JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Rational Design of Single-Composition ABC Collagen Heterotrimers

Jorge A. Fallas,[†] Michael A. Lee,[‡,§] Abhishek A. Jalan,[†] and Jeffrey D. Hartgerink*[,†,‡]

Departments of [†]Chemistry and [‡]Bioengineering, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: Design of heterotrimeric ABC collagen triple helices is challenging due to the large number of competing species that may be formed. Given the required one amino acid stagger between adjacent peptide strands in this fold, a ternary mixture of peptides can form as many as 27 triple helices with unique composition or register. Previously we have demonstrated that electrostatic interactions can be used to bias the helix population toward a desired target. However, homotrimeric assemblies have always remained the most thermally stable species in solution and therefore comprised a significant component of the peptide mixture. In this work we incorporate complementary modifications to this triple-helical design strategy to destabilize an undesirable competing state while compensating for this destabilization in the desired ABC composition. The result of these modifications is a new ABC triple-helical system with high thermal stability and control over composition, as observed by NMR. An additional set of modifications, which exchanges aspartate for glutamate, results in an overall lowering of stability of the ABC triple helix yet shows further improvement in the system's specificity. This rationally designed system helps to elucidate the rules governing the self-assembly of synthetic collagen triple helices and sheds light on the biological mechanisms of collagen assembly.

The relationship between a protein's amino acid sequence and three-dimensional structure, while intensively studied for decades, has only been reliably solved for a few relatively simple protein folds. Foremost among these is the α-helical coiled coil, which can be reliably designed using hydrophobic and polar residue patterning as well as stabilizing side-chain contacts.[1−3] While collagen also has some straightforward design rules, such as its requirement for glycine every third amino acid and a high percentage of the cyclic imino acids proline and hydroxyproline, these design rules only aid in the preparation of homotrimeric helices.[4] Natural collagen, however, can be homotrimeric or heterotrimeric of the AAB or ABC variety depending on how many unique peptide strands are involved in its self-assembly. The design rules to control these heterotrimeric assemblies, both in nature and in synthetic analogues, are not well understood. Here we utilize a recently derived sequence-structure relationship between the positively charged amino acid lysine and the negatively charged amino acids aspartate and glutamate to rationally engineer specific salt-bridges into heterotrimeric assemblies that destabilize undesirable competing states while engaging in stabilizing pairwise interactions in the target state. This multistate approach to rational design results in the self-assembly of high-stability single-composition ABC-type collagen heterotrimers in which the designed interactions are satisfied.

The principles required to design sequences that fold into homotrimeric triple helices are straightforward. They follow natural collagen's primary structure, made up of Xxx-Yyy-Gly triplets, and amino acid composition, with a high content of imino acids. For example, the sequence $(POG)_{10}$ forms a highly stable homotrimeric helix, **O·O·O**, with a melting temperature of 67.5 °C.[5] (Here we use single-letter amino acid code and "O" for hydroxyl proline, while bold letters designate the peptides listed in Table 1.) This sequence utilizes the most common amino acids found in the Xxx and Yyy positions of natural collagens, proline and 4*R*-hydroxyproline. P and O contribute to the stability of the folded state by restricting the main-chain dihedrals to values close to those found in the folded structure.[6] Also, O is thought to participate in interchain water-mediated contacts that further stabilize the folded state.[7] The glycine residues pack tightly in the core of the helix enabling the canonical hydrogen-bonding network of this fold, which is perpendicular to the helical axis and goes from the amide nitrogen of glycine to the carbonyl of an amino acid in the Xxx position in an adjacent strand.[8] In order to fulfill this hydrogen-bonding pattern the chains assemble staggered by one amino acid differentiating a leading, middle, and lagging strand.

Table 1. Peptide Sequences and Abbreviations

abbreviation	sequence[a]
A	$YG(PKGPKGPKGDKGPKG)_2$
B	$YG(DOG)_{10}$
B1	$YG(EOG)_{10}$
C	$YG(POGPKGPOGPOGPOG)_2$
O	$(POG)_{10}$

[a]All peptides include a ^{15}N-labeled glycine at position 17 and are free amines at the N-terminus and amides at the C-terminus.

However, the design principles for obtaining sequences that self-assemble into heterotrimeric helices are not well understood. This protein fold lacks a well-defined hydrophobic core that can be used to bias the self-assembly toward a particular structure, as is the case in other protein structural motifs such as coiled coils.[1] With this constraint, only solvent exposed side-chain interactions can be used to direct the assembly toward a

Received: October 14, 2011
Published: December 22, 2011

   dx.doi.org/10.1021/ja209669u | *J. Am. Chem.Soc.* 2012, 134, 1430−1433

particular state. Furthermore, because of the one residue stagger seen in collagen, an AAB heterotrimer can potentially form three different triple helices, each one referred to as a different register of the same helical composition. Similarly, an ABC heterotrimer can populate six distinct registers, which can be thought of as cyclic permutations of the same structure making them hard to differentiate energetically. A total of 27 different triple helices are possible in a ternary mixture of peptides when including all compositions and registers, making this a challenging system for rational design. Our initial approach to generate ABC-type triple helices using electrostatic interactions[5,9,10] relied upon charge pairing between two chains, each with +10 or −10 charge, while using the neutral $(POG)_{10}$ sequence to complete the helix. This approach can be rationalized by analyzing the stabilizing effects of pairwise interactions in triple-helical proteins, where it has been shown that oppositely charged amino acids engage in the most stabilizing contacts in this protein fold.[11] Although successful at generating high-stability heterotrimers, this approach has the major drawback that the most thermally stable helix within the system is the **O·O·O** homotrimer. This results in a mixed composition ensemble where both homotrimeric and heterotrimeric states are populated.[12] A recent study that also uses electrostatics to impart specificity, and a computational methodology for sequence selection, generated a single-composition ABC heterotrimeric system,[13] as determined by circular dichroism (CD) melting studies. However, the presence of competing species is easily missed using CD alone.[12] Additional atomic resolution information would be desirable to corroborate the composition of the system and validate the pairwise interactions that give rise to the observed specificity.

In order to improve upon the specificity toward heterotrimers within our zwitterionic system, the stability of the **O·O·O** homotrimer needs to be reduced. Work by Brodsky *et al.* shows that the substitution of hydroxyproline by any other amino acid leads to a loss of thermal stability in the resulting triple helix.[14] This element of negative design is particularly attractive since homotrimeric assemblies will include 3 times the number of substitutions relative to their heterotrimeric counterparts. Peptide **C** (Table 1) is based on the $(POG)_{10}$ template but includes two substitutions in the Yyy position of the second and seventh triplets (O7K and O22K). This peptide, as well as all others in this study, includes an N-terminal tyrosine to facilitate accurate calculations of concentration. The resulting peptide forms a homotrimeric triple helix **C·C·C**, as evidenced by the sigmoidal transition observed in CD thermal unfolding experiments. Figure 1 shows the first derivative of the CD melting curve of **C·C·C**, with a transition temperature of 51 °C, which is 14 °C lower than that of the parent helix **O·O·O**. To aid the study of the molecular conformation of the peptide in solution, an ^{15}N-labeled glycine was included in the fifth triplet of the peptide (G17). The ^{1}H,^{15}N-heteronuclear single-quantum coherence (HSQC) spectrum of the peptide is shown in Figure 2A. The peak corresponding to the **C·C·C** homotrimer (G_H) has the expected chemical shift, similar to that of **O·O·O**. Two additional peaks are observed (G_M), and we have assigned them to monomeric peptide with different cis−trans isomerization states of the prolyl-peptide bonds surrounding the labeled glycine.[15]

