Protein Engineering, Design & Selection vol. 22 no. 3 pp. 217–220, 2009
Published online October 28, 2008 doi:10.1093/protein/gzn059

SHORT COMMUNICATION

Sequence determinants of protein aggregation in human VH domains

Kip Dudgeon[1,3], Kristoffer Famm[2,3] and Daniel Christ[1,4]

[1]Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, NSW 2010, Australia and [2]Centre for Protein Engineering, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, UK

[3]Both authors contributed equally to this work

[4]To whom correspondence should be addressed.
E-mail: d.christ@garvan.org.au

Human antibody variable heavy (VH) domains tend to aggregate upon denaturation, for instance, by heat or acid. We have previously demonstrated that domains resisting protein aggregation can be selected from CDR-only repertoires by phage display. Here we analysed their sequences to identify determinants governing protein aggregation. We found that, while many different CDR sequences conferred aggregation–resistance, certain physico-chemical properties were strongly selected for. Thus, hydrophobicity and beta-sheet propensity were significantly lower among the selected domains, whereas net negative charge was increased. Our results provide guidelines for the design of human VH repertoires with reduced levels of protein aggregation.

Keywords: antibodies/combinatorial repertoire/molecular evolution/phage display/protein aggregation

Many human proteins aggregate, particularly in response to non-physiological conditions such as heat and acidic pH. This is also the case for human antibody domains, such as variable heavy (VH) chains, which tend to aggregate in isolation. This is in marked contrast to VH from other species, such as camels, which naturally occur in the absence of light-chain partners. Structural studies of camelid single domains have revealed specific features that are believed to improve their solubility (Desmyter *et al*., 1996; Spinelli *et al*., 1996). These include increased hydrophilicity of surface patches that form the light chain interface in other species. Consequently, mutation of residues in this interface has been shown to improve solubility and to reduce aggregation of human VH domains (Davies and Riechmann, 1994; Barthelemy *et al*., 2008).

However, other studies have shown that similar aggregation–resistant properties can be observed among human VH domains lacking mutations in the light chain interface (Jespers *et al*., 2004b). These domains had been derived from a synthetic antibody library based on a single human framework, but different CDR regions, highlighting the effect of variable regions on the biophysical properties of human VH domains. Other studies have reported similar results, with CDR sequences shown to affect expression yields, retention on gel-filtration columns and thermodynamic stability (Jespers *et al*., 2004a; Famm *et al*., 2008; Barthelemy *et al*., 2008).

We have previously reported a method that allowed the selection of human variable heavy chains resisting aggregation from a synthetic single framework CDR-only library (Jespers *et al*., 2004a; Famm and Winter, 2006). This method was based on the display of VH domains on phage, followed by heat-induced unfolding and capture of aggregation–resistant domains using protein A, a conformation-specific superantigen. The domains selected by this method not only resisted aggregation when heated above their unfolding temperature, but also displayed favourable expression and purification yields (Jespers *et al*., 2004a). However, while we noted that many of the selected domains exhibited acidic isoelectric points, detailed sequence characteristics remained unknown. Here we describe the CDR-sequence determinants governing protein aggregation in human VH domains.

Material and Methods

Selection for aggregation–resistance on phage

Phages from a VH CDR-only library based on the human V3-23/D47 framework [library 1, 2×10^9 clones (Jespers *et al*., 2004a)] were purified from culture supernatant by two precipitations with polyethylene-glycol/NaCl solution [as previously described (Lee *et al*., 2007)]. Phages were subjected to a single round of heat treatment [in contrast to three rounds of treatment used by Jespers *et al*. (Jespers *et al*., 2004a)]. Approximately 10^{10} phages were heated in 10 μl of phosphate-buffered saline (PBS) at 80°C for 10 min and cooled to 4°C (heat-selected repertoire). As a control, phages were incubated in PBS at room temperature (unselected repertoire). Phages from both repertoires were diluted into PBS supplemented with 2% (w/v) marvel milk powder (MPBS) and captured on a protein A coated MaxiSorp Immuno-plate (Nunc) (which had been coated over night with protein A at 10 μg/ml in PBS at room temperature and blocked using 2% MPBS). The selected phages were added to the blocked wells for 2 h at room temperature. After 10 washes with PBS, protein A-bound phages were eluted in 1 mg/ml trypsin in PBS for 10 min at room temperature. The eluted phages were used to infect exponentially growing *Escherichia coli* TG1 bacteria (Gibson, 1984) (with an OD_{600} of approximately 0.5) by incubation for 30 min at 37°C. Bacteria were pelleted by centrifugation, plated on TYE agar plates supplemented with 15 μg/ml tetracycline and grown at 37°C over night.

