

SUPPLEMENTARY APPENDIX

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Mechanisms of Collagen IV345 assembly and dysfunction in Goodpasture's and Alport diseases: II. Crystal structure of α 345 hexamer reveals bioactive sites.

Table of Contents

	Page
Supplementary Section 1: Composition and arrangement of chains within the α345 hexamer (Supplementary Figs. 1 and 2)	2
Supplementary Section 2: Crystal structure of the NC1 α345 hexamer (Supplementary Fig. 3, Supplementary Tables 1-3)	4
Supplementary Section 3: The chloride ions at the NC1 α345 trimer-trimer interface (Supplementary Fig. 4, Supplementary Table 4)	7
Supplementary Section 4: Crevices, pockets, inner cavities, loops and surface-exposed reactive residues in the α345 hexamer (Supplementary Figs. 5-6)	9
Supplementary Section 5: Analysis of known Alport variants in COL4A3, COL4A4, and COL4A5 (Supplementary Figs. 7-9, Supplementary Tables 5 and 6)	11

Supplementary Section 1: Composition and arrangement of chains within the α 345 hexamer

>human_a345NC1

MRAWIFFLLCLAGRALAAPLA**DYKDDDDK**LAGATWTTTRGFVFTTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAH
GQDLGTLGSLQRFTTTPFLFCNVNDVCNFASRNDYSYWLSTPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAV
HSQTTDIPPCPHGWISLWKGFSSFIMFTSAