doi:10.1016/j.jmb.2003.10.040 J. Mol. Biol. (2004) 335, 41–48

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

Optimization of the Antibody C_H3 Domain by Residue Frequency Analysis of IgG Sequences

Stephen J. Demarest*, Jeff Rogers and Geneviève Hansen

Torrey Mesa Research Institute 3115 Merryfield Row, San Diego, CA 92130, USA

In an attempt to enhance the overall assembly, yield and half-life of recombinant antibody proteins, we have cloned and expressed several IgG1 C_H3 domains and examined their folding/refolding characteristics. We utilized a cytoplasmic bacterial expression system with a thioredoxin reductase knock-out strain of BL21(DE3) to produce bovine, murine and human C_H3. Under identical conditions, expression of bovine C_H3 resulted consistently in the highest yields of properly folded/oxidized protein. Circular dichroism and fluorescence experiments demonstrate that oxidized bovine and murine C_H3 have surprisingly similar structures and stabilities, considering the marginal sequence conservation between the two molecules. Residue frequency analysis using a limited data set of 36 unique Fc sequences originating from 19 different mammalian species targeted five specific sites for optimization within bovine C_H3. Combination of three of these mutants increased the thermal stability of the molecule to 86 °C. Comparison of this approach to similar studies using larger sequence databases and/or different selection criteria suggests sequence database design can increase the success rate for identifying residue sites worth optimizing. This optimized C_H3 domain can be used as a particularly stable platform for functional design and can be grafted into full-length antibody sequences to enhance their thermodynamic parameters and shelf-life.

*Corresponding author

Keywords: antibody; immunoglobulin; C_H3; protein design; protein folding

Monoclonal antibody-based pharmaceuticals represent a large and growing fraction of all drug candidates currently entering clinical trials.[1] Current technologies have enabled the production of antibodies capable of recognizing virtually any antigen or molecular target with extremely high affinity. Additionally, antibodies can identify and neutralize pathogens (bacteria, viruses), toxins and cancerous cell types. Thus, antibodies have emerged as powerful diagnostic tools in all aspects of life sciences, as well as an attractive means of creating highly specific therapeutics.

Mass production of antibodies is costly and requires tedious processing before the product is ready for market.[2] For these reasons, many recombinant methodologies for producing monoclonal antibodies are being investigated, including production in mammalian, plant, yeast and bacterial cell systems.[3,4] The ability to increase the yield and molecular half-life of fully assembled and functional antibody products within these expression systems is also a major concern.[5]

Here, we describe our approach to optimize the portion of the IgG Fc responsible for heavy chain association (i.e. the C_H3 domain; Figure 1). Traditionally, research has focused on the study of antibody variable domains to understand their

Supplementary data associated with this article can be found at doi: 10.1016/y.jmbi.2003.10.040

Present addresses: S. J. Demarest, J. Rogers, G. Hansen, Diversa Corp., 4955 Directors Place, San Diego, CA 92121, USA.

Abbreviations used: ASA, accessible surface area; bC_H3, bovine C_H3; Fc, antibody heavy chain C_H2C_H3 dimer; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; LB, Luria broth; hC_H3, human C_H3; mC_H3, murine C_H3; NCBI, National Center for Biotechnology Information; scFv, single-chain antibody variable domains; T_m, temperature midpoint of thermal denaturation.

