# Collagen Type XI α1 Chain Amino Propeptide Structural Model and Glycosaminoglycan Interactions *in Silico*

Chris Mallory[1,3], Owen McDougal[1,3], Julia Thom Oxford[2,3]

[1]Department of Chemistry and Biochemistry, [2]Department of Biological Sciences, [3]Biomolecular Research Center, Boise State University, Boise, ID 83725, USA

**Abstract-***Modeling of the collagen α1(XI) amino propeptide (NPP) domain was performed to better understand how dimerization and glycosaminoglycan binding are coordinated. The program MODELLER was used to generate a homology model of collagen α1(XI) NPP domain based on the crystal structure of the closely related NC4 domain of collagen α1 (IX) (PDB:2UUR) to a root mean square deviation (rmsd) of 0.785 Å resolution. A model of collagen α1(XI) NPP domain dimer was constructed in two alternative templates; 1) the thrombospondin dimer template (PDB:1Z78), and 2) by submission of two monomer subunits based on PDB:2UUR to ClusPro. Calculation of relative binding energy for the interaction between each collagen α1(XI) NPP model and glycosaminoglycans as ligands was performed using AutoDock4. Results support a higher affinity between heparan sulfate and the dimer compared to the monomer. Sequential point mutation studies in the putative binding site (147-KKKITK-152) indicated the importance of each basic lysine residue in the binding of heparan sulfate. Two orders of magnitude change in binding affinity was predicted when comparing wild type to the mutation K152A.*

**Keywords:** collagen, heparin, molecular interaction, glycosaminoglycan, protein

## 1 Introduction

Collagen is a triple helical protein comprising approximately 25% of the protein contained in the human body. The triple helix of collagen is unique due to its formation from three left-handed helical strands to compose the right-handed triple helix. The strands that make up this triple helix have the sequence Gly-Xxx-Yyy; where approximately 30% of the Xxx and Yyy are proline and hydroxyproline, respectively [1-3]. To date more than 27 different collagens have been reported in the literature, of which 16 have non-collagenous domains attached to the extended collagen triple helix. Collagen type XI is a minor fibrillar collagen involved in regulating the diameter of collagen fibrils [1,4]. Composed of three different left-handed helical strands α1, α2, and α3, each of the alpha chains contains non-collagenous domains. The α1 amino terminal non-collagenous domain (NPP) is proteolytically cleaved at a much slower rate than α2 or α3 and is therefore resident on the surface of collagen fibrils for an extended period of time in tissues [1]. Currently, a protein data base structure file is not publically available for the Npp α1, but a recently published structure for the NC4 domain of collagen IX is available that demonstrates remarkable structural similarity to collagen XI [5]. The NPP domain is included in a family of laminin, neurexin, sex hormone binding globulin (**LNS**) domains, in which a crystal structure of thrombospondin has also recently appeared in the literature in a monomer and dimer form [6].

There is experimental evidence to support the hypothesis that the Npp domain of collagen α1(XI) interacts with glycosaminoglycans such as heparan sulfate. This interaction is proposed to be significant in determining the thickness of the fibril as it forms [4]. These interactions occur at a glycosaminoglycan binding within collagen α1(XI) consisting of 147-KKKITK-152. Independently, experimental evidence indicates the formation of an NPP α1(XI) dimer. Using a combination of homology modeling and protein-protein docking, a computational model for the α1(XI) NPP was created. Models were used in the docking program AutoDock4 to calculate the energy of interaction between collagen α1(XI) NPP (monomer and dimer) with the heparan sulfate ligand [7]. The results of the docking study provide a

theoretical inhibition constant ($K_i$), molecular binding energies, and predicted atomic interactions such as hydrophobic, electrostatic, and hydrogen bonding.

The conservation and importance of each basic lysine residue in the putative binding site (147-KKKITK-152) were further evaluated by point mutations. Each positively charged residue was suspected to be critical for glycosaminoglycan binding and was analyzed by computational docking analysis of the point mutants in AutoDock4.

