8-2022

Determining the Full-Length Structure of Collagenase H using Small-Angle X-ray Scattering

Josie Carson
University of Arkansas, Fayetteville

Determining the Full-Length Structure of Collagenase H using Small-Angle X-ray Scattering

An Honors Thesis submitted in partial fulfillment of the requirements for Honors Studies in Biochemistry

By

Josie Carson

Spring 2022

Department of Biochemistry

J. William Fulbright College of Arts and Sciences

The University of Arkansas

Acknowledgements

I would like to extend a special thanks to my thesis director, Dr. Josh Sakon, and his graduate assistant, Cody Brazel, for sharing their knowledge and for their steadfast commitment toward my success throughout my undergraduate research. I also wish to express my gratitude to my thesis committee, Dr. Frank Millett and Dr. Kirsten Erickson, for their time, support, and interest in my project. Finally, I would like to thank the University of Arkansas Honors College, the Sturgis family, and University of Arkansas's Chemistry Department for giving me the opportunity to pursue my academic aspirations.

Table of Contents

Abstract

Known to cause gas gangrene, *Hathewaya histolytica* secretes two sister collagenases, collagenase G (Col G) and collagenase H (Col H), to degrade the triple helical structure of collagen to further infection in a host. Individual domains of Col H have been crystalized in previous studies[1, 2, 3], but methods in x-ray crystallization of full-length Col H have been unsuccessful. Using Small Angle X-Ray Scattering (SAXS) data, atomistic modeling was used to generate multiple conformations of Col H while accounting for flexibility between domains. Full-length Col H was found to adopt a two-state conformational model exhibiting a majority compact and a minority elongated form regardless of calcium concentration. This suggests Col H may become more flexible in lower calcium concentrations but does not elongate as previously suspected. Determination of full-length Col H could significantly impact drug delivery design and therapeutic agents concerning procedures such as skin debridement and pancreatic islet extraction.[4, 5, 6, 7]

1. Introduction

1.1 Collagenases

Collagen, the most plentiful protein found in animals[8], is composed of a triple helical motif that commonly contains three, parallel polypeptide chains. Tropocollagen utilizes this triple helical structure to resist degradation by proteases as the scissile peptide bonds are protected. Because of this, bacteria have evolved to produce collagenases that degrade collagen fibrils made of tightly packed tropocollagen molecules.

Hathewaya histolytica, a bacterium known to cause gas gangrene, secretes two sister collagenases, collagenase G and collagenase H to further infection in a host. Col H consists of four domains: a catalytic module (S1), and a three-part binding segment consisting of two domains and a collagen binding domain (S2A, S2B, and S3).[1] The S2A and S2B domains resemble polycystic kidney disease-like domains (PKD).[9] The full-length structure of Col H has yet to be determined.

Previous literature suggests a strong correlation between protein flexibility and structure with calcium concentration.[10] Bacterial collagenases require calcium to bind collagen and reach full catalytic activity.[11, 12, 3] As the calcium concentration increases, Col H adopts a more contoured, rigid, and functional shape. When the calcium concentration decreases, the protein may become more linear and flexible.[2, 10-12] This elongation at lower calcium concentrations could potentially assist in the secretion of Col H into the extracellular matrix.

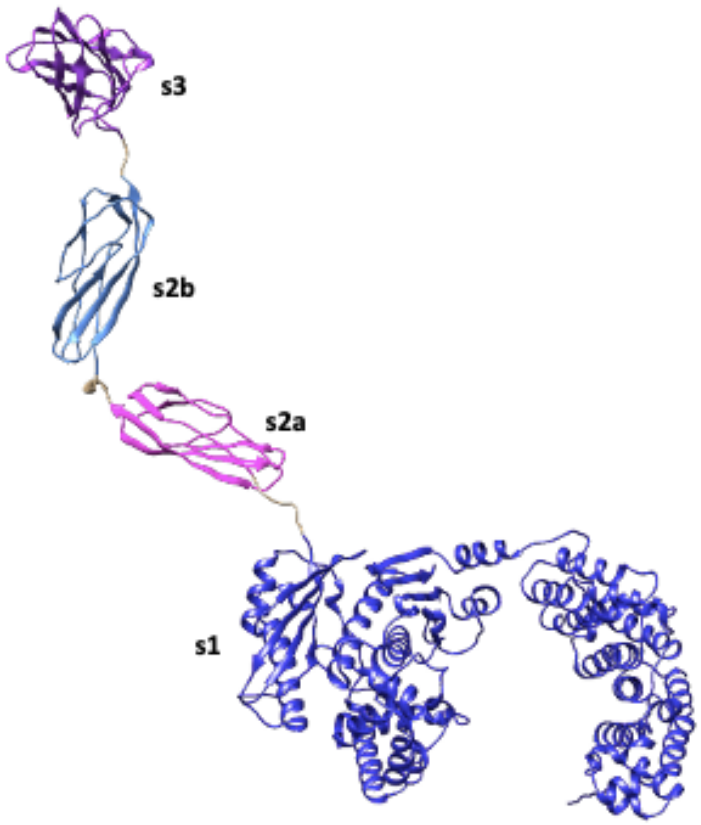

Figure 1: Domain arrangement of Collagenase H.

1.2 Small Angle X-ray Scattering

The individual domains of Col H have been crystalized in previous studies[1,2, 3], but methods in x-ray crystallization of full-length Col H have been unsuccessful. Small Angle X-ray Scattering (SAXS), a Low-Resolution Technique, allows an estimation of the shape of a molecule in solution without crystallization. SAXS has been used to observe the structure of Col H related to varied calcium concentrations. Preliminary data, generated using FOXS to compute single state models of Col H in various calcium concentrations, produced high global fit values that showed no correlation between theoretical modeling and SAXS experimental curves. Conversely, MultiFoXS utilizes a population-weighted computation, beginning from a single input structure, by fitting to

the SAXS scattering curve of the protein. Multi-state modeling, using MultiFoXS, can result in significantly improved global fit values.[13] The combination of SAXS and atomistic modeling can generate multiple conformations of Col H while accounting for its flexibility between domains. The method by which *Hathewaya histolytica* secretes Col H, as well as Col H's binding mechanism, processivity along collagen fibers, and possible synergy with collagenase G, are still unknown. Determination of the full-length structure of Col H will provide insight into these mechanisms.

1.3 Significant Impacts

Collagenases are currently used in a wide variety of therapeutic agents. For the last 50 years, collagenase H and G have been used in an ointment, SANTYL, that debrides dead tissue post trauma, allowing fresh skin to grow and accelerating the healing process.[6] Injections of Col G and Col H can also be used to treat the excessive growth of connective tissue found in Dupuytren's contracture.[4] Collagenases break down the extra collagen and bring temporary relief from the stiffness and constrictions experienced by affected individuals. Pancreatic islet injections and isolation transplantations employing the sister collagenases can be used to treat type 1 diabetes.[7] The islets are derived from the patient but are found in deeper tissues which makes extraction difficult without the use of Col H and Col G. Col H and Col G can also be used to extract pluripotent stem cells from adipose tissue that can then be used to treat various injuries.[5, 14] However, as of now, cell isolation from collagen rich tissue—for example, from the lung, tendons, or bone—is not time efficient. Further characterization of Col H could optimize cell extraction and the degradation of difficult tissues.

Col H combined with fusion proteins can be used in systematic applications to treat osteoporosis and to prevent and treat alopecia.[15] Tumor treatments can also utilize Col H to help with drug penetration and to assist in tumor removal; however, after as soon as the third injection, researchers saw antibodies arise that were capable of neutralizing collagen.[16] Collagenases can also be used in tandem with growth factors; the collagenases help prevent growth from becoming malignant.[17] The noncatalytic segments can also aid in the attachment of drugs to the site of interest which consequently decreases the drug dosage needed. Unfortunately, Col H is an immunogenic enzyme leading to inflammation and reduced effectiveness. Elucidation of the full-length structure of Col H could help decrease dosage and side effects while increasing efficacy as a therapeutic agent.