E-mail address of the corresponding author: stephen.demarest@diversa.com

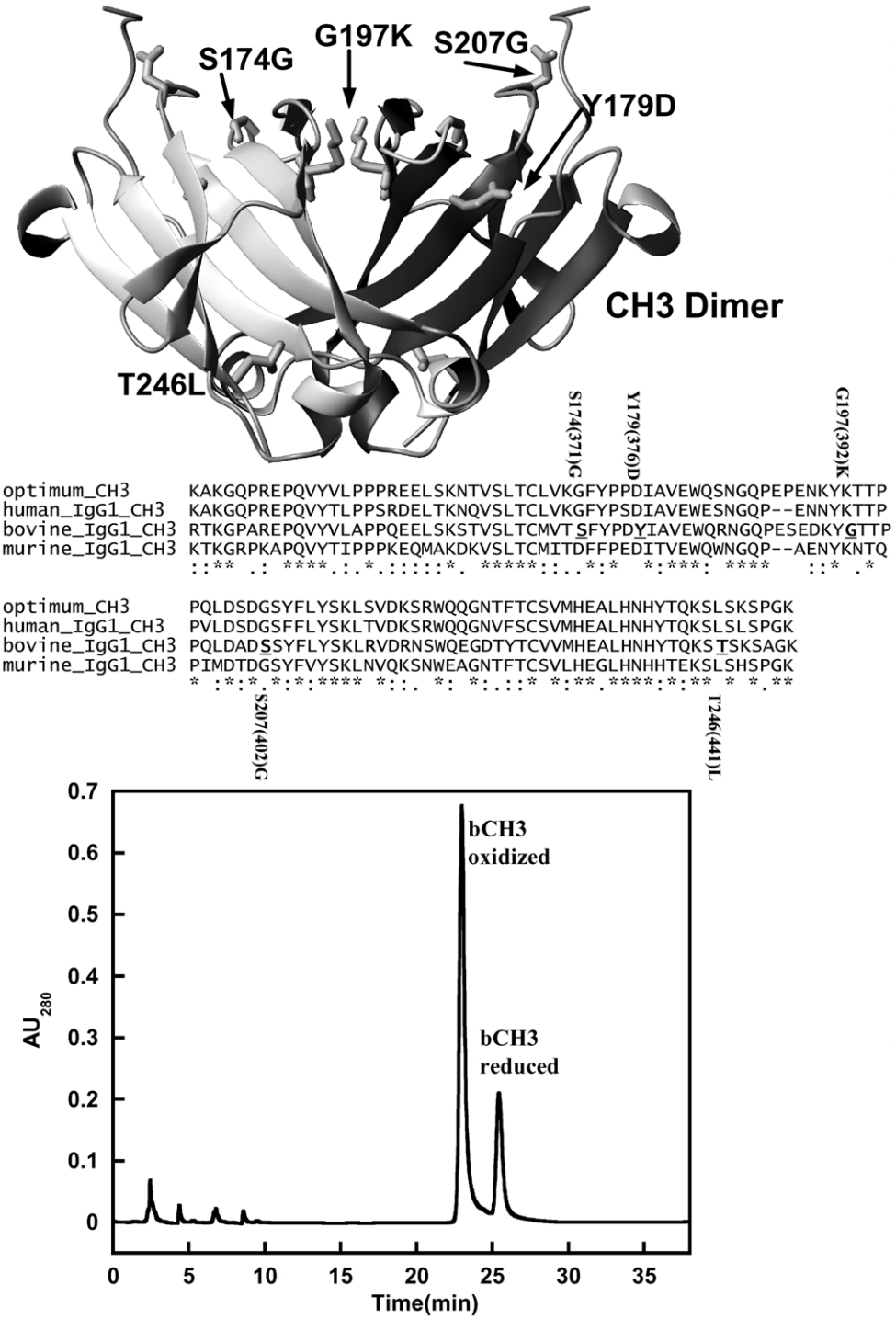

Figure 1. Top: A ribbon diagram of the X-ray crystal structure of human C_H3.[20,39] Residue positions identified as potential sites for optimization on the basis of residue frequency analysis are displayed in stick format. The ribbon diagram was rendered using MolMol.[32] Middle: Alignment of the consensus sequence from a non-redundant database of 36 NCBI mammalian IgG Fc sequences against the human, bovine and murine IgG1 C_H3 sequences. The human C_H3 domain was cloned from Genbank source #BC026038 purchased from the ATCC. Bovine C_H3 was cloned from in-house bovine constructs derived from a bovine spleen cDNA library purchased from Stratagene. Murine C_H3 was cloned from an in-house synthetic murine IgG1 gene based on the murine IgG1 Kabat database consensus sequence.[33] C_H3 PCR inserts and commercial pET21a plasmid (Novagen) were digested using the designed *Bam* HI and *Xho* I restriction sites (enzymes from Invitrogen). Inserts were ligated into the vector using the H.C. T4 DNA ligase (Invitrogen). The QuickChange® Site-Directed Mutagenesis kit (Stratagene) was used to create mutations in the bovine C_H3 pET21a vector. Residue positions within the bovine C_H3 domain where directed mutations were made are indicated. Below the alignments, (.), conservative variation; (:), strict variation; (*), no variation. All inserts and mutations were confirmed by DNA sequencing. BL21-trxB(DE3) (Novagen) was utilized as the host for IPTG-inducible cytoplasmic expression. BL21(DE3) did not yield significant amounts of oxidized protein and Rosetta-Gami (Novagen) had a significantly longer growth phase without providing a significant increase in C_H3 oxidation over the BL21-trxB(DE3)