Phage ELISA for aggregation–resistance

Aggregation–resistance of clones was analysed by measuring retention of signal in a phage ELISA format (McCafferty *et al*., 1990). Wells of a Nunc Maxisorp Immuno-plate were coated over night with protein A at a concentration of 2.5 μg/ml in PBS. The plate was washed

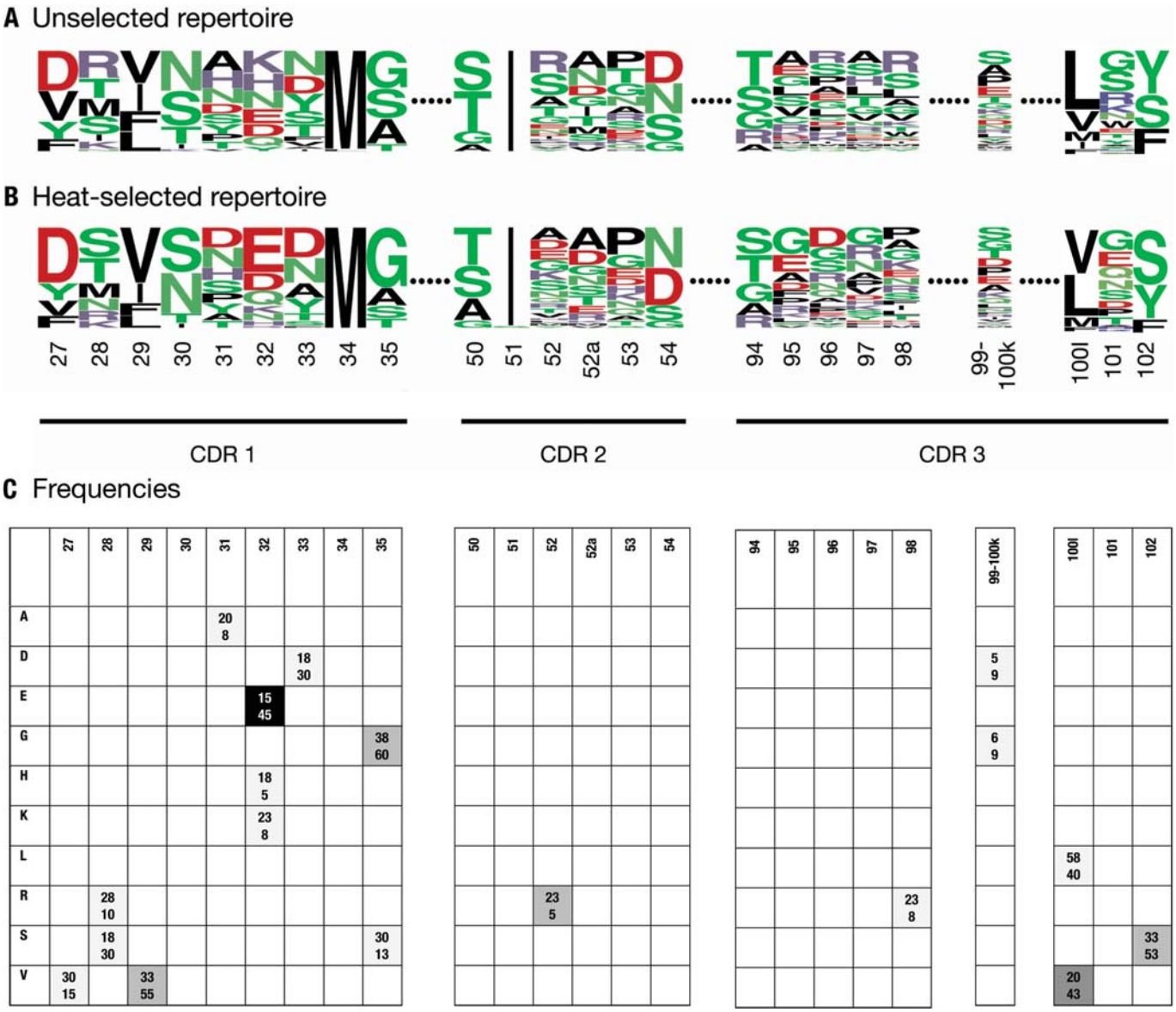

Fig. 1. CDR amino acid frequencies of VH repertoires. Frequencies were calculated from sequences of 40 clones from the unselected VH repertoire (**A**), as well as 40 clones selected by heating on phage (**B**). Amino acids are indicated by single letter codes and frequencies correlate with letter size. Amino acids are colour-coded according to physico-chemical properties: green (G, S, T, Y, C, Q, N) indicates polar residues, blue (K, R, H) basic residues, red (D, E) acidic residues, black (F, V, L, A, M, I, W, P) hydrophobic residues. Numbering of CDR residues according to Kabat and Wu (Kabat and Wu, 1991). Residues 99–100 k are shown as a single, averaged position to account for variable CDR 3-length in repertoires. (A) and (B) were generated using WebLogo (Crooks *et al.*, 2004). (**C**) shows amino acid frequencies (in percent) at positions displaying significant sequence bias. Upper values refer to the unselected repertoire, lower values to the heat-selected repertoire. Shading indicates statistically significant bias as follows: light gray $p < 0.05$; gray $p < 0.01$; dark gray $p < 0.001$; black $p < 10^{-6}$. p-values were calculated using χ^2 statistics with one degree of freedom; amino acids rare at a given position in the unselected repertoire (frequency of 10% or less) were excluded from the analysis.

once with PBS and blocked with 2% MPBS. Single colonies were picked from agar plates and grown over night in 2 × TY medium supplemented with 15 μg/ml tetracycline in 96-well round bottom plates shaking at 37°C. Cells were removed by centrifugation and phages were biotinylated directly in the culture supernatant by adding biotin-PEO4-N-hydroxysuccinimide (Pierce; 50 μM final concentration). For heat selection, supernatant was incubated at 80°C for 10 min and then cooled to 4°C. Supernatant was added to the blocked ELISA wells. After three washes with PBS, bound phage particles were detected using an Extravidin-HRP conjugate (Sigma) and 3,3′,5,5′-tetramethylbenzidine (TMB) substrate.

Sequencing of repertoires

Regions of the phage genome encoding the VH domains were amplified by PCR from bacterial cultures using primers 5′-CGGTTTATCAGCTTGCTTTCG-3′ and 5′-GTTTTGTCG TCTTTCCAGACG-3′ and subjected to DNA sequence analysis.

