

DISS. ETH NO. 17906

Identification and Characterization of Collagen Glycosyltransferases of Human and Viral Origin

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
BELINDA SERAINA PAMELA SCHEGG

Dipl. Natw. ETH
born on 17.12.1979
citizen of Oberriet-Montlingen SG

accepted on the recommendation of
Prof. Dr. Markus Aebi, examiner
Prof. Dr. Sabine Werner, co-examiner
Prof. Dr. Thierry Hennet, external examiner

2008

Summary

Collagens are a superfamily of glycoproteins mainly found in the extracellular matrix. They are the most abundant proteins in the human body. Collagens are characterized by a right handed triple helix formed out of three left-handed α -chains representing repeats of the motif G-X-Y, where (hydroxy)proline and (hydroxy)lysine are often found at positions X and Y. To act as a functional collagen, selected hydroxylysine residues have to be further modified by the addition of either galactose or the disaccharide glucosylgalactose. This glycosylation of collagen takes place in the endoplasmic reticulum before the formation of the triple helix and is mediated by specific $\beta(1-O)$ galactosyl- and $\alpha(1-2)$ glucosyltransferase enzymes. The molecular nature of these glucosyltransferases has remained unknown to date. The present study describes the identification of collagen galactosyltransferase enzymes using a strategy based on affinity chromatography and protein sequencing by mass spectrometry. Three structurally related candidate genes were cloned and expressed in Sf9 insect cells using the baculovirus system. Two of the three candidate glucosyltransferases (GLT25D1 and GLT25D2) were confirmed to be active collagen galactosyltransferases. The collagen galactosyltransferase genes are differentially expressed in human tissues, suggesting that these enzymes may show preference for different types of collagens or contribute to the varying extent of collagen glycosylation throughout tissues. This was supported by showing a selective preference of GLT25D1 and GLT25D2 for collagen type III and collagen type IV acceptors. GLT25D1 showed a higher enzymatic activity on deglycosylated collagen type I to type V than GLT25D2.

Collagen glycosylation is conserved in animals and collagen is also found in several prokaryotic genomes. Proteins sharing structural similarity with the collagen galactosyltransferases have been found in prokaryotes and even in virus. The *Acanthamoeba polyphaga* mimivirus has been detected as a unique member of the nucleo-cytoplasmic large DNA virus family being clearly a virus but also showing features never seen before in viruses. Mimivirus is the largest known DNA virus with a 1.2 Mbp linear dsDNA ge-

nome. It was reported that mimivirus encodes eight proteins with a collagen triple helix motif. These collagens are most probably found in the fibrils of the virus capsid. These fibrils cover the whole icosahedral virus capsid, which are specific for mimivirus. As the Gram staining of mimivirus is positive, it is supposed that the viral fibrils are glycosylated. This suggests that the mimiviral collagens might be post-translationally modified by hydroxylation and subsequent glycosylation.

In this study, the protein L230 was identified as a mimiviral collagen glucosyltransferase transferring glucose on the acceptor hydroxylysine on animal and on mimiviral collagens. This addition of glucose on the acceptor hydroxylysine in collagen has not been reported up to now. It seems that collagen glycosylation in mimivirus is different than in animal collagens.

In conclusion, this work reports the identification and characterization of the two human collagen galactosyltransferases GLT25D1 and GLT25D2 and the identification of the mimiviral collagen glucosyltransferase L230.