antigen-binding capabilities.[6] However, effector functions that allow antibodies to involve additional factors within the immune system are embedded within the constant domains.[7,8] For these reasons, the past few years have witnessed an explosion in the sequencing of antibody constant domains. We utilize the publicly available Fc sequences to find residue positions within the bovine IgG1 C_H3 fold that are poorly represented and describe a limited set of mutations at these sites designed to enhance the thermal stability of the domain.

The C_H3 domain of bovine IgG1 was chosen as our model system on the basis of analysis of a condensed set of 36 IgG Fc sequences. These sequences were collected by performing the Tera-Blast™ routine on the non-redundant protein sequence database maintained by NCBI (The National Center for Biotechnology Information) using bovine, murine and human Fc sequences as bait. Other than primate IgG sequences, bovine IgG1 demonstrated the closest resemblance to the consensus sequence, while exhibiting a few definitive sites that, once mutated, might lead to increases in thermal stability. The sequences of bovine, human and murine IgG1 C_H3 (denoted bC_H3, hC_H3 and mC_H3, respectively) are shown in Figure 1. Proteins with disulfides have traditionally been expressed in the periplasmic space of bacteria.[38] However, cytosolic yields tend to be much higher; therefore, bC_H3 was expressed in the cytoplasm of several bacterial cell lines. The best results were obtained in BL21-trxB(DE3), a bacterial cell line carrying a mutation in the thioredoxin reductase gene, *trxB*.[10] bC_H3 was expressed at 20–40 mg/l in a shaker flask in BL21-trxB(DE3). The protein was affinity-purified on Ni^{2+}-NTA resin (Qiagen) using the C-terminal hexahistidine tag. Oxidized protein was separated from reduced material using reverse-phase HPLC (Figure 1). Purified yields of folded/oxidized protein were approximately 70% of the expressed material. The exact mass of oxidized bC_H3, 15064.9402 AMU, was determined by electrospray, time-of-flight (TOF) mass spectroscopy. The reduced fraction (analyzed in the presence of DTT) yielded a major peak with a molecular mass of 15066.9403 AMU, exactly 2 AMU higher than the oxidized form, due to the protonation of the free thiol groups. The theoretical masses of oxidized and reduced bC_H3 are 15066.6 and 15068.6, respectively, and are indistinguishable within the error of the measurement from the experimentally derived values. Oxidized mC_H3 could be obtained in only small quantities (~1 mg per 1 l shaker flask) and oxidized hC_H3 could not be separated from its reduced form using our purification scheme.

The C_H3 domain is known to form a specific non-covalent dimer.[11] Oxidized bC_H3 runs appropriately as a dimer on a BioSpec2000 (Pharmacia) HPLC gel-filtration column. Molecular mass standards included bovine serum albumin (66 kDa), carbonic anhydrase (28 kDa), lysozyme (14 kDa) and ubiquitin (8.5 kDa). The elution time of bC_H3 was constant over a 200 nM–200 μM injected range of concentrations. Similar to bC_H3, oxidized mC_H3 appears to be a dimer, as judged by gel-filtration. Purified hC_H3 elutes in several forms, predominantly trimer, dimer and monomer, an indication that it is not wholly folded/oxidized. Repeated expressions/purifications of hC_H3 never produced viable protein.

All three C_H3 domains exhibited significant secondary structure, as judged by their circular dichroism (CD) spectra (Figure 2A). Far-UV CD spectra of bC_H3 solubilized from cell pellets using 8 M urea and subsequently Ni^{2+}-purified in urea and refolded were identical with the bC_H3 spectra solubilized and purified using strictly aqueous buffers. bC_H3, mC_H3 and hC_H3 all have minima at 217 nm, a characteristic of β-sheet structures consistent with the antibody fold being composed of a twofold β-sheet sandwich. The hC_H3 spectrum was significantly different in intensity and shape from the spectra of bC_H3 and mC_H3, but was very similar to what was observed for reduced bC_H3 (data not shown). Surprisingly, the bC_H3 and mC_H3 spectra were nearly superimposable, indicating that their structures must be nearly identical. This was unexpected considering bC_H3 and mC_H3 are only 54% identical in sequence.