Results

We first selected aggregation–resistant domains by heating a large CDR-only VH library (2×10^9 clones) (Jespers *et al.*, 2004a) on phage: the phages were heated to 80°C, followed

Table I. Physico-chemical properties of VH domains and their CDR regions

	Repertoire	Net charge	Hydrophobicity	β-sheet propensity
VH	Unselected	1.823 (1.840)	−0.001 (0.036)	0.328 (0.005)
	Heat-selected	−0.707 (1.579)***	−0.038 (0.034)***	0.336 (0.007)***
CDR 1 only	Unselected	−0.336 (1.076)	−0.071 (0.269)	0.314 (0.052)
	Heat-selected	−1.393 (1.058)***	−0.258 (0.256)**	0.348 (0.056)**
CDR 2 only	Unselected	−0.174 (0.902)	−0.152 (0.345)	0.318 (0.064)
	Heat-selected	−0.646 (0.970)*	−0.255 (0.299)	0.351 (0.065)*
CDR 3 only	Unselected	0.361 (1.391)	−0.099 (0.263)	0.321 (0.039)
	Heat-selected	−0.640 (1.848)**	−0.224 (0.228)**	0.351 (0.041)**

Properties were calculated from sequences of 40 clones from the unselected repertoire, as well as 40 clones selected by heating on phage. Shown are population means with standard deviations displayed in brackets. *p*-values were calculated using two-tailed heteroscedastic *t*-tests. CDR positions included in the analysis were CDR 1 (27–35), CDR 2 (50–52, 52a, 53, 54), CDR 3 (94–98, 99–100 k, 100l, 101, 102). Net charges were calculated at pH 7.4 based on Henderson–Hasselbach equation using side-chain pKa values (Fersht, 1998). Hydrophobicity values were calculated at pH 7.5 according to Cowan and Whittaker (Cowan and Whittaker, 1990). Beta-sheet propensities were calculated according to Chiti (Chiti *et al*., 2003) (proline residues were excluded from the analysis); propensities range from 0 (high) to 1 (low). *Indicates $p < 0.05$; **Indicates $p < 0.01$; ***Indicates $p < 0.001$.

by capture of domains resistant to heat-induced aggregation by protein A (heat-selected repertoire). As a control, the VH library was incubated at room temperature and captured by protein A (unselected repertoire). We next analysed the aggregation propensity of the two repertoires by phage ELISA. For this purpose, phages were heated and aggregation–resistant clones detected with an anti-phage conjugate after protein A capture. Using this assay, we identified 40 clones from each repertoire, which differed considerably in their aggregation–resistance (with an average retention of ELISA signal after heating of 88% for the heat-selected clones and 42% for the unselected clones).