# 2 Methods

## 2.1 Collagen α1(XI) NPP homology model monomer

The sequence for Collagen α1(XI) NPP was submitted to BLAST against available PDB structures for template identification [8]. Blast returned two possible templates; thrombospondin (PDB:1Z78) and the NC4 domain of collagen α1(IX) (PDB:2UUR) [5,6,8]. The sequences for each template were aligned independently to the Collagen α1(XI) NPP domain, and then collaboratively using the '***salign***' command in MODELLER [9]. Homology models for Collagen α1(XI) NPP domain using both 2UUR and 1Z78 as templates were then created using the MODELLER method with disulfide bonding specified between Cys 25–Cys 207 and Cys 146–Cys 200 [1,9]. The returned models were evaluated using PROCHECK for areas of high energy, restricted points on the Ramachandran plot, and residue clashes [10]. Loop rebuilding and energy minimization were performed in MODELLER, while corrections to amino acids contained in restricted areas of the Ramachandran plot were amended in Chimera [11]. An rmsd was calculated for the homology model generated from each template (2UUR & 1Z78), using Chimera [11]. The homology model created from 2UUR (HM1) was selected as the best model based on sequence alignment, query match, E value, and an rmsd of 0.785Å from template structure. Swiss Deep View was used to make the following single-point mutations into the Collagen α1(XI) NPP domain putative binding site: K147A, K148A, K149A, and K152A [12].

## 2.2 Collagen α1(XI) homology model dimer

A dimer model of the Collagen α1(XI) NPP was created with MODELLER and ClusPro using the thrombospondin dimer (PDB:2ES3) as a template. The homology model was created by the same process as the monomer model with the exception that a repeated sequence of Collagen α1(XI) NPP was used. Model evaluation was performed using PROCHECK, Verify3d, and Ramachandran plot analysis. A second dimer model was created by submitting two identical homology model monomers to the ClusPro server [13]. The overall lowest energy model was selected from the balanced interaction cluster.

## 2.3 Docking Studies

Docking studies between the Collagen α1(XI) NPP monomer homology model, single-point putative binding site mutants, MODELLER dimer, and the ClusPro dimer with the glycosaminoglycan heparan sulfate ( dissacharide, decasaccharide, and Arixtra, a low molecular weight heparin) as the ligand were performed using AutoDock4 [7]. The AutoDock standard conditions for a large run were used with a grid spacing of 0.379 Å. Binding energies were evaluated using WordPad, with residue interactions visualized using Chimera [11].

# 3 Results

## 3.1 Homology model

BLAST search results indicated that the NC4 domain of collagen IX (PDB:2UUR) was the best template with an E value of $5.0 \times 10^{-5}$. Alignment and model creation was performed using MODELLER with PDB:2UUR as a template. The computational model of the Collagen α1(XI) NPP monomer from 2UUR was found to have an rmsd of 0.785 Å (Fig. 1a) [5,11]. Point mutations were introduced into the putative binding site sequentially using Swiss Deep View. Docking interactions using a heparan sulfate disaccharide and Collagen α1(XI) NPP monomer models indicated that all lysine residues were necessary for best binding and lowest estimated inhibition constant (Table I). Submission of the wild type structural monomer to ClusPro, resulted in clusters returned based on hydrophobic, electrostatic, van der Waals + electrostatic, and balanced interactions. Without previous dimerization knowledge the lowest energy balanced model with a weighted score of -918.2 was selected, as recommended by ClusPro (Fig. 1b) as the optimal dimer model [13-16].

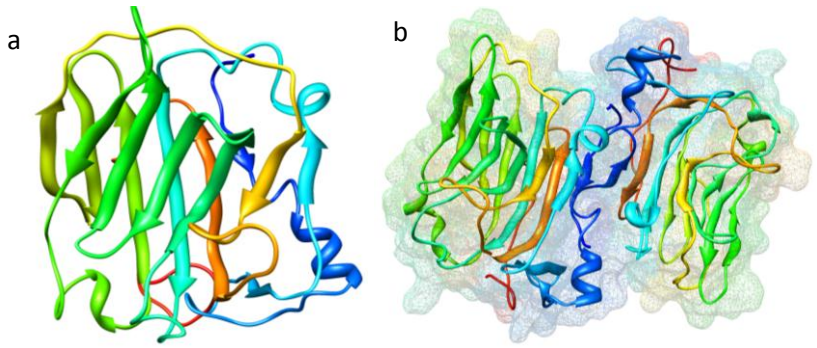


**Fig. 1:** Structural depictions of: a) Homology model monomer built from the template 2UUR; b) Lowest energy balanced interaction dimer from ClusPro [5,13].

**Table I**
Monomer homology model and mutant interactions between heparan sulfate disaccharide.

| Model | Estimated Free Energy (kcal/mol) | Estimated $K_i$ | Hydrogen bond residue interactions | Polar residue interactions |
|---|---|---|---|---|
| Wild Type | -8.08 | 1.20 μM | ARG119, LYS149 | ARG119, LYS149 |
| K147A | -8.37 | 727.94 nM | PHE118, LYS152 | PHE118, LYS152 |
| K148A | -8.41 | 687.57 nM | LYS152 | LYS152 |
| K149A | -7.42 | 3.64 μM | LYS148 | LYS148 |
| K152A | -5.34 | 121.85 μM | - | LYS149 |

*The estimated inhibition constant ($K_i$) describes the binding affinity for the ligand to the receptor. This is not to be confused with the dissociation constant ($K_d$) that provides insight into how easily the receptor-ligand complex separates into its individual components, the receptor and ligand.

### 3.2 Binding Studies

Docking results for heparan sulfate and each dimer model confirmed the hypothesis that dimerization increases affinity for glycosaminoglycans relative to the monomer, as shown in Table II.

**Table II**
Interactions between ClusPro dimer and a heparan sulfate disaccharide, decasaccharide, and Arixtra[7,13-16].

| Ligand | Est. Free Energy of Binding (kcal/mol) | Est. Ki | Electrostatic Energy (kcal/mol) | Total Intermolecular Energy (kcal/mol) | Hydrogen bond residue interactions | Polar residue interactions |
|---|---|---|---|---|---|---|
| Heparan sulfate disaccharide | -17.53 | 142.31 fM | -10.57 | -12.65 | - | LYS73, ASN122, LYS149 |
| Heparan sulfate decasaccharide | -21.13 | 327.17 aM | -22.22 | -23.18 | LYS148 | LYS73, ARG97, ARG119, LYS148, TYR197 |
| Arixtra | -7.93 | 1.53 μM | -5.07 | -7.92 | LYS73 | LYS72, ARG97, ASP125 |

## 4 Discussion

Point mutation docking studies revealed that K148A had the lowest estimated binding energy, providing a preliminary hypothesis that K148 is the least critical residue involved in binding of glycosaminoglycans. Furthermore, K152 appears to be the most important residue in the putative binding site involved in glycosaminoglycan

interaction as evidenced by a predicted two orders of magnitude increase in the $K_i$. Initial experimental results suggest that a greater affinity for glycosaminoglycans occurs upon dimerization or oligomerization of Collagen α1(XI) NPP. Binding results for heparan sulfate dissacharide, decasaccharide, and Arixtra support the experimental results of increasing affinity for glycosaminoglycans upon dimerization of the Collagen α1(XI) NPP. While it is clear that they are coordinated, it is still unclear whether dimerization is facilitated upon interactions with glycosaminoglycans or alternatively, if dimerization occurs prior to binding, and subsequently facilitates interaction with glycosaminoglycans.

## 5 Conclusion

To date there is no crystal structure available for the Collagen α1(XI) NPP domain. Using computational modeling, we have created a homology model based on the NC4 domain of collagen IX (PDB:2UUR) as a template. The resulting model was used investigate glycosaminoglycan binding and provided insight into protein:glycosaminoglycans interactions. *In silico* prediction provided preliminary insight into the importance of K152 in the binding interactions with glycosaminoglycans, from a 100 fold decrease in binding affinity compared to that of the wild type.

Furthermore, dimerization of the Collagen α1(XI) NPP domain has been observed experimentally to increase the binding affinity to glycosaminoglycans. This experimental result was successfully replicated through the modeling of a dimer and docking interactions of heparan sulfate and its derivatives. Additional studies are being conducted to determine if glycosaminoglycan binding induces dimerization, or alternatively, if increased affinity for glycosaminoglycans is a result of Collagen α1(XI) NPP dimerization.

## Acknowledgments

Authors wish to acknowledge technical and editorial support from Luke Woodbury. This work was supported in part by grants from the National Institutes of Health/National Center for Research Resources Grant P20RR16454, NIH/NICHD R15HD059949, NASA (NNX10AN29A), Arthritis Foundation, Research Corporation Cottrell College Scholars, and Mountain States Tumor Medical Research Institute, National Institutes of Health/NIAMS RO1AR47985 and KO2AR48672, the M. J. Murdock Charitable Trust, and Lori and Duane Stueckle Professorship.

## References

[1] A. Fallahi, B. Kroll, L.R. Warner, R.J. Oxford, K.M. Irwin, L.M. Mercer, S.E. Shadle, J.T. Oxford, "Structural model of the amino propeptide of collagen XI α1 chain with similarity to the LNS domain," *Prot. Sci.*, *14*, 1526-1537, 2005.

[2] N.P. Morris, H.P. Bachinger, "Type XI Collagen Is a Heterotrimer with the Composition (1α, 2α, 3α) Retaining Non-triple-helical Domains," *J. Biol. Chem.*, *262*, 11345-11350, 1987.

[3] M.K. Gordon, R.A. Hahn, "Collagens," *Cell Tissue Res.*, *339*, 247-257, 2010.

[4] D.R. Keene, J.T. Oxford, N.P. Morris, "Ultrastructural Localization of Collagen Types II, IX, and XI in the Growth Plate of Human Rib and Fetal Bovine Epiphyseal Cartilage: Type XI Collagen is Restricted to Thin Fibrils," *J. Histochem. Cytochem.*, *43*, 967-979, 1995.

[5] K. Tan, M. Duquette, J.H. Liu, K. Shanmugasundaram, A. Joachimiak, J.T. Gallagher, A.C. Rigby, J.H. Wang, J. Lawler, "Heparin-induced cis- and trans-dimerization modes of the thrombospondin-1 N-terminal domain," *J Biol Chem.*, 283:3932-41, 2008.

[6] V.–M. Leppanen, H. Tossavainen, P. Permi, L. Lehtio, G. Ronnholm, A. Goldman, I. Kilpelainen, T. Pihlajamaa, "Crystal structure of the N-terminal NC4 domain of collagen IX, a zinc binding member of the laminin-neurexin-sex hormone binding globulin (LNS) domain family," *J. Biol. Chem.*, *282*, 23219-23230, 2007.

[7] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function," *J. Comput. Chem.*, *19*, 1639-1662, 1998.

[8] Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T. L. "NCBI BLAST: A Better Web Interface," *Nucleic Acids Res.*, *36*, W5-W9, 2008.

[9] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, "Comparative Protein Structure Modeling of Genes and Genomes," *Annu. Rev. Biophys. Biomol. Struct.*, *29*, 291-325, 2000.

[10] R.A. Laskowski, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Cryst.*, *26*, 283-291, 1993.

[11] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, "UCSF Chimera-A Visualization System for Exploratory Research and Analysis," *J. Comput. Chem.*, *25*, 1605-1612, 2004.

[12] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling," *Bioinformatics*, *22*, 195-201, 2006.

[13] D. Kozakov, D.R. Hall, D. Beglov, R. Brenke, S.R. Comeau, Y. Shen, K. Li, J. Zheng, P. Vakili, I.C. Paschilidis, S. Vajda, "Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19," *Proteins: Struct.,Funct., Bioinf.*, *78*, 3124-3130, 2010.

[14] D. Kozak, R. Brenke, S.R. Comeau, S. Vajda, "PIPER: An FFT-Based Protein Docking Program with Pairwise Potentials," *Proteins: Struct.,Funct., Bioinf.*, *65*, 392-406, 2006.

[15] S.R. Comeau, D.W. Gatchell, S. Vajda, "Camacho, C. J. ClusPro: An Automated Docking and Discrimination Method for the Prediction of Protein Complexes," *Bioinformatics*, *20*, 45-50, 2004.

[16] S.R. Comeau, D.W. Gatchell, S. Vajda, C.J. Camacho, "ClusPro: A Fully Automated Algorithm for Protein-Protein Docking," *Nucleic Acids Res.*, *32*, W96-W99, 2004.