Leather Related Collagen Modeling: The Challenges of Modeling Hierarchical Structures*

by

Lorenz Siggel[1]**, Rosa E. Bulo[2], Ferenc Molnar[1], Horst Weiss[1] and Tilman Taeger[1]

[1]*BASF Aktiengesellschaft*

Ludwigshafen, Germany

[2]*Department of Chemistry, Free University of Amsterdam,*

The Netherlands

Collagen is the most abundant protein in mammals. It covers the body in the form of skin and is crucial in virtually all connective tissues. Disruption of collagen biosynthesis from vitamin-C deficiency leads to scurvy, which decimated ships crews prior to the 18th century on long sea voyages where fresh fruits and vegetables were not carried. Healthy collagen is essential for a healthy life, yet this is often not appreciated.

Collagen in the form of animal hides has been an important raw material for man since the earliest days. Skins protected the body and feet against the elements. It is not known when early man first discovered tanning as a method of preserving and altering the properties of animal hides but it certainly goes back many thousands of years. The discovery of Oetzi "the iceman" in 1991 in Oetztal, Italy was a landmark discovery. Oetzi died of wounds received in ca. 3200 B.C. high in the Alps and was preserved in "deep freeze" for 5300 years along with his complete clothing and equipment. Among his leather goods were "composite shoes" made of a combination of leather outers (bearskin, calf and deer), plant fiber and straw inners for protection against the cold and rocks. It is clear that "primitive man" man was anything but primitive and maximized the performance of the available materials with their technological know-how in the same way that modern man still does, such as leather technology.

Tanning is the stabilization of leather against water, swelling and rot.[1] There are several ways to tan hides but of greatest significance today by far is chrome tanning which uses salts of Chromium III and was discovered in the mid 19th century. Alternatively there is vegetable tanning which is composed of large polyphenolic molecules found in the bark and roots of certain plants and has been known for hundreds if not thousands of years. Oetzis' leather articles were tanned, demonstrating a high level of sophistication for his culture.

"Leather" is a 40 billion dollar industry and leather chemicals have sales of roughly 3 billion dollars p.a. In order to improve and optimize the products for the modern market place ongoing research and development is crucial. Research over the past 100 years has brought great strides forward in the understanding of all aspects of leather production. However, there is still a lack of detailed understanding at the molecular level of tanning and the interaction of leather processing chemical with the leather, i.e. collagen molecules. Such an understanding could greatly contribute to a more rational design and development of chemicals for all aspects of leather chemistry. One goal could be the replacement of chromium salts with environmentally friendlier molecules while retaining their performance advantages.

Leather is an incredible material with performance attributes that rival modern synthetic materials: and this stretching back into antiquity! Leather is tough, abrasion resistant, and waterproof yet breathes, etc. It owes these attributes to its' hierarchical structure. By hierarchical we mean that substructure motifs are combined to form larger structures with different physical and mechanical properties. For example: tropo-collagen strands coalesce to micro fibrils and these to collagen fibers and these in turn to leather. The bulk structure of leather and the collagen fibers has been well studied by optical and electron microscopy (Figure 2) down to the nanometer scale. Also the interaction of many leather chemicals and finishing products can be studied and optimized at this level of detail, e.g. surface finishes, etc.. In order to gain an understanding of the structure of collagen at the molecular level several techniques have played crucial roles: the aforementioned high resolution electron microscopy (TEM, SEM), atomic force microscopy (AFM), x-ray crystallography, protein sequencing and theoretical methods.

Type-1 collagen consists of linear triple helix strands in a "coiled-coil" geometry analogous to a three stranded rope with the three helically coiled amino acid chains (two identical alpha-1 chains and an alpha-2 chain). This motif was first elucidated in the pioneering work of R.N. Ramachandran in the 1950's.[2] Type-1 collagen is ca 1000 amino acids long and consists of GLY-X-Y triplets, i.e. every third amino acid is glycine. This is important so that the repeating helix structure can exist. The other two amino acids (X and Y) in the triplets consist of ca.10% proline, ca.10% hydroxyproline, ca.10% alanine and the rest a distribution of the other naturally

* The Editor offers this Heidemann manuscript because it is expected to be of interest to our readers. This review, presented by Dr. Siggel at the XXIX Congress/103rd ALCA Annual Meeting on June 21, 2007, includes some figures that were first published in *Macromolecular Bioscience* **2007**, 7, 234-24. These figures are reproduced with permission from Wiley-VCH Verlag GmbH & Co.