To complement the negative design element we included a positive design component to widen the stability gap between the desired state and competing states. This was achieved by pairing the lysine residues in chain **C** with aspartate residues in an adjacent strand. Since the **A·B·C** register of the heterotrimer is the target state, two aspartate residues were placed in the Xxx position of the fourth and ninth triplets of the $(PKG)_{10}$ template sequence (P12D and P27D) to make peptide **A**. This sequence arrangement (K at position *n* in the lagging strand and D at position *n*+5 in the leading strand) was chosen on the basis of modeling and previous NMR studies,[12] which indicate that this places the charged side-chain moieties in an ideal position to engage in interstrand ionic hydrogen bonds. To utilize this relationship we complete our heterotrimeric system with peptide **B**, which follows the $(DOG)_{10}$ template.

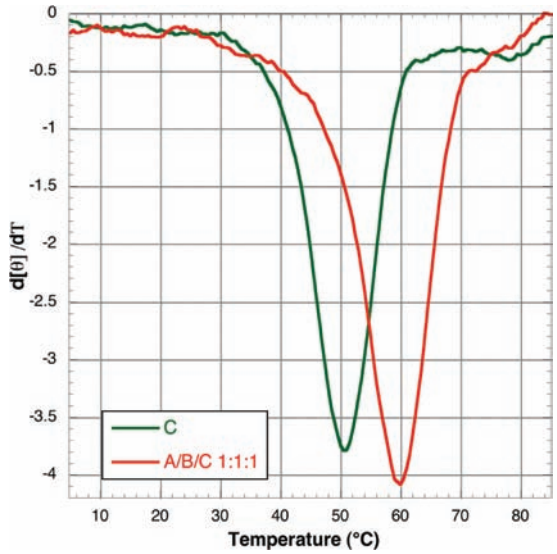

Figure 1. First derivative of the CD thermal unfolding curve for the **C·C·C** homotrimer and **A·B·C** heterotrimer.

A 1:1:1 mixture of peptides **A**, **B**, and **C** forms a highly stable ABC-type heterotrimer as evidenced by CD thermal unfolding studies. The first derivative of the CD melting curve of an annealed mixture of the three peptides is shown in Figure 1. Neither peptide **A** nor **B** forms a homotrimer, and the melting temperature of the ternary mixture, at 60 °C, is approximately 8 °C higher than that of the **C·C·C** homotrimer, as well as that of any of the binary mixtures (see Supporting Information (SI)), making the ABC heterotrimer the most thermally stable species in the redesigned system.

Despite the apparent success observed by CD (a clean single transition in the ternary mixture), we are interested in studying the molecular conformation of the mixture. Particularly, we would like to verify that the difference in stability between the ABC heterotrimer and the homotrimer is sufficient to preclude self-assembly of the latter. The ^{1}H,^{15}N-HSQC spectrum of the mixture (Figure 2B) shows the three peaks expected from the heterotrimer, with chemical shifts comparable to those of the template sequences.[12] Importantly, it also lacks the peak corresponding the **C·C·C** homotrimer. Thus, within the experimental limits of CD and NMR, we only observe single-composition ABC heterotrimers, validating our design protocol. However, in addition to the peaks corresponding to the main register of the ABC system, an additional cross peak is observed in the spectrum (G'_B), which we assigned to a competing register of the ABC helix. This assignment is based on a combination of nuclear Overhauser effect spectroscopy (NOESY) and ^{1}H,^{15}N-HSQC experiments (available in the SI) at 45 °C to avoid ambiguity about composition since the highest stability AAB heterotrimer within this system has a

