

## Supplemental Information

### The cysteine rich region of type VII collagen is a cystine knot with a new topology

Henrik Wegener<sup>a,c</sup>, Hauke Paulsen<sup>b,c</sup>, and Karsten Seeger<sup>a\*</sup>

<sup>a</sup> Institute of Chemistry

<sup>b</sup> Institute of Physics

University of Lübeck

Ratzeburger Allee 160

23562 Lübeck

Germany

<sup>c</sup> HW and HP contributed equally to this work.

Correspondence: Institute of Chemistry, University of Lübeck, 23562 Lübeck, Germany

email: karsten.seeger@chemie.uni-luebeck.de

## Amino acid sequence of CysvWFA2-THC peptide

CysvWFA2-THC-peptide is a fusion protein based on the Col7 subdomain vWFA2, the cysteine rich region and on eight triplets of the collagenous domain. To account for the decreased stability due to the lack of hydroxyprolines, additional six GPP triplets have been added to the sequence (italic). After a glycine spacer, a tyrosine and a histidine tag were incorporated to simplify purification. The amino acid sequence corresponding to Col7-THC-peptide is underlined.

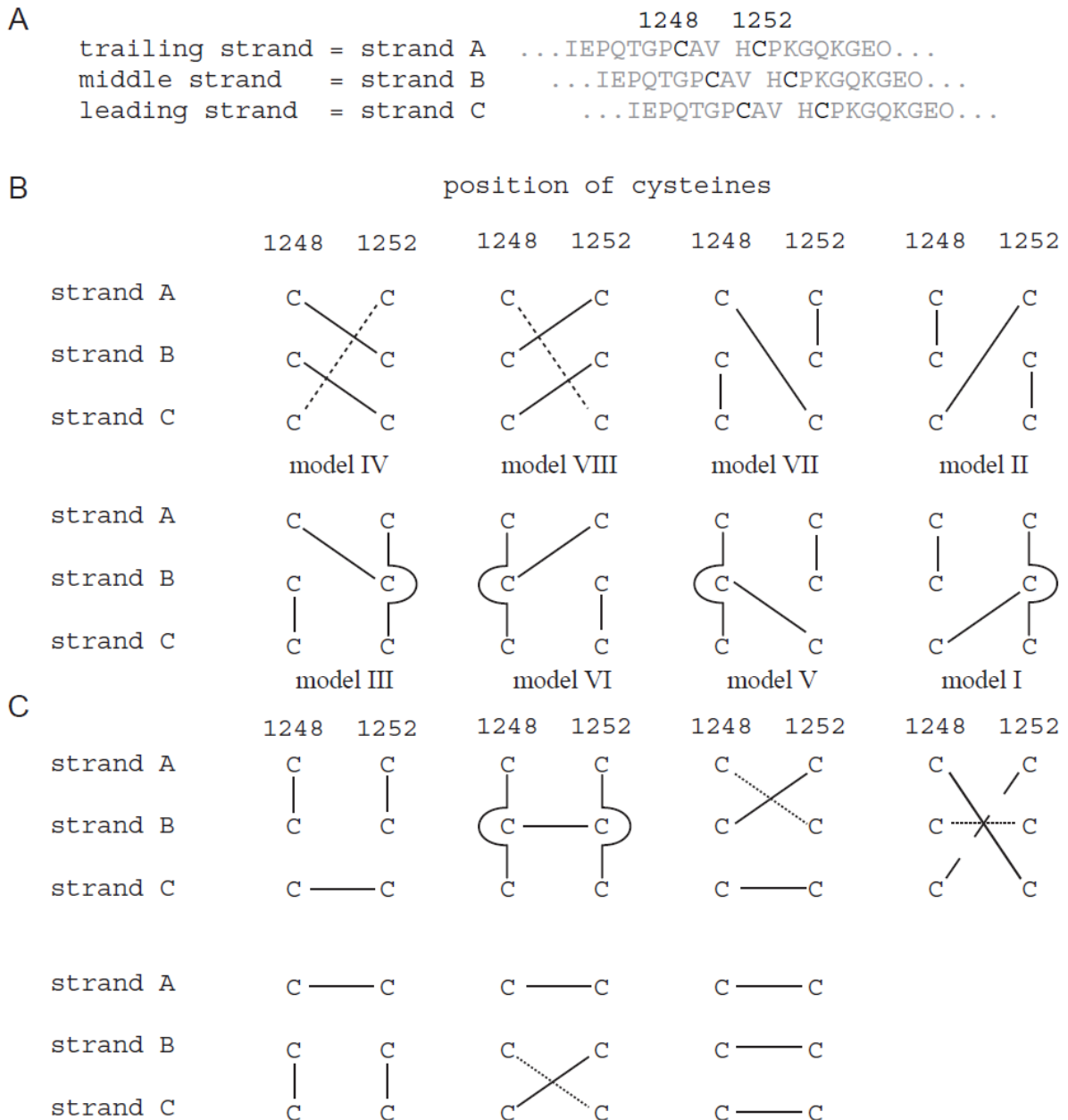
```
      10          20          30          40          50          60
CSHGPDVDFV LLHATRDNAH NAEAVRRVLE RLVSALGPLG PQAAQVGLLT YSHRPSPLFP

      70          80          90         100         110         120
LNSSHDLGII LRKIRDIPYV DPSGNNLGTA VTTAHRYLLA SNAPGRRQQV PGVMVLLVDE

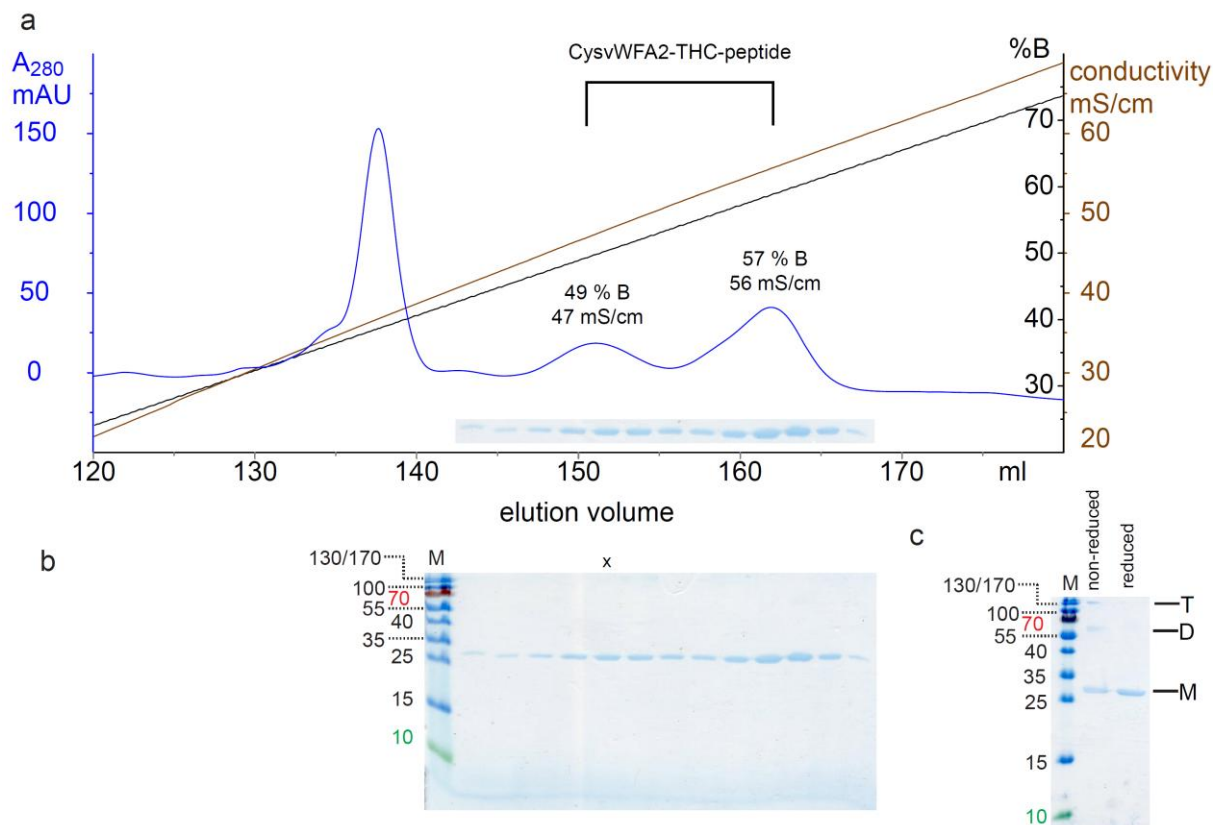
     130         140         150         160         170         180
PLRGDILSPI REAQTSGLKV MALSLVGADP EQLRRLAPGT DPIQNFFAVD NGPGLDRAVS

     190         200         210         220         230         240
DLAVALCQAA VTIEPQTGPC AVHCPKGQKG EPGVTGLQGQ AGPPGPPGLP GPPGPPGPPG

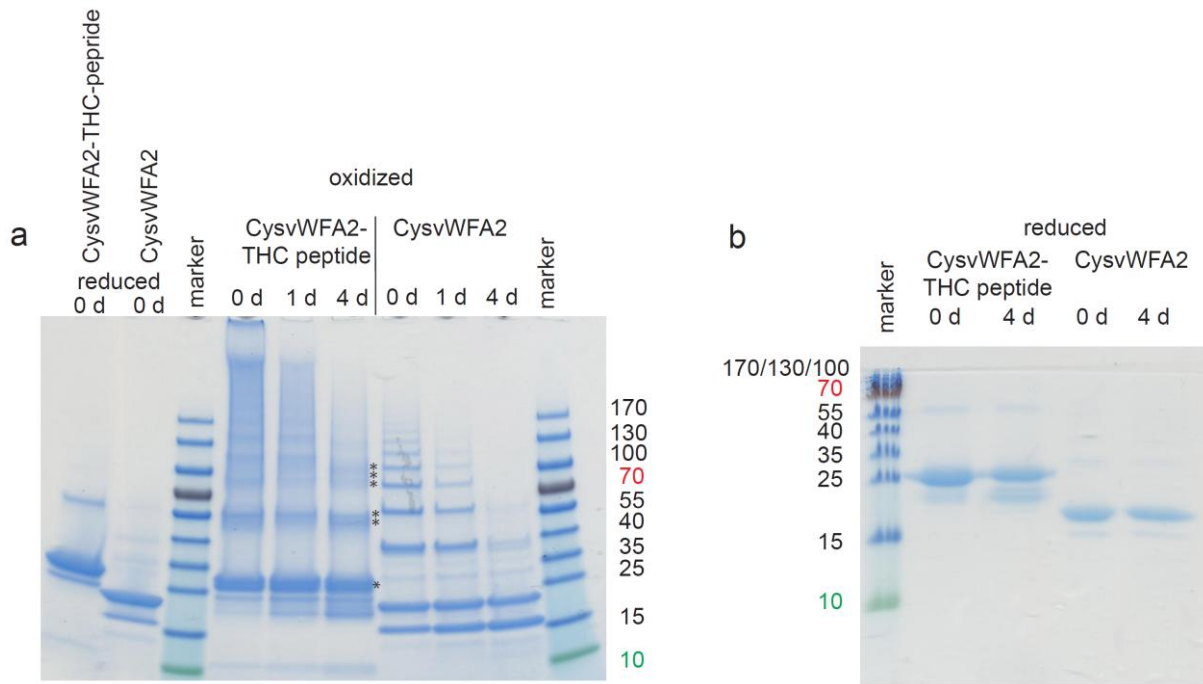
     250         259
PPGPPGPPGG GYMHHHHHH
```



**Supplemental Scheme 1: Possible topologies for interchain disulfide bridges.** **A** The three chains within a collagen triple helix are not identical and can be referred to as leading (C-terminal overhang), middle and trailing strand (N-terminal overhang, cf. Bella, J. (2010) A new method for describing the helical conformation of collagen: Dependence of the triple helical twist on amino acid sequence. *J. Struct. Biol.* **170**, 377-391). There are 15 different patterns of disulfide bridges possible. **B** 8 disulfide patterns are exclusively interchain disulfide bridges resulting theoretically in a collagen trimer. **C** 7 patterns have at least one intrachain disulfide bridge resulting in monomeric and dimeric forms.



**Supplemental figure 1: Elution profile of the cation exchange chromatography of CysvWFA2-THC-peptide.** **a)** CysvWFA2-THC-peptide elutes due to triple helix formation at unusually high salt concentrations. Two peaks containing CysvWFA2-THC-peptide are observed, probably due to different patterns of disulfide bridges. The insert shows the protein bands in an SDS-PAGE of the fractions. The complete SDS-PAGE (15%, Coomassie stained) is shown in **b)**. The fraction marked with an x is again shown in **c)** under reducing and non-reducing conditions. Under oxidative conditions three bands, corresponding to monomeric (M), dimeric (D) and trimeric (T) species are observed.



**Supplemental figure 2: Air oxidation of CysvWFA2-THC-peptide and CysvWFA2 after reduction with TCEP.** SDS-PAGE (4-20 %, Coomassie stained) analysis shows that air oxidation results in several additional bands at higher molecular weight that diminish with time. The sample of CysvWFA2-THC-peptide **a**) at day 4 shows the presence of very low amounts of trimers (\*\*\*) and dimers (\*\*) and as major species monomers (\*). CysvWFA2 shows only monomeric species after 4 days of incubation. Samples taken at day 0 and day 4 have been analyzed also under reducing conditions in the SDS-PAGE **b**) (15 %, Coomassie stained). Air oxidation leads to an incorrect pattern of the disulfide bridges whereas the glutathione redox system allows reshuffling of incorrect formed disulfide bridges resulting in a higher yield of trimer.