** Corresponding author - Lorenz.siggel@basf.com, BASF AG, Dept. GVC/C, D67056 Ludwigshafen, Germany

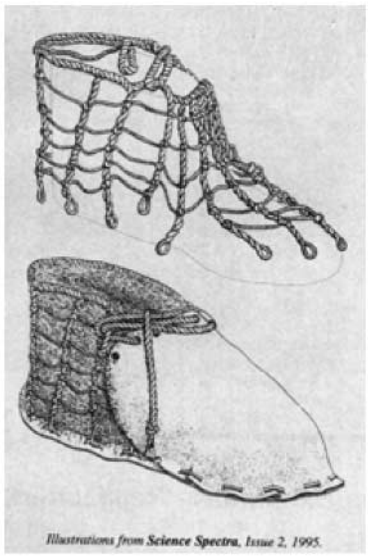

Figure 1: Oetzis shoes: leather is partially removed to illustrate the composite construction.

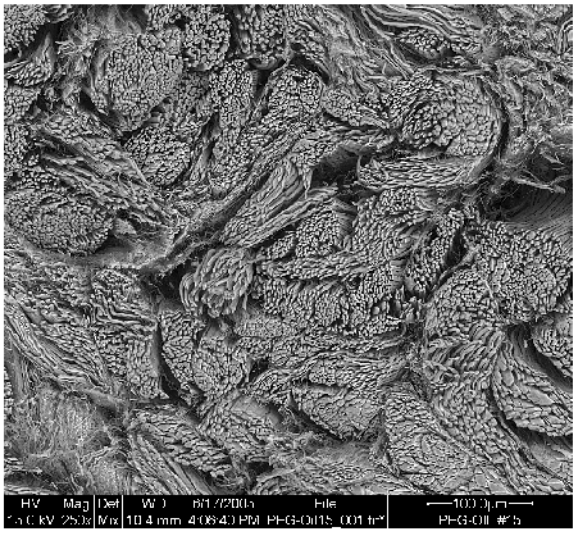

Figure 2: Cross-section of leather treated with PEG and silicone oil (courtesy of USDA)

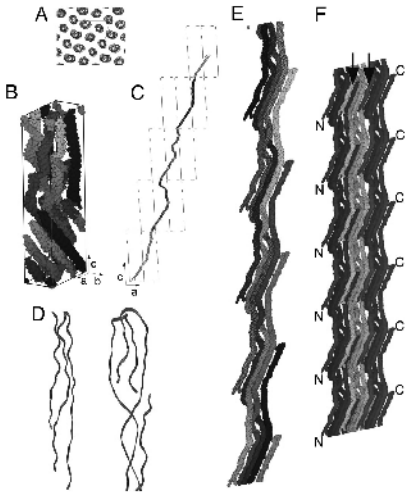

Figure 3: Type-1 collagen crystal structure. Orgel, Joseph P. R. O. et al. (2006) Proc. Natl. Acad. Sci. USA 103, 9001-9005

occurring amino acids. The end regions of the collagen molecule consist of short sections of less ordered telopeptides. The role of the hydroxyproline (formed from proline and the co-factor ascorbic acid) is to stabilize the fibrils via interstrand hydrogen bonding interactions rather than intrafibril stabilization. The proline provides the chain with stiffness but not interstrand stability. The tropocollagen molecules are offset to their neighbors along the axis of the fibril. This is seen as regular bands in the electron micrographs. In 2006 Orgel, et al.[3] published a crystal structure of the complete collagen type-1 fibril, including the telopeptide region. The structure shows an interesting packing pattern of the tropocollagen along the length of the fiber and a hexagonal packing in cross-section (Figure 3). The Orgel stucture confirms the packing pattern first reported by Wess et al. in 1998.[4] The primary amino acid sequences of the different types of collagen are known and deposited in sequence databases such as SwissProt or PIR.[5]

Individual tropocollagen molecules are too long to currently simulate in total. However, short segments have been simulated.[6-10] Short microfibrils consisting of five triple helix strands of (Gly-X-Y)[6-12] triplets in the "Smith-like" pentagonal arrangement have been studied computationally since the 1980's.[11-16] Early studies by Docherty, et.al., Brown, et. al. were carried out with forcefield energy minimization techniques and semi-empirical quantum mechanics. The emphasis was on the structure of the model tropocollagen strands and their interaction with one another via side chain functionality. The possible interactions of acidic side chains with chromium clusters were also briefly examined based on geometric arguments and forcefield methods. These studies provided the ground work for understanding the intra- and interstrand interactions such as hyrogen bonding and the role of tightly bound water.