dx.doi.org/10.1021/ja209669u | *J. Am. Chem.Soc.* 2012, 134, 1430−1433

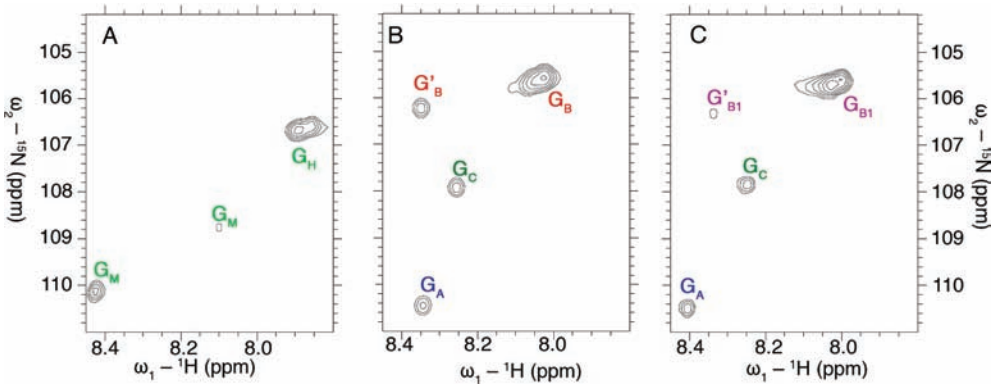

Figure 2. ^{1}H,^{15}N-HSQC spectra of the (A) **C**·**C**·**C**, (B) **A**·**B**·**C**, and (C) **A**·**B1**·**C** triple helices at 25 °C. Peaks are from labeled glycine residues of the indicated peptide strand (**A**, **B**, **B1**, or **C**). H indicates homotrimer while M indicates monomer of **C**·**C**·**C**.

melting temperature of 46 °C. The peaks corresponding to the other two chains of the competing ABC register were not identified due to chemical shift overlap with the peaks arising from the main register.

Our structure-based design strategy was successful in generating a single-composition ABC heterotrimer, but NMR revealed that it resulted in a mixed system with respect to chain register. Because we are ultimately interested in generating self-assembled single-composition, single-register heterotrimers, the effect of replacing aspartate with glutamate in the Xxx position of the B chain on the specificity of the system was explored (peptide **B1**, Table 1). Previously, we observed that replacing aspartate with glutamate in heterotrimeric triple-helical systems decreases the melting temperature of heterotrimers,[9] which could lead toward improved specificity in a redesigned system. CD demonstrates that a 1:1:1 mixture of peptides **A**, **B1**, and **C** also forms an ABC heterotrimer albeit with reduced thermal stability (52 °C, similar to that of the homotrimer of **C**, see SI). Yet, the ^{1}H,^{15}N-HSQC spectrum of the system (Figure 2C) lacks the peak corresponding to the **C**·**C**·**C** homotrimer, confirming that this system is also composed solely of ABC heterotrimers (within the limit of detection by NMR) despite the similar melting temperature of the **C**·**C**·**C** homotrimer. The lack of **C**·**C**·**C** homotrimer can be rationalized by considering the chemical potential of the solution and the relative stability of the different available states. The mixture will seek to minimize the chemical potential to reach equilibrium, and this is achieved by populating only heterotrimeric helices, as the self-assembly of **C**·**C**·**C** would force the two remaining peptide chains to fold into unstable AAB heterotrimers. The pattern observed in this spectrum is very similar to that of the mixture containing **B**, with the main difference being that the peak corresponding to the secondary is register is significantly weaker. We interpret this as a reduction in the relative population of this state. Thus, by including both K/E and K/D charge pairs, we have improved upon the specificity toward a particular register with the caveat that the overall stability of the system is decreased. This trade-off between the overall stability in a designed protein system and the specificity toward a particular state is not unusual[3,16] but has only been recently been explored for triple-helical systems.[13]

The CD melting studies corroborate the success of our negative design component, but in order to confirm the success of our multistate strategy, the structure-based positive design element needs to be validated using solution NMR. Because of the symmetry of the triple helix, only a single set of chemical shifts is expected for each of the newly incorporated charged residues,[17] facilitating the analysis. Figure 3A shows two sections of the NOESY spectrum of a 3 mM 1:1:1 mixture of peptides **A**, **B1**, and **C** at 45 °C. The cross peak at 8.61 and 3.02 ppm arises from the asparate amide proton in chain **A** and the lysine ε-methylene in chain **C**, which shows a single chemical