Zusammenfassung

Kollagene sind Glykoproteine, die hauptsächlich in der extrazellulären Matrix vorkommen und sind daher die am häufigsten vorkommenden Proteine im menschlichen Körper. Kollagene haben eine charakteristische rechts drehende tripel-helikale Struktur. Diese Helix wird aus drei links drehenden α -Ketten aufgebaut, welche aus dem repetitiven Sequenzmotif Gly-X-Y bestehen. Häufig stehen an den Positionen X und Y (Hydroxy)Prolin oder (Hydroxy)Lysin. Damit das Kollagen funktionell aktiv sein kann, muss es mit Galaktose oder Glukosylgalaktose post-translationell modifiziert werden. Diese Kollagen Glykosylierung findet im Endoplasmatischen Retikulum statt bevor die Tripel-Helix gebildet wird und wird von einer spezifischen $\beta(1-O)$ Galactosyltransferase und einer $\alpha(1-2)$ Glucosyltransferase durchgeführt. Diese beiden Glykosyltransferasen waren bis heute noch nicht bekannt. Diese Arbeit beschreibt die Identifizierung dieser Kollagen Galaktosyltransferasen, welche mit Hilfe einer Affinitätschromatographie und anschließender Proteinsequenzierung mittels Massenspektroskopie identifiziert werden konnten. Daraufhin wurden drei strukturell verwandte Kandidatengene kloniert und mittels Bakuloviren in Sf9 Insektenzellen exprimiert. Zwei der drei gefundenen Kandidaten (GLT25D1 und GLT25D2) zeigten in einem Aktivitätstest eine Kollagen-Galaktosyltransferase-Aktivität.

Die Kollagen Galaktosyltransferasen sind in menschlichen Geweben unterschiedlich exprimiert. Dies lässt vermuten, dass diese Enzyme eine Präferenz für verschiedene Kollagen Typen zeigen könnten, oder dass sie dazu beitragen, dass die Kollagene in den verschiedenen Geweben verschieden stark glykosyliert werden. Diese Vermutung wurde bestärkt indem gezeigt wurde, dass GLT25D1 und GLT25D2 präferentiell Kollagen Typ III und Kollagen Typ IV glykosylieren. GLT25D1 hingegen zeigte eine höhere enzymatische Aktivität auf deglykosyliertem Kollagen Typ I bis Kollagen Typ V als GLT25D2.

Kollagen Glykosylierung in Tieren ist ein stark konservierter Prozess. Kollagen selbst findet man jedoch nicht nur in höheren Lebewesen, sondern auch in verschiedenen prokaryontischen Organismen. Daher ist es nicht erstaun-

lich, dass auch Proteine mit strukturellen Ähnlichkeiten zu den Kollagen Galaktosyltransferasen in Prokaryonten und sogar in Viren gefunden wurden. Ein Beispiel dafür ist das *Acanthamoeba polyphaga* Mimivirus. Es ist ein einzigartiges Wesen, das auf der einen Seite klar ein Virus ist, jedoch auf der anderen Seite auch charakteristische Merkmale hat, die nie zuvor in Viren, gefunden worden sind. Mit einem 1.2 Mbp grossen linearen doppelsträngigen DNS Genom ist das Mimivirus das grösste bekannte DNS Virus. Es wurde berichtet, dass das Mimivirus acht Proteine kodiert, welche ein Tripel-Helix Kollagen Motiv haben. Diese Kollagene findet man mit grosser Wahrscheinlichkeit in den Fibrillen auf dem Viruskapsid. Die für das Mimivirus spezifischen Fibrillen bedecken das ganze ikosahedrische Viruskapsid. Die Gram-Färbung des Mimivirus ist positiv, was darauf hindeutet, dass die mimiviralen Kollagene durch Hydroxylierung und anschliessende Glykosylierung post-translationell modifiziert werden.

In dieser Arbeit wurde das Protein L230 als mimivirale Kollagen Glukosyltransferase identifiziert. Wir konnten zeigen, dass L230 die Übertragung von Glukose auf den Hydroxylysin-Akzeptor sowohl in tierischen als auch in mimiviralen Kollagenen katalysiert. Dieser Transfer von Glukose auf den Hydroxylysin-Akzeptor in Kollagen wurde bis jetzt noch nicht beschrieben. Daraus lässt sich schliessen, dass sich die Kollagen Glykosylierung im Mimivirus von der Kollagen Glykosylierung in Tieren unterscheidet.

Zusammenfasst zeigt diese Arbeit die Identifizierung und Charakterisierung der zwei humanen Kollagen Galaktosyltransferasen GLT25D1 und GLT25D2 sowie die Identifizierung der mimiviralen Kollagen Glukosyltransferase L230.