The question of how chromium salts interact with collagen has been qualitatively been examined as described above. In order to properly describe the system more sophisticated methods are called for, such as density functional theory. With todays computing power relevant calculations including chromium salts are clearly do-able but have not been done to the best of our knowledge. Parthasarathi et.al. published an ab initio study (Hartree-Fock and Density functional theory) on the stability of collagen triplets in 2003.[17] In this paper they examined the stability and free energy differences of different GLY-X-Y triplets in an extended and helical conformations. Despite the detailed information provided by this study they are too limited for general applicability to realistic collagen models.

Leather chemists are interested in the interaction of process chemicals, tanning agents, pickling salts, dyes, etc. with collagen. Therefore, in order for theoretical models to have relevance for lab chemists the models must be extended. Cappelli and Monti et al. have reported on the stability and interactions of formaldehyde and gallic acid with smith-like collagen bundles (5 TH's consiting of 24 GLY-X-Y triplets) surrounded with explicit water in the form of a periodic box and mixtures of water/formaldehyde and water/gallic acid.[18] The influence of the additives on the collagen micro fibril

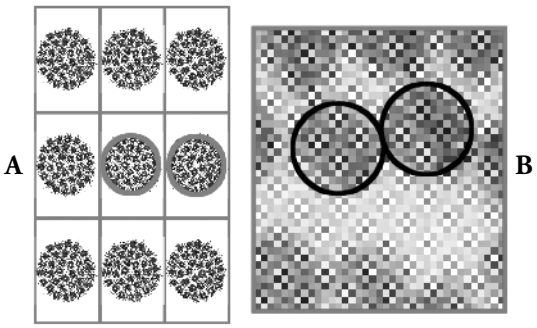

Figure 4. A) periodic fibril model,
B) TEM fibril cross-section of type-1 collagen (enlargement)

structure was examined with molecular dynamics. Simply stated, molecular dynamics is a force field coupled with Newtons' laws of motion so that one can study the development and changes of a molecular ensemble over time and differing temperatures. A force field views molecules as charged balls - coulombic interactions- coupled with "springs" for bond distances, bond angles and torsions. They are empirically parameterized to reproduce crystal structures or physical properties. Cappelli and Monti examined which interactions are favorable between water, formaldehyde and gallic acid and the collagen amino acids. These small molecules can be viewed as models for polyfunctional aldehydes or polyphenolic tanning agents and demonstrate the differing interactions with side chains: formaldehyde is a hydrogen bond acceptor and interacts strongest with amino side chains whereas gallic acid has hydrogen bond donor and acceptor attributes and interact with a greater proportion of the collagen surface. Pure water perturbed the fibril structure the most, while formaldehyde and gallic acid stabilized it.

An extension of this work was published in 2005 where the authors looked at the influence of solvent mixtures (water, water/formaldehyde and water/gallic acid) on two adjacent smith-like collagen bundles (5TH's each 23 amino acids long), again in an explicit periodic solvent box.[19] Four different two fibril models were constructed and studied with molecular dynamics at acidic and neutral pH. The two collagen microfibrils were within interaction distance so that the effect of solvent mixtures on the interfibril interactions (primarily hydrogen bonding) could be studied. The effect of the different solvent systems paralled those found in the initial paper described above.

It is evident that with increasing computing power and more efficient algorithms the size and complexity of the systems that can be studied with molecular dynamics and quantum chemical approaches is rapidly expanding. BASF became interested in leather simulations in 2003. At that time we set out to develop a model of a collagen microfibril with the following requirements (see Figure 4): a) a periodic model in all three dimensions so that there are no end effects in the fibril along the fibril axis (we did not model the telopeptide end sections and ignored the gap region) and that neighboring fibrils are taken into account without explicitely having to build them (this improves computational efficiency and allows

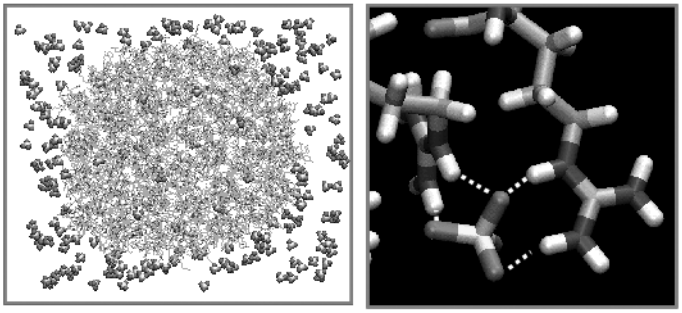

Figure 5: Salt interactions. Anion interaction of sulfate with collagen
NOTE: Sodium ions have been omitted for clarity.

