decay [43].

2.2. Stability measurements

Melting temperatures were assessed in triplicate by the Thermofluor assays using Sypro Orange (Thermo Fisher Scientific, Waltham, USA) at a nanobody concentration of 0.5 mg/ml. Fluorescence was detected every 45 s using a LightCycler 480 (Roche Diagnostics, Mannheim, Germany) at a heating rate of 0.5 °C/min. Fluorescence traces were evaluated using a Matlab script, determining the transition point by finding the maximum of the numerical derivation of the smoothed trace data. Smoothing was performed by a linear box filter of 200 data points width. N-terminal variants were measured in triplicate by DSF using a Prometheus NT.48 instrument with back-reflection optics (NanoTemper Technologies, Munich, Germany) after dialysis against a PBS solution at pH 7.4 (Slide-A-Lyzer MINI, 3.5 kDa MWCO; Thermo Fisher Scientific, Waltham, USA). The molar protein concentration was adjusted to 32.7 μM (corresponding to 0.5 mg/ml), if not otherwise stated. The heating rate was 0.5 °C/min. T_m values were determined using the product software package based on first derivative analysis. Gibbs free energy of unfolding was measured by equilibrium denaturation on the same Prometheus NT.48 instrument using guanidinium chloride as denaturant and fitting the data to a two-state model of equilibrium unfolding [44]. Guanidinium chloride concentrations were determined using refractive indices according to Santoro and Bolen [45]. Since protein stickiness had been observed in the fluorescence cuvette filled with nanobody NbD1 or its N-terminal variant at a concentration of 2 μM, the protein concentration was adjusted to 30 μM to avoid artifacts resulting from surface binding. For both the Thermofluor and the DSF assays, the differences measured for mutations Q1E and Q5V were substantiated statistically by applying a paired *t*-test and a Wilcoxon signed-rank test.

2.3. Nanobody structure determination

Nanobody NbD2 (PDB ID: 5M7Q) was crystallized in 100 mM MES, pH 6.5, 1.6 M magnesium sulfate heptahydrate and was cryo-preserved in 100 mM MES pH 6.5, 1.375 M magnesium sulfate heptahydrate and 25% glycerol. Data were collected on beamline PROXIMA 1 at the SOLEIL synchrotron and processed with iMOSFLM [46]. Space group determination, scaling and merging was performed by means of the Pointless/Aimless routines of the CCP4 suite [47]. PHASER [48] provided a good molecular replacement solution using RCSB protein database entry 4W6X as a model. Refmac [49] was used for structure refinement while model building of the two copies contained in the asymmetric unit was performed with coot [50]. Data collection and refinement statistics are given in Supplementary Table 1.

2.4. Global Sequence Signature analysis

The GSS analysis [51] is based on the concept that any property of a