2. Materials and Methods

SAXS Sample Preparation of full-length Col H: HBS-EGTA, HBS-Ca, and HBS—three different HEPES buffered saline (HBS) stock solutions—were prepared consisting of 10mM HEPES, 100mM NaCl, and 0.4mM EGTA. This solution was also made in five different concentrations of glycerol: 0%, 0.5%, 1%, 1.5%, and 2%. HBS-Ca is comprised of 10mM HEPES, 100mM NaCl, and 2.4mM $CaCl_2$. HBS consists of 10mM HEPES and 100mM NaCl. The web application MAXCHELATOR was used to mix the five solutions to a final volume of 100mL with differing calcium concentrations: pCa 3, 4, 5, 6 and 7. Buffer exchange of full-length Col H were performed using Centricon—centrifugal units with a MWCO of 10 or 30 kDa. For each pCa, each glycerol concentration was analyzed using Dynamic Light Scattering (DLS) and UV-Vis Spectrometry. Samples showing no aggregation were selected for data collection at the SIBYLS beamline.

SAXS Data Collection: Synchrotron SAXS data from solutions of Full-length Col H in 50 mM Hepes, 100 mM NaCl, 5 mM $CaCl_2$, and pH 7.5 were collected on the 12.3.1 (SIBYLS) beamline at the Advanced Light Source (ALS; Berkeley, CA, USA) using a Pilatus3 X 2M detector at a wavelength of $\lambda = 0.113$ nm (I(s) vs s, where $s = 4\pi\sin\theta/\lambda$ and 2θ is the scattering angle). Five Col H samples were prepared in differing calcium concentrations as pCa 3-7. Protein concentrations ranging between 1 and 5 mg/ml were measured at 10°C. 33 successive 0.300 second frames were collected. The data was normalized to the intensity of the transmitted beam and radially averaged; the scattering of the solvent-blank was subtracted, and the different curves were scaled for protein concentration. The low angle data collected at lower concentrations was extrapolated to infinite dilution and merged with the higher concentration data to yield the final composite scattering curve.

Scattering for samples with concentrations of 5 mg/ml, 3 mg/ml and 1 mg/ml were averaged. Each of the average scattering curves were used to extrapolate towards infinite dilution.

Primary Data Analysis, SCATTER: SAXS data was analyzed using SCATTER and the ATSAS Suite. Buffer subtracted data for each concentration was subjected to a series of manual optimizations in SCATTER. Guinier analysis was utilized to remove outlying residuals; the q*Rc limit range was found at the first plateau of the Rc plot; flexibility and volume analysis were used together to give a maximum volume estimate of the molecule; P(r) analysis was used to estimate the Dmax and calculated using GNOM from the ATSAS suite. Visualization of data was performed using UCSF Chimera V 1.14.

Multi-State Modelling and MultiFoXS: Refined SAXS profiles were submitted to MultiFoXS calculations which uses Rapid Random Trees (RRT) to derive and average multiple conformations of Col H. For MultiFoXS, the flexible residues were identified using hingeProt. Full-length structure was predicted from the primary sequence using AlphaFold multimer V 2.0[18] and subsequently used as the input structure for MultiFoXS calculations.

3. Results

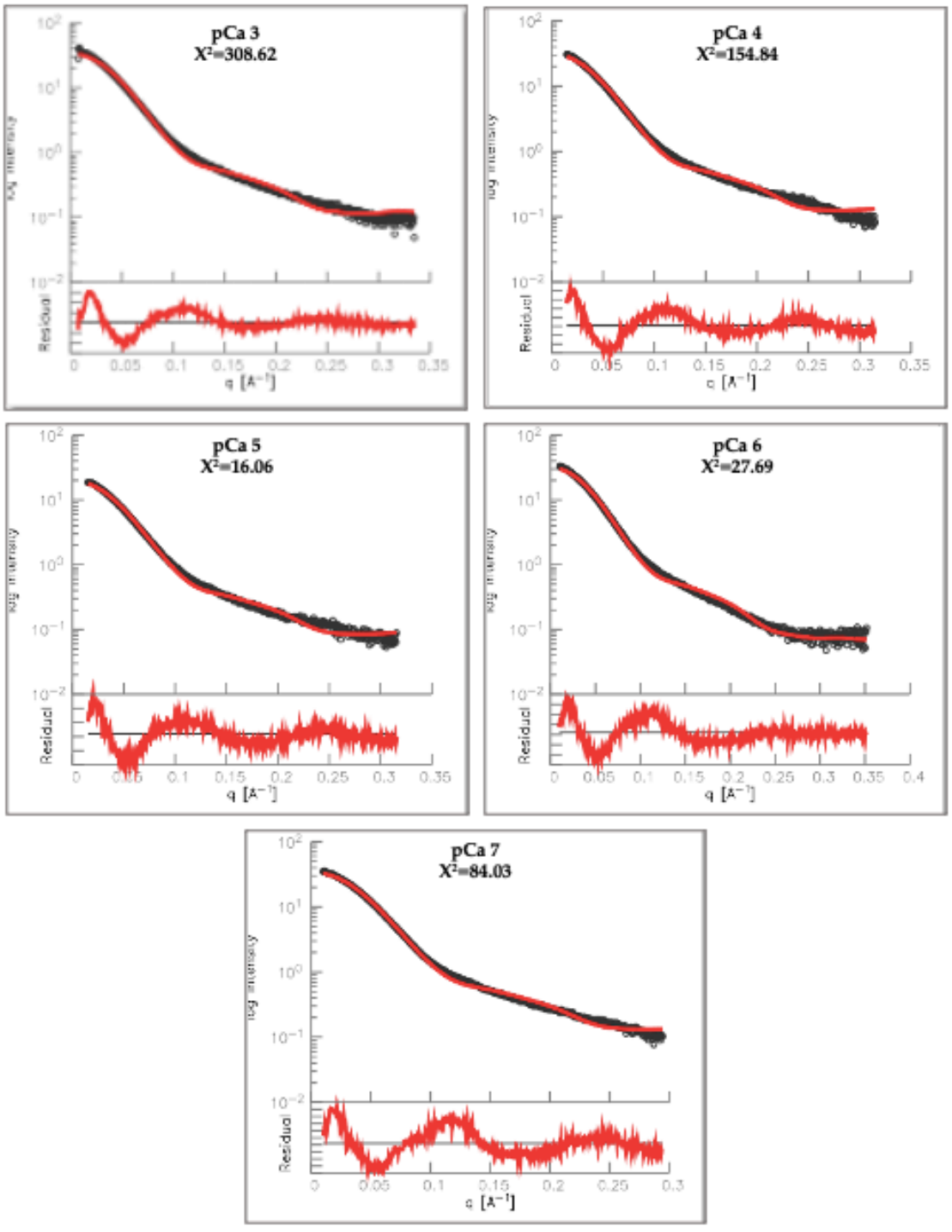

Figure 2: Single State Modelling of Full-Length Col H via FoXS. Significant deviation from theoretical (red) and experimental (black) scattering curve per pCa sample. Deviations suggest samples are not rigid systems.

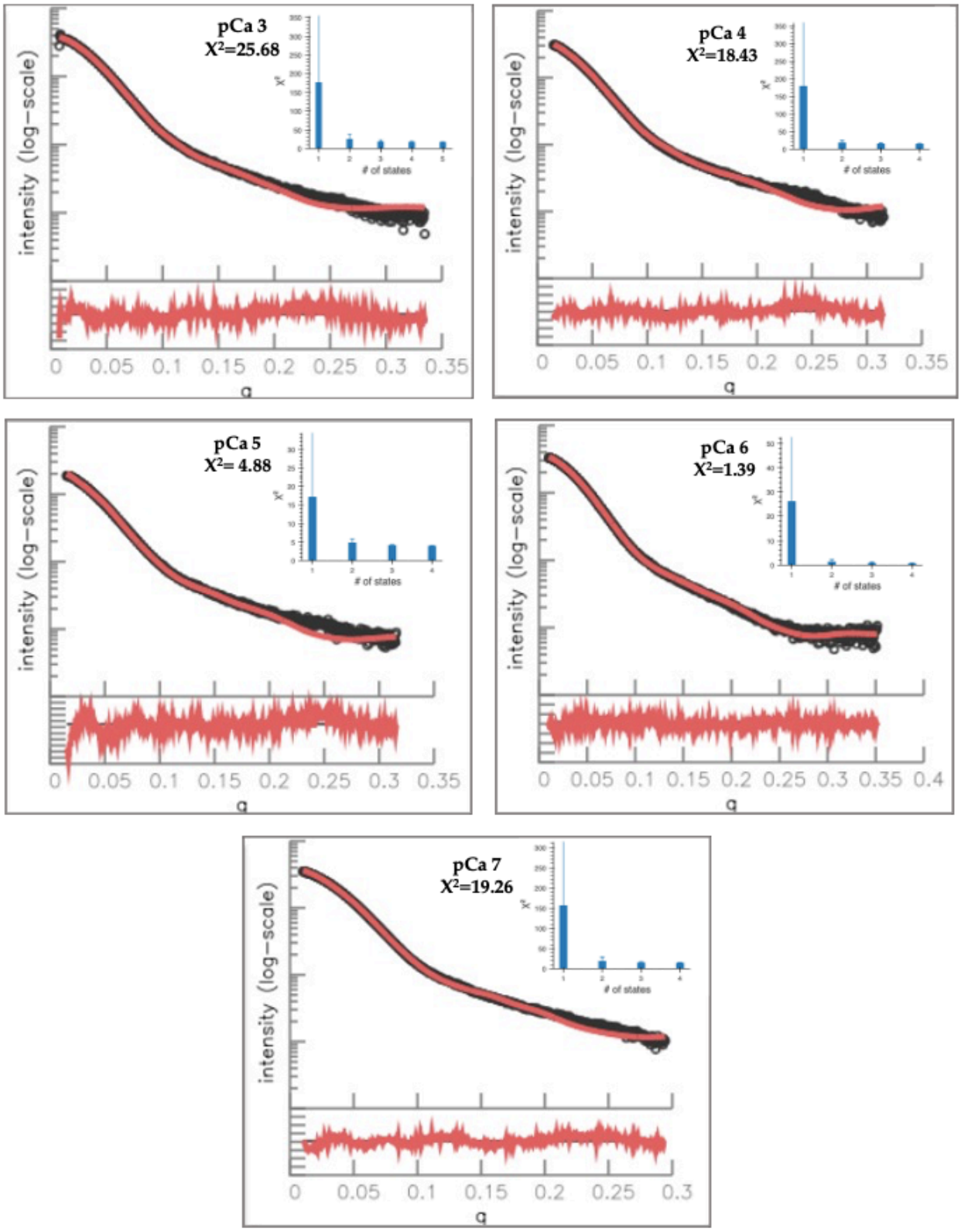

Figure 3: Multi-State modelling of Full-length Col H via MultiFoXS. Multi-State Ensembling reduces deviation between theoretical (red) and experimental scattering (black). Profiles represent the two-state conformational model for each pCa.

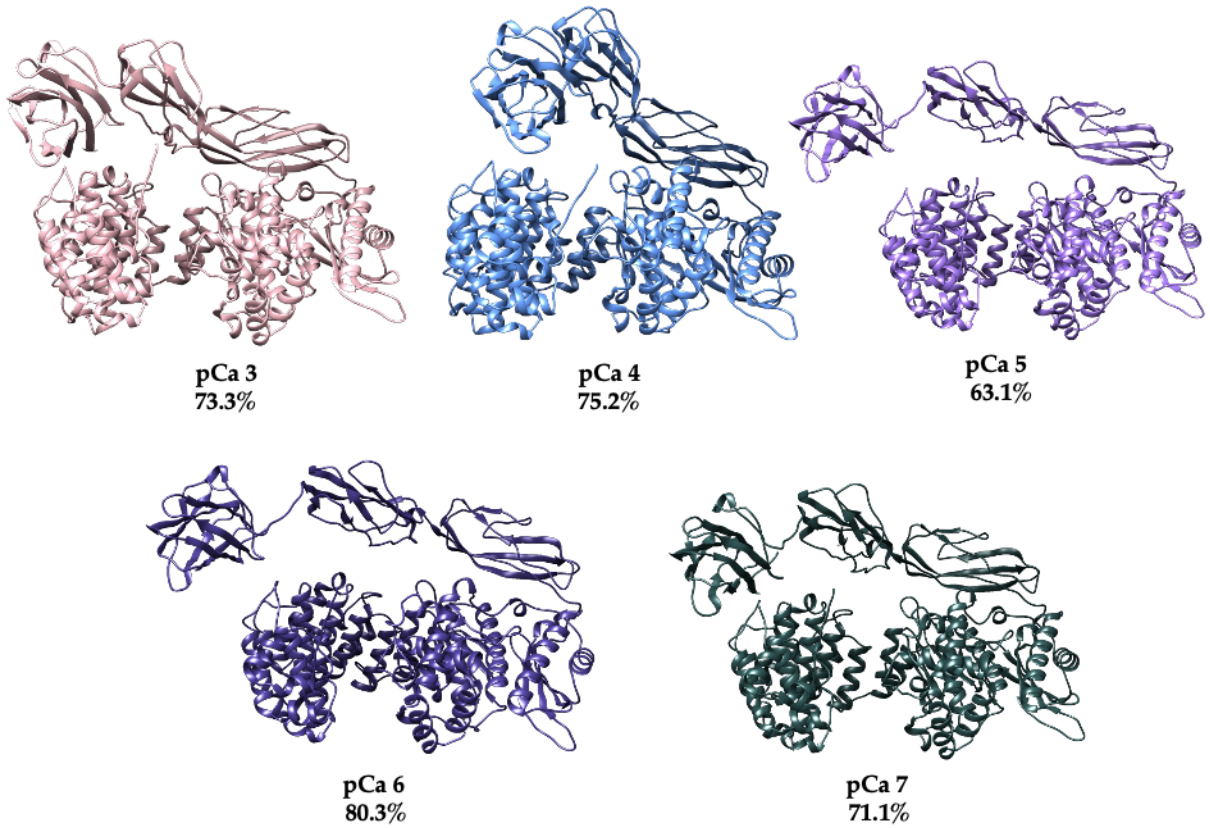

Figure 4: Highest Population-Weighted Ensemble Models via MultiFoXS.

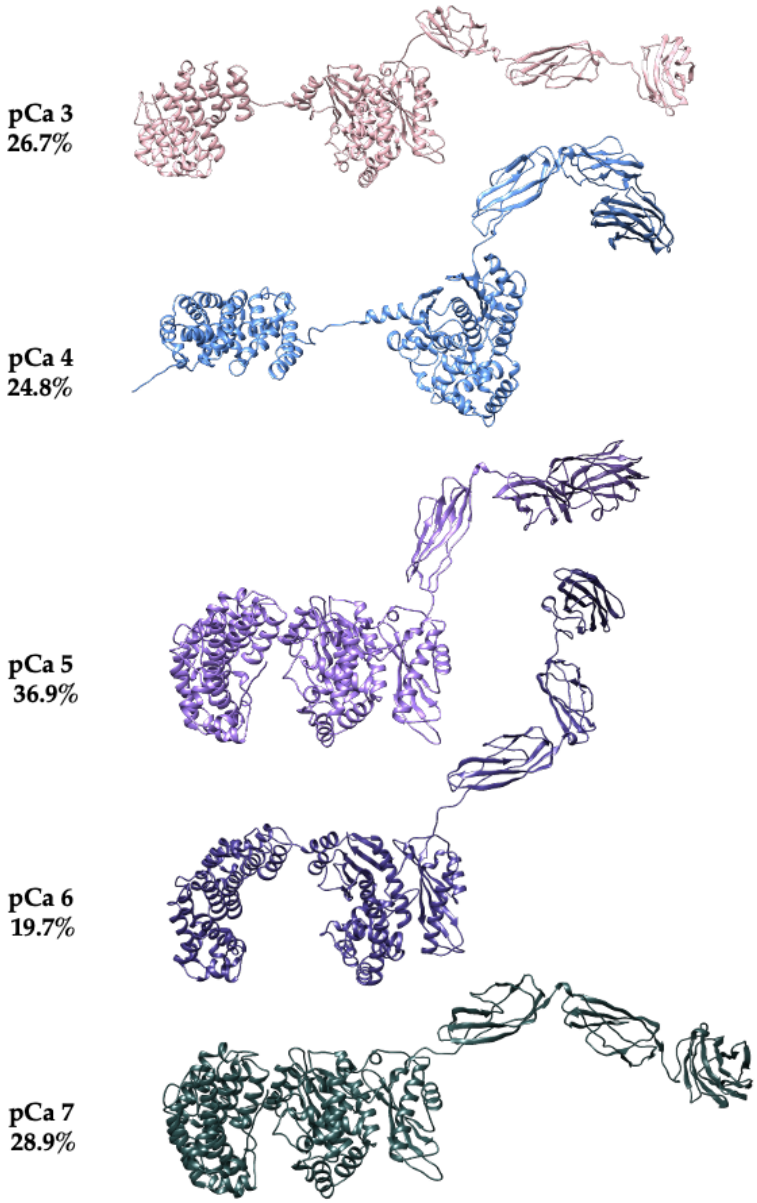

Figure 5: Second Highest Population-Weighted Ensemble via MultiFoXS.