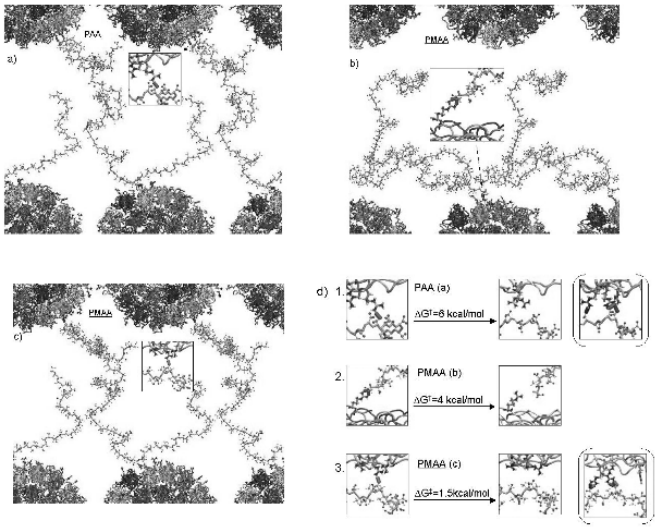

Figure 6. a) PAA-collagen hydrogen bond between a central carboxylate group and an arginine residue.
b) PMAA-collagen hydrogen bond between the mobile tail of the polymer and an arginine residue.
c) PMAA-collagen hydrogen bond of the system constructed from a bonded PAA conformation.
d) Barriers for breaking of the hydrogen bonds.

us to increase fibril size), b) fibril diameter should be large enough to differentiate between TH's surrounded only by other TH's and TH's at the fibril solvent interface, c) the water box should constructed so that fibril fibril interactions are intact in one dimension (X-axis) and that there is a water buffer between the fibril in the other direction, d) the fibrils should be long enough to allow the addition of leather process chemicals. The amide side chains (GLN and ASN) were hydrolysed to the amines (GLU and ASP) as happens during the stringent conditions of leather production.

In 2004 we initially reported on molecular dynamics simulations on the initial models (27 TH's each 28 AA's long, MD program was NAMD[20]) at pH 1,7 and 14 with the appropriate protonation and counterions to ensure charge neutrality.[21] To validate our model we ran MD simulations at constant pressure and temperature (300K, 1 bar) to check the equilibrium density of the neutral molecule. The simulation box had a density of 1.44 g/cm^3 which is in good agreement with an experimental value of 1.35 g cm^3 (density varys over the hide and from animal to animal). A further validation of

the model came from simulations of swelling at the different pH's and pickling with various salts (NaCl, $CaCl_2$, Na_2SO_4, see Figure 5) as well as urea (1% and 12% w/w).[22] Results show that addition of salt prevents swelling at very low and high pH to varying degrees, just as one would expect in the lab. We see the strongest interaction with sulfate rather than chloride. This is in accordance with the Hofmeister series for the interaction of salts with proteins (F^-,SO_4^{2-} > HPO_4^{2-} > acetate > Cl^- > NO_3^- > Br^- > ClO_3^-).[23] Urea was found to concentrate at the collagen surface during the course of the simulation, however, the simulation time of 2ns was not sufficient to see the beginnings of denaturation with the 12% solution that would be expected and much longer simulations would be required.

This question of simulation time of large systems with many degrees of freedom is of general applicability in molecular dynamics. Conformational changes, large translocations, etc. generally take place at time scales that are not accessible to molecular dynamics, with it's 1 femtosecond step size. Fortunately, methods are being developed to overcome some of the shortcomings. In the group of Parrinello the so called "metadynamics" method has been implemented to force molecules to cover the desired phase space and one can extract the relative free energies for the process under investigation.[24-25] In the case of collagen simulations it is possible with the use of metadynamics to simulate polymeric retanning agents. We examined the effect of adding polyacrylic acid (PAA) and poly methyl acrylic acid (PMAA) as 200mer polymers to the 34 TH fibril model at pH=7. PMAA is a retanning agent and PAA does not work in this application. PMAA is more hydrophobic and stiffer with a larger radius of gyration than PAA. Otherwise they are structurally quite similar and the question was whether-or-not a difference could be detected and explain the difference in tanning properties.