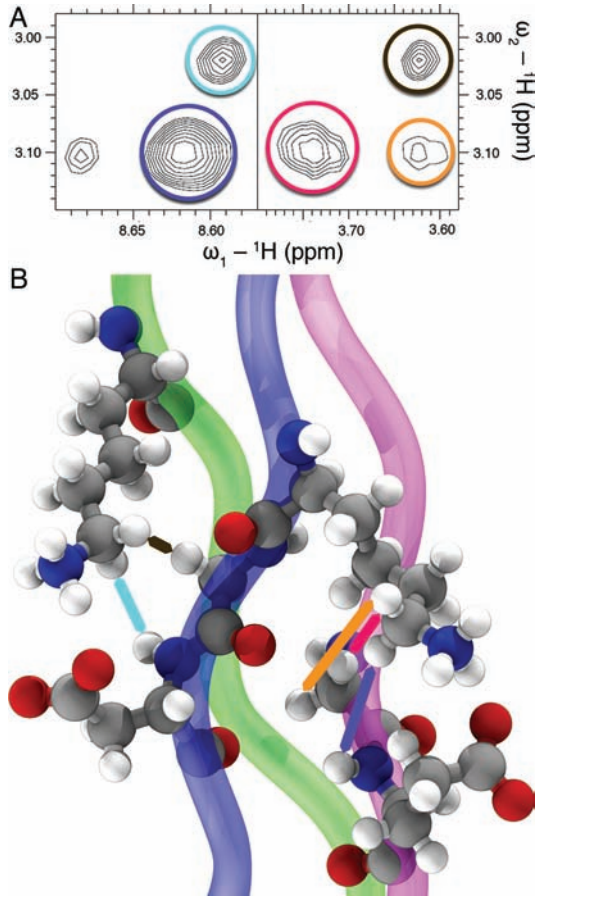

Figure 3. ^{1}H,^{1}H-NOESY spectrum and model. (A) NOESY spectrum of the **A**·**B1**·**C** heterotrimer highlighting interstrand interactions involving the lysine side chains and aspartate/glutamate and glycine backbone atoms. (B) *In silico* molecular model (chain **A** is blue, **B1** is purple, and **C** green) showing the protons that give rise to the resonance in (A); circles in (A) correspond to lines of the same color in (B).

dx.doi.org/10.1021/ja209669u | *J. Am. Chem.Soc.* 2012, 134, 1430–1433

shift for both diastereotopic protons. This cross peak is characteristic of salt bridges in triple-helical peptides[12,18] and therefore validates our positive design strategy. Furthermore, peaks at 3.68 and 3.02 ppm, arising from lysine ϵ-protons in chain **C** and the glycine α-protons preceding the aspartate in chain **A**, are also observed. The corresponding resonances between chains **A** and **B1** are also highlighted in the figure. These peaks are important, not only because they validate our positive design strategy but also because they serve to unambiguously determine the register of the triple helix as **A·B1·C**. Given the relative position of the charged amino acids in peptides **A** and **C**, the only register in which those residues can come close enough to one another to generate NOEs is the target state, the **A·B1·C** register of the heterotrimeric helix. Figure 3B shows a triple-helical structure generated using the flexible backbone modeling capabilities[19] of the Rosetta macromolecular modeling suite[20] in addition to biasing the charged side chains' conformation using constraints derived from the observed NOEs. This highlights the atoms that give rise to the resonances mentioned above. A comparable analysis for the remaining system (available in the SI) leads toward the same conclusion: the main component in the mixture corresponds to the target state, the **A·B·C** register of the heterotrimeric helix.

The synthesis of a single-register self-assembling ABC triple helix of high stability is a major challenge for *de novo* protein design. Such a system is desirable because it can be used as a scaffold in host−guest peptides to study the structure, biochemistry, stability, and multistate self-assembly of heterotrimeric collagenous proteins, mirroring what has been done with homotrimeric triple helices, which has been pivotal in our understanding of the most abundant protein family in the human body.[21] This study successfully improves upon previous efforts by generating a high-stability single-composition ABC heterotrimeric system utilizing a rational multistate design strategy that exploits novel sequence−structure relationships in triple-helical proteins. Despite populating more than one register of the desired heterotrimer, NMR studies on the system are able to confirm that the major component in the mixture corresponds to the target state and that the stabilizing pairwise interactions that were included on the basis of structural modeling are satisfied. Furthermore, we are able to reduce the relative population of alternative registers by modifying the amino acid composition of one the peptide chains within the system. In order for future systems to improve upon selectivity toward a particular register, the stability of competing states of the same composition as the target state needs to be included in the design protocol.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Experimental details and characterization data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

jdh@rice.edu

Present Address

[§]Department of Chemical Engineering, Princeton University, Princeton, NJ

■ ACKNOWLEDGMENTS

This work was funded in part by NSF CAREER Award (DMR-0645474), the Robert A. Welch Foundation (Grant No. C1557), and the Norman Hackerman Advanced Research Program of Texas.

■ REFERENCES

(1) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645−11651.
(2) Nautiyal, S.; Alber, T. *Protein Sci.* **1999**, *8*, 84−90. Havranek, J. J.; Harbury, P. B. *Nat. Struct. Biol.* **2003**, *10*, 45−52. Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2006**, *7*, 691−695.
(3) Grigoryan, G.; Reinke, A. W.; Keating, A. E. *Nature* **2009**, *458*, 859−864.
(4) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929−958.
(5) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683−2690.
(6) Shoulders, M. D.; Satyshur, K. A.; Forest, K. T.; Raines, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 559−564.
(7) Bella, J.; Brodsky, B.; Berman, H. M. *Structure* **1995**, *3*, 893−906.
(8) Fallas, J. A.; O'Leary, L. E.; Hartgerink, J. D. *Chem. Soc. Rev.* **2010**, *39*, 3510−3527.
(9) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034−15041.
(10) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7509−7515.
(11) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2005**, *44*, 1414−1422.
(12) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851−26859.
(13) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *133*, 15260−15263.
(14) Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. *J. Biol. Chem.* **2005**, *280*, 19343−19349.
(15) Buevich, A. V.; Dai, Q. H.; Liu, X. Y.; Brodsky, B.; Baum, J. *Biochemistry* **2000**, *39*, 4299−4308.
(16) Bolon, D. N.; Grant, R. A.; Baker, T. A.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12724−12729.
(17) Li, M.-H.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377−7387.
(18) O'Leary, L. E.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432−5443.
(19) Humphris, E. L.; Kortemme, T. *Structure* **2008**, *16*, 1777−1788.
(20) Leaver-Fay, A. *Methods Enzymol.* **2011**, *487*, 545−574.
(21) Kramer, R. Z.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. *Nat. Struct. Biol.* **1999**, *6*, 454−457. Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960−14967. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *J. Mol. Biol.* **2002**, *316*, 385−394. Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell* **2000**, *101*, 47−56. Farndale, R. W.; Lisman, T.; Bihan, D.; Hamaia, S.; Smerling, C. S.; Pugh, N; Konitsiotis, A.; Leitinger, B.; de, G., PG; Jarvis, G. E.; Raynal, N. *Biochem. Soc. Trans.* **2008**, *36*, 241−250. Carafoli, F.; Bihan, D.; Stathopoulos, S.; Konitsiotis, A. D.; Kvansakul, M.; Farndale, R. W.; Leitinger, B.; Hohenester, E. *Structure* **2009**, *17*, 1573−1581. Kar, K.; Ibrar, S.; Nanda, V.; Getz, T. M.; Kunapuli, S. P.; Brodsky, B. *Biochemistry* **2009**, *48*, 7959−7968.