As the clones from the two repertoires strongly differed in their resistance to aggregation, we wondered whether this property was reflected in their sequence characteristics. We therefore determined their sequences and calculated amino acid frequencies at all individual CDR positions. This revealed considerable diversity at all of the analysed positions (Fig. 1A and B). However, further analyses found that the frequencies of amino acids at several positions differed significantly between the two repertoires, indicating selection pressure associated with the heat-denaturation step (Fig. 1C). This included strong enrichment ($p < 0.01$) of valine at position 29, glycine at position 35 and serine at position 102. A considerable preference ($p < 0.001$) was observed for valine at position 100l and in particular for glutamate at position 32 ($p < 10^{-6}$). We also found that certain amino acids were less frequent among the selected domains, in particular arginine and lysine. However, with the exception of arginine at position 52, these biases were generally weak ($p < 0.05$). Many of these positions are located within CDR 1. In total, we detected six positions, three in CDR 1 (29, 32, 35), one in CDR 2 (52) and two in CDR 3 (100l, 102) that differed strongly ($p < 0.01$) in their amino acid composition between the two repertoires.

We next determined physico-chemical properties of longer stretches of the proteins, both at the domain-level (Table I, first row), as well as for individual CDRs (Table I, rows 2–4). This revealed strong increase of net negative charge for the heat-selected VH domains, as well as reduced hydrophobicity. This is in agreement with the enrichment of acidic and polar residues at individual positions (as above). We also observed that beta-sheet propensity of the domains was significantly reduced. Similar trends were observed for individual CDRs, particularly CDR 1 and, to a lesser degree, CDR 3 and CDR 2 (Table I; rows 2–4).

In addition to heat-induced aggregation, we also analysed the sequence features of a repertoire of VH domains that had previously been selected by incubation at acidic pH (Famm *et al*., 2008). In contrast to the domains selected by heat, we did not observe physico-chemical properties to differ significantly from the unselected repertoire (data not shown). This is in agreement with our earlier findings that these proteins resist acid-induced aggregation mainly through the selection of specific point mutations (such as Arg28) (Famm *et al*., 2008).

Discussion

The use of phage display for the selection of antibody domains from a large combinatorial repertoire has allowed us to identify CDR-sequence determinants governing aggregation–resistance. For the VH domains that had been selected by heating on phage, we detected significant enrichment of amino acids at several positions, particularly within CDR 1. This included glycine 35, a residue previously shown to increase the solubility of VH domains (Jespers *et al*., 2004b; Barthelemy *et al*., 2008). However, other substitutions had not been previously described, including a strong preference for glutamate at position 32.

We also wondered whether we could detect changes that may not be obvious from the analyses of amino acid frequencies at single positions. Such changes may originate from the effect of combined amino acid changes at multiple positions or may be obscured by co-variation. Further analysis did indeed reveal properties that had not been obvious from the analyses of single positions. Thus, hydrophobicity and beta-sheet propensity were significantly lower among the selected domains, while net negative charge was increased. All of the observed properties have previously been implicated in protein aggregation and the formation of amyloid-like aggregates (Dobson, 1999). These properties are also the basis of algorithms that aim to predict the formation of protein aggregates (DuBay *et al*., 2004; Fernandez-Escamilla *et al*., 2004). Indeed, we observed that these algorithms predicted lower levels of protein aggregation for heat-selected VH repertoires (data not shown).

The effect of charge on protein aggregation and solubility has long been known and proteins display minimal solubility near their isoelectric points (Loeb, 1921). Our observation that increased negative charge plays a major role in aggregation–resistance is in agreement with reports on the engineering of proteins other than VH. In a recent study, Whitesides and co-workers reported that increasing negative charge in alpha-amylase (by chemical acetylation of lysines) significantly improved resistance against thermal and chemical inactivation (Shaw *et al*., 2008). Liu and co-workers reported the engineering of green fluorescent protein (GFP) into a highly aggregation–resistant form by the introduction of negatively charged surface residues (Lawrence *et al*., 2007). In contrast to our results, they also observed similar effects for a highly positively charged GFP variant (although not for other proteins, such as glutathione-S-transferase).