Temperature-induced unfolding of the C_H3 domains was monitored separately using fluorescence, near-UV CD and far-UV CD. The fluorescence excitation wavelength was 278 nm with a 3 nm bandwidth and no emission cutoff filter; the near-UV CD signal was monitored at 278 nm using a 2 nm bandwidth; and the far-UV CD signal was monitored at 217 nm using a 1 nm bandwidth. bC_H3 and mC_H3 undergo single unfolding transitions with midpoints (T_m) at 348.8(±0.5) K and 346.8(±0.5) K, respectively. Refolding of both bC_H3 and mC_H3 was reversible. The CD spectra at 278 K (5 °C) prior to and one hour after temperature-denaturation were virtually unchanged.

The T_m of both bC_H3 and mC_H3 were unaffected by concentration. Fluorescence measurements were performed using 0.5 μM and 4 μM protein,

strain. Bottom: Purification of bC_H3. Bacterial pellets containing C_H3 constructs were dissolved in 8 M urea (pH 8), spun down and the supernatants were passed over Ni^{2+}-NTA resin (Qiagen). Protein was eluted from the resin by adding 8 M urea (pH 4). Reverse-phase HPLC was performed on a Dionex DX500 chromatography system using a Phenomenex C_5-Prep column with water/acetonitrile gradients and an absorbance detector set to 280 nm. The bottom panel displays the chromatograph of bC_H3. The peak eluting at 22.9 minutes is oxidized C_H3, while the peak eluting at 25.4 minutes is reduced C_H3.

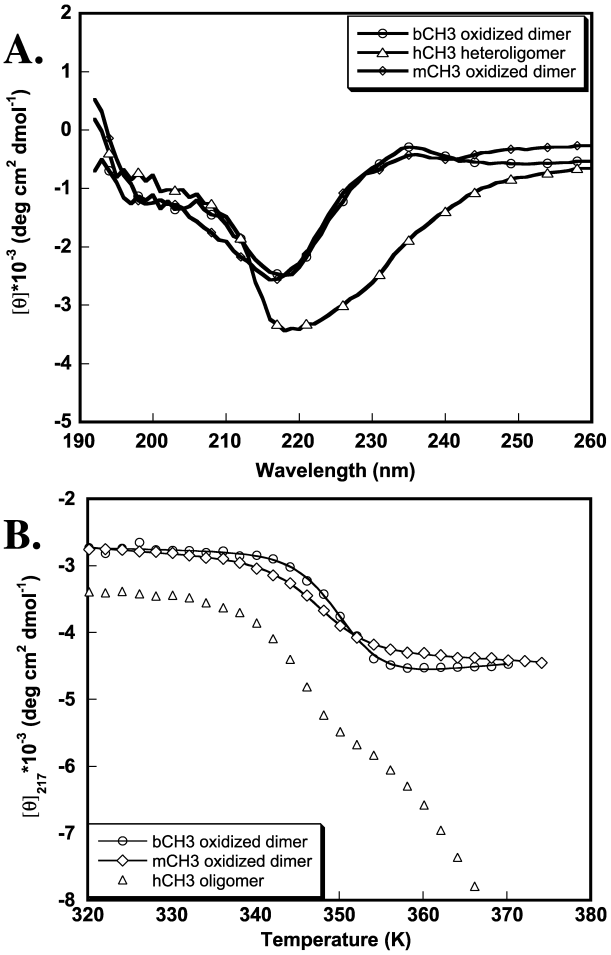

Figure 2. Structure and stability analyses of bC_H3, mC_H3 and hC_H3. A, The CD spectra of bC_H3, mC_H3 and hC_H3 at 5 °C. The CD and fluorescence measurements were performed using an Aviv model 215 spectrometer equipped with a thermoelectric cuvette-holder and the total fluorescence accessory. The buffer contained 2 mM phosphate, borate, citrate, 10 mM NaCl (pH 7.5). All final spectra were the average of at least four scans utilizing a signal averaging time of 2 s/λ. Signal averaging times for all melting experiments were 50 s/deg. C. Protein concentrations were determined using the method of Pace and co-workers.[34] B. Temperature-denaturation of bC_H3, mC_H3 and hC_H3 monitored by the far-UV CD signal at 217 nm. All melts increased the temperature at 2 deg. C intervals for CD and fluorescence sampling and utilized a 180 s equilibration period between data acquisitions. The melting temperature of bovine and murine C_H3 did not vary with concentration; therefore the far and near-UV CD curves and the fluorescence curves were all fit to a two-state, single-molecule unfolding model.[35] The stability of each variant was qualitatively ranked using T_m values. A theoretical $\Delta C_p°$, 1727 cal/mol K (1 cal = 4.184 J), was calculated based on an estimate of the change in accessible surface area (ΔASA) between the folded and unfolded states of the C_H3 dimer.[36,37]