$$\chi = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{I_{\exp}(q_i) - cI(q_i)}{\sigma(q_i)}\right)^2}$$

Equation 1: Global fit value expression.[13]

4. Conclusion

The models constructed using a single state of Col H resulted in high, triple digit global fit values (figure 2). Significantly high fit values, as seen in FoXS results, indicates that the one-state models may not approximate experimental scattering results. The presence of multiple conformations of Col H is evidenced by the reductions in global fit values of the MultiFoXS data (figure 3) versus the FoXS data (figure 2). The Col H scattering data coincides better with the multistate modelling, indicating that the protein may be a flexible system. Interestingly, each pCa concentration exhibited similar compact structure as the major conformation. These results are counter to the elongation that was expected to complement flexibility. Differentiating elongation from overall dynamics is difficult to perform in SAXS alone and may require additional experimental methods to solve. A trypsin limited proteolysis of Col H, at differing pCa concentrations, may elucidate the change in dynamics that SAXS results are suggesting.

MultiFoXS results describe a flexible protein—one that may change dynamics in response to varying calcium concentrations and provides a possible explanation as to the secretion of the protein. Potentially, lower calcium concentrations of cell cultures during expression will help facilitate the secretion, and subsequent yield, of Col H. In addition, higher calcium concentrations may be useful in preserving the integrity of purified proteins, thus ensuring activity and structure are kept constant over periods of long-term storage.

References

(1) Eckhard, U.; Schönauer, E.; Brandstetter, H. Structural Basis for Activity Regulation and Substrate Preference of Clostridial Collagenases G, H, and T. *Journal of Biological Chemistry* **2013**, *288* (28), 20184-20194. DOI: 10.1074/jbc.m112.448548 (acccessed 2021-05-12t15:46:05).

(2) Bauer, R.; Wilson, J. J.; Philominathan, S. T. L.; Davis, D.; Matsushita, O.; Sakon, J. Structural Comparison of ColH and ColG Collagen-Binding Domains from Clostridium histolyticum. *Journal of Bacteriology* **2013**, *195* (2), 318-327. DOI: 10.1128/jb.00010-12 (acccessed 2021-05-12t15:37:09).

(3) Wilson, J. J. A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. *The EMBO Journal* **2003**, *22* (8), 1743-1752. DOI: 10.1093/emboj/cdg172 (acccessed 2021-05-14t02:11:33).

(4) Gaston, R. G.; Larsen, S. E.; Pess, G. M.; Coleman, S.; Dean, B.; Cohen, B. M.; Kaufman, G. J.; Tursi, J. P.; Hurst, L. C. The Efficacy and Safety of Concurrent Collagenase Clostridium Histolyticum Injections for 2 Dupuytren Contractures in the Same Hand: A Prospective, Multicenter Study. *J Hand Surg Am* **2015**, *40* (10), 1963-1971. DOI: 10.1016/j.jhsa.2015.06.099.

(5) Carson, C. r. Analysis of the clinical safety of intralesinal injection of collagenase Clostridiumhistolyticum (CCH) for adults with Peyronie's disease (PD). Sadeghi-Nejad, H., Ed.; BJU Int.: BJU Int., 2015.

(6) Shi, L.; Carson, D. Collagenase Santyl ointment: a selective agent for wound debridement. *J Wound Ostomy Continence Nurs* **2009**, *36* (6 Suppl), S12-16. DOI: 10.1097/WON.0b013e3181bfdd1a.

(7) Li, D.; Peng, S.-Y.; Zhang, Z.-W.; Feng, R.-C.; Li, L.; Liang, J.; Tai, S.; Teng, C.-B. Complete disassociation of adult pancreas into viable single cells through cold trypsin-EDTA digestion. *Journal of Zhejiang University SCIENCE B* **2013**, *14* (7), 596-603. DOI: 10.1631/jzus.b1200226 (acccessed 2021-05-12t15:58:26).