In summary, we found that CDRs with highly diverse sequences, but common physico-chemical properties, can give rise to human variable domains resisting protein aggregation. This is in agreement with previous results, demonstrating that repertoires of aggregation–resistant domains can be generated by combinatorial assembly of pre-selected CDR regions (Christ *et al*., 2006). Although good antigen binders were obtained in these experiments, pre-selection did lead to a considerable loss of CDR-diversity (Christ *et al*., 2007). Our results here suggest an alternative path to the construction of such repertoires. Thus, the identification of sequence determinants governing protein aggregation may open up the path towards the design of human VH repertoires with reduced levels of protein aggregation. This may include the use of trinucleotide mutagenesis (Virnekas *et al*., 1994) to limit diversity at restrictive positions (such as CDR 1, residue 32) and the targeting of diversity to other regions (such as CDR 2) that display little sequence bias in our analyses.

Funding

The work was funded by the United Kingdom Medical Research Council (K.F.), the Garvan Institute of Medical Research, the Australian National Health and Medical Research Council and the Cancer Institute NSW (K.D. and D.C.).

References

Barthelemy,P.A., Raab,H., Appleton,B.A., Bond,C.J., Wu,P., Wiesmann,C. and Sidhu,S.S. (2008) *J. Biol. Chem.*, **283**, 3639–3654.
Chiti,F., Stefani,M., Taddei,N., Ramponi,G. and Dobson,C.M. (2003) *Nature*, **424**, 805–808.
Christ,D., Famm,K. and Winter,G. (2006) *Nucleic Acids Res.*, **34**, e108.
Christ,D., Famm,K. and Winter,G. (2007) *Protein Eng. Des. Sel.*, **20**, 413–416.
Cowan,R. and Whittaker,R.G. (1990) *Pept. Res.*, **3**, 75–80.
Crooks,G., Hon,G., Chandonia,J. and Brenner,S. (2004) *Genome Res.*, **14**, 1188–1190.
Davies,J. and Riechmann,L. (1994) *FEBS Lett.*, **339**, 285–290.
Desmyter,A., Transue,T.R., Ghahroudi,M.A., Thi,M.H., Poortmans,F., Hamers,R., Muyldermans,S. and Wyns,L. (1996) *Nat. Struct. Biol.*, **3**, 803–811.
Dobson,C.M. (1999) *Trends Biochem. Sci.*, **24**, 329–332.
DuBay,K.F., Pawar,A.P., Chiti,F., Zurdo,J., Dobson,C.M. and Vendruscolo,M. (2004) *J. Mol. Biol.*, **341**, 1317–1326.
Famm,K. and Winter,G. (2006) *Protein Eng. Des. Sel.*, **19**, 479–481.
Famm,K., Hansen,L., Christ,D. and Winter,G. (2008) *J. Mol. Biol.*, **376**, 926–931.
Fernandez-Escamilla,A.M., Rousseau,F., Schymkowitz,J. and Serrano,L. (2004) *Nat. Biotechnol.*, **22**, 1302–1306.
Fersht,A.R. (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Freeman.
Gibson,T.J. (1984) PhD thesis, University of Cambridge, Cambridge.
Jespers,L., Schon,O., Famm,K. and Winter,G. (2004a) *Nat. Biotechnol.*, **22**, 1161–1165.
Jespers,L., Schon,O., James,L.C., Veprintsev,D. and Winter,G. (2004b) *J. Mol. Biol.*, **337**, 893–903.
Kabat,E.A. and Wu,T.T. (1991) *J. Immunol.*, **147**, 1709–1719.
Lawrence,M.S., Phillips,J.P. and Liu,D.R. (2007) *J. Am. Chem. Soc.*, **129**, 10110–10112.
Lee,C.M., Iorno,N., Sierro,F. and Christ,D. (2007) *Nat. Protoc.*, **2**, 3001–3008.
Loeb,J. (1921) *J. Gen. Physiol.*, **4**, 547–555.
McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) *Nature*, **348**, 552–554.
Shaw,B.F., Schneider,G.F., Bilgicer,B., Kaufman,G.K., Neveu,J.M., Lane,W.S., Whitelegge,J.P. and Whitesides,G.M. (2008) *Protein Sci.*, **17**, 1446–1455.
Spinelli,S., Frenken,L., Bourgeois,D., de Ron,L., Bos,W., Verrips,T., Anguille,C., Cambillau,C. and Tegoni,M. (1996) *Nat. Struct. Biol.*, **3**, 752–757.
Virnekas,B., Ge,L., Pluckthun,A., Schneider,K.C., Wellnhofer,G. and Moroney,S.E. (1994) *Nucleic Acids Res.*, **22**, 5600–5607.

Received September 30, 2008; revised September 30, 2008; accepted October 6, 2008

Edited by P. Holliger