far-UV CD measurements were performed using 10 μM protein and near-UV CD measurements using 30 μM protein (see Figure 2B for far-UV CD temperature-denaturation). Over this 60-fold range of concentrations, no change in the T_m was observed, as might be expected for a folded dimer to unfolded monomer two-state transition.[12] Buchner and co-workers, however, observed a concentration-dependent change in the guanidinium–HCl denaturation midpoint of the C_H3 domain from the murine monoclonal antibody MAK33 typical for a native dimer to unfolded monomer transition.[13,14] Thermal unfolding transitions of the dimeric Arc Repressor protein and the dimeric four-helix bundle protein, ROP, both display concentration-dependent T_m shifts.[15,16] The lack of a temperature shift between different concentrations of protein could be explained if the folding rate of the protein is not dominated by the rate of dimer association, but instead dominated by the *trans–cis* isomerization of P177(374). Another possibility could be that the dimer dissociates before reaching the temperature at which the protein unfolds. Due to the uncertainty in the model that should be applied for quantitative interpretation of the free energy of folding, we use the T_m value of bC_H3 as a standard to assess the relative stability changes induced by the designed mutants described below.

Our goal was to discover residue positions within bC_H3 not fully optimized for protein stability. We performed positional frequency and positional entropy analyses on our limited IgG data set (36 sequences from 19 mammalian species) utilizing a *Perl* program written in-house. A ClustalW alignment of the 36 sequences is provided in the Supplementary Material†.[17] All sequences were at least 95% unique (i.e. at least ten residue variations out of ~220 residues from any other sequence within the data set). Without the 5% cutoff, the dataset burgeons to over 100 sequences. The redundancy creates considerable bias and is therefore eliminated. The residue frequency, $p_i(r)$, for each position, i, in an individual sequence is simply the number of times that a particular residue-type (r = A,C,D…V,W,Y) is observed within the dataset divided by the total number of sequences. The positional entropy, $N(i)$, was calculated as a measure of every residue position's variability.[18,19] The positional entropy is a function of the Shannon informational entropy, $H(i)$:

$$N(i) = e^{H(i)}, \quad H(i) = -\sum_{r=A}^{Y} p_i(r)\ln(p_i(r)) \qquad (1)$$

Data from the residue frequency analysis were used to design mutants capable of optimizing bC_H3 stability. Mutations were selected based on four criteria. The first is that the bovine sequence's residue frequency at a given position is much smaller than the most common residue frequency, $mcp_i(r)$, (i.e. $p_i(r)/mcp_i(r) \leq 0.2$). Next, mutations must be conservative based on C_H3 structure analysis and residue type. Thus, no charge reversals

† http://www.ebi.ac.uk/clustalw

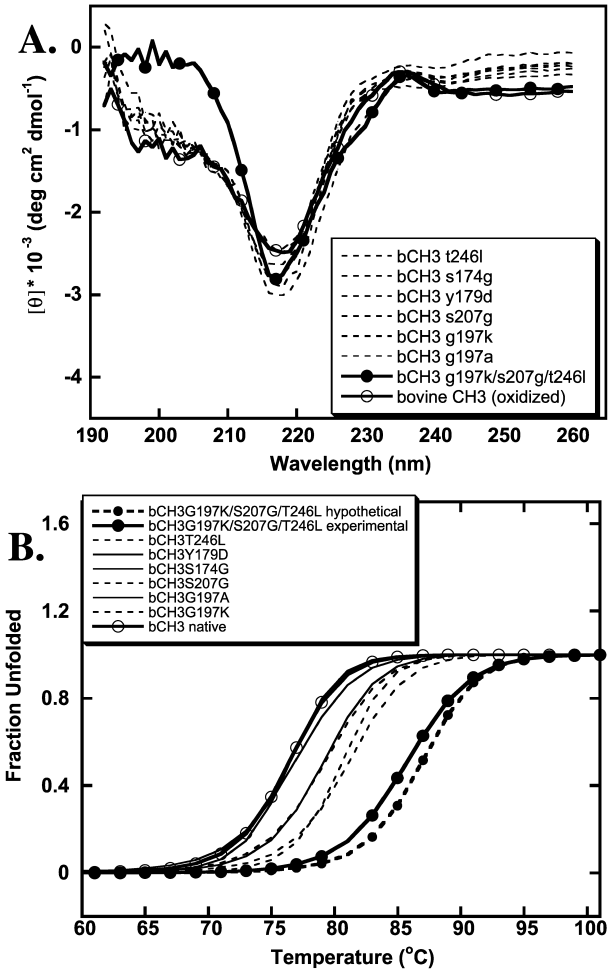

Figure 3. Structure and stability analyses of bC_H3 and the mutant constructs of the domain. A, The CD spectra of native and mutant bC_H3 constructs at 5 °C, pH 7.5. B, Plots of the unfolded fractions of the bC_H3 variants between 325 and 375 K.

were made and no large, buried hydrophobic amino acids were converted to charged residues or Gly. The third criterion is that there must be an absence of co-variation at this position with other residue positions throughout the sequence. Lastly, the residue to be mutated must not be conserved within the individual species IgG subclass (i.e. in bovine IgG2 or IgG3). This final criterion is related to the no-covariation criterion.

Five potentially non-ideal residue sites were identified when pitting the bovine C_H3 sequence against the Fc data set. These positions have been rendered onto the structure in Figure 1 and are indicated in the sequence alignment below the structure. Several C_H3 domains, including bovine C_H3, contain insertions or deletions resulting in an ambiguous numbering system when compared to the human IgG sequence. Therefore, we chose to number according to our Fc data set with residue G35 corresponding to residue G237, the first residue of the Fc structure of human IgG1.[20] Numbers in parentheses refer to the standard full-length human IgG residue numbering, while the numbers directly to the left of the residue letter refer to numbering originating from our Fc data set. The sites are S174(371), Y179(376), G197(392), S207(402) and T246(441). Each residue site is spatially distant from the four other sites, on the basis of the structure of hC_H3,[20] and is thus expected to have an independent and additive effect on the stability of the domain. The following point mutations were made based on the residue frequency analysis of bC_H3: S174G, Y179D, G197K, G197A (conservative mutation), S207G and T246L. Four of these five mutant positions are at residue sites of reasonable heterogeneity (i.e. the measured positional entropies are in the top 70% for all residue positions). Mutation at S207(402) was the only position that was highly conserved; only four of the 36 sequences have a residue other than Gly, with Ser being represented twice.

All mutant proteins display nearly identical CD spectra compared to the native protein (Figure 3A), indicating that no major structural change was induced by any of the point mutations. Each mutant domain exhibits a reproducible T_m, as judged by both near and far-UV CD (Figure 3B). Near and far-UV CD denaturations were performed on all the mutants, using the same concentrations that were used to study the wild-type protein (Table 1). G197K, S207G and T246L mutations had a stabilizing affect on bC_H3. Combination of these mutants into a single construct led to near-additive increases in the thermal stability of the domain. The experimentally determined

Table 1. Results of fitting the CD temperature melts of bCH3, the six single-mutant variants and a triple-mutant variant to a two-state unfolding model

Construct	T_m (K) fluor. Ex. λ_{280}	T_m (K) CD $[\theta]_{217}$	T_m (K) $[\theta]_{280}$	T_m (K) Avg.	ΔT_m (K)
bCH3	348.0 ± 0.3	349.9 ± 0.3	349.0 ± 0.3	348.8 ± 0.5	–
S174G	–	348.7	349.6	349.1	0.3
Y179D	–	349.3	349.9	349.6	0.8
G197K	–	354.2	354.2	354.2	5.4
G197A	–	351.2	352.2	351.7	2.9
S207G	–	352.6	353.7	353.2	4.4
T246L	–	352.2	352.3	352.3	3.5
G197K/S207G/T246L	–	–	358.8	358.8	10.0
*G197K/S207G/T246L	–	–	–	*361.1	*13.3

The asterisk (*) indicates a hypothetical triple mutant, assuming the individual mutant stabilizations are additive.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.