Our simulations indicate that PMAA forms a better buffer between collagen microfibrils, due to stiffnes and less hydrogen bonding, whereas PAA forms many, relatively strong hydrogen bonds with the fibrils which could allow the fibrils to come into close proximity forming a parchment -like structure. More simulations are needed to confirm these results. Even with the use of metadynamics there is still a lot of phase space that would need to be covered before these questions can be definitively answered. The good news is that computing power is continuously going up and the price coming down. Also the software is being optimized for highly efficient parallel processing. These facts combine to make clusters of several hundred linux based PC's running efficient academic code such as NAMD financially accessible to even small organizations interested in extending their knowledge of these highly complex systems.

REFERENCES

1. Fundamentals of Leather Manufacturing, E. Heidemann, Eduard Roether KG, Darmstadt, **1993**.
2. a) G. N. Ramachandran, G. Kartha, *Nature* **1954**, *174*, 269-270; b)G. N. Ramachandran, G. Kartha, *Nature* **1955**, *176*, 593-595; c) M. Bansal, *Resonance*, **2001**, 38.
3. J. P. R. O. Orgel, T. C. Irving, A. Miller, T. J. Wess, *Procs. Natl. Acad. Sci. USA* **2006**, *103(24)*, 9001-9005.
4. T.J. Wess, A.P. Hammersley, L. Wess and A. Miller, *J. Struct. Biol.*, **1998**, *122*, 92-100
5. Protein Information resource at http://www.pir.georgetown.edu and http://ca.expasy.org/
6. A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick, B. Brodsky, *Biochemistry* **2005**, *44*, 1414-1422
7. R. Z. Kramer, M. G. Venugopal, J. Bella, P. Mayville, B. Brodsky, H. M. Barman, *J. Mol. Biol.* **2000**, *301*, 1191-1205
8. M. J. Buehler, *J. Mat. Res.* **2006**, *21(8)*, 1947-1961; b) Buehler M.J., P. Natl. Acad. Sci. USA, **2006**, 103, 12285-12290
9. S. D. Mooney, P. A. Kollman, T. E. Klein, *Biopolymers* **2002**, *64*, 63-71
10. S. D. Mooney, C. C. Huang, P. A. Kollman, T. E. Klein, *Biopolymers* **2002**, *58*, 347-353.
11. R. Docherty, D. Buttar, L. Tricker, R. M. Swart, *JALCA*. **1997**, *92*, 185.
12. J. M. Chen, S. H. Feairheller, E. M. Brown, *JALCA*. **1991**, *86*, 475.
13. E. M. Brown, G. King, *JALCA*. **1996**, *91*, 161.
14. J. M. Chen, C. E. Kung, S. H. Feairheller, M. Brown, *J. Prot. Chem*. **1991**, *10(5)*, 535.
15. E. M. Brown, J. M. Chen, S. H. Feairheller, *JALCA*. **1993**, *88*, 2.
16. G. King, E. M. Brown, J. M. Chen, *Prot. Engineering*, **1996**, *9(1)*, 43.
17. R. Parthasarathi, B. Madhan, V. Subramanian and T. Ramasami, *Theor. Chem. Acta*, **2003**, *110*, 19.
18. S. Bronco, C. Capelli, S. Monti, *J. Phys. Chem. B.* **2004**, *108,* 10101-10112.
19. S. Monti, S. Bronco, C. Capelli, , *J. Phys. Chem. B.* **2005**, *109*, 11389-11398.
20. L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comp. Phys.*, **1999**, *151*, 283-312.
21. L. Siggel and F. Molnar, *JALCA*, **2006**, *101*, 179.
22. R.-E. Bulo, L. Siggel, F. Molnar and H. Weiss, *Macromol. Biosci.* **2007**, *7*, 234.
23. W. Kunz, P. Lo Nostro, B. W. Ninham, *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 19-37.
24. A. Stirling, M Ianuzzi, A. Laio, M. Parrinello, *ChemPhysChem*, **2004**, *5*, 1558-1568.
25. A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, M. Parrinello, *J. Phys. Chem. B*, **2005**, *109*, 6714-6721.