(8) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annual Review of Biochemistry* **2009**, *78* (1), 929-958. DOI: 10.1146/annurev.biochem.77.032207.120833 (acccessed 2021-05-14t02:10:32).

(9) Bauer, R.; Janowska, K.; Taylor, K.; Jordan, B.; Gann, S.; Janowski, T.; Latimer, E. C.; Matsushita, O.; Sakon, J. Structures of three polycystic kidney disease-like domains fromClostridium histolyticumcollagenases ColG and ColH. *Acta Crystallographica Section D Biological Crystallography* **2015**, *71* (3), 565-577. DOI: 10.1107/s1399004714027722 (acccessed 2021-05-12t15:38:13).

(10) Ohbayashi, N.; Matsumoto, T.; Shima, H.; Goto, M.; Watanabe, K.; Yamano, A.; Katoh, Y.; Igarashi, K.; Yamagata, Y.; Murayama, K. Solution structure of clostridial collagenase H and its calcium-dependent global conformation change. *Biophys J* **2013**, *104* (7), 1538-1545. DOI: 10.1016/j.bpj.2013.02.022.

(11) Caviness, P.; Bauer, R.; Tanaka, K.; Janowska, K.; Roeser, J. R.; Harter, D.; Sanders, J.; Ruth, C.; Matsushita, O.; Sakon, J. Ca 2+ -induced orientation of tandem collagen binding domains from clostridial collagenase ColG permits two opposing functions of collagen fibril formation and retardation. *The FEBS Journal* **2018**, *285* (17), 3254-3269. DOI: 10.1111/febs.14611 (acccessed 2021-05-12t15:43:48).

(12) Philominathan, S. T. L.; Matsushita, O.; Jordan, J. B.; Sakon, J. H-1, C-13 and N-15 resonance assignments of Ca2+ bound collagen-binding domain derived from a clostridial collagenase. *Biomolecular Nmr Assignments* **2008**, *2* (2), 127-129. DOI: 10.1007/s12104-008-9102-z.

(13) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Research* **2016**, *44* (W1), W424-W429. DOI: 10.1093/nar/gkw389 (acccessed 2020-11-13t18:16:53).

(14) Riley, K. N.; Herman, I. M. Collagenase promotes the cellular responses to injury and wound healing in vivo. *J Burns Wounds* **2005**, *4*, e8.

(15) Katikaneni, R.; Ponnapakkam, T.; Suda, H.; Miyata, S.; Sakon, J.; Matsushita, O.; Gensure, R. C. Treatment for chemotherapy-induced alopecia in mice using parathyroid hormone agonists and antagonists linked to a collagen binding domain. *Int J Cancer* **2012**, *131* (5), E813-821. DOI: 10.1002/ijc.27379. Katikaneni, R.; Ponnapakkam, T.; Matsushita, O.; Sakon, J.; Gensure, R. Treatment and prevention of chemotherapy-induced alopecia with PTH-CBD, a collagen-targeted parathyroid hormone analog, in a non-depilated mouse model. *Anti-Cancer Drugs* **2014**, *25* (1), 30-38. DOI: 10.1097/CAD.0b013e3283650bff. Katikaneni, R.; Seymour, A. W.; Gulati, R.; Ponnapakkam, T.; Gensure, R. C. Therapy for Alopecia Areata in Mice by Stimulating the Hair Cycle with Parathyroid Hormone Agonists Linked to a Collagen-Binding Domain. *J Investig Dermatol Symp Proc* **2015**, *17* (2), 13-15. DOI: 10.1038/jidsymp.2015.32.

(16) Dolor, A.; Szoka, F. C. Digesting a Path Forward: The Utility of Collagenase Tumor Treatment for Improved Drug Delivery. *Molecular Pharmaceutics* **2018**, *15* (6), 2069-2083. DOI: 10.1021/acs.molpharmaceut.8b00319 (acccessed 2021-05-12t16:36:12).

(17) Ponnapakkam, T.; Katikaneni, R.; Sakon, J.; Stratford, R.; Gensure, R. C. Treating osteoporosis by targeting parathyroid hormone to bone. *Drug Discovery Today* **2014**, *19* (3), 204-208. DOI: 10.1016/j.drudis.2013.07.015.

(18) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **2022**, *50* (D1), D439-D444. DOI: 10.1093/nar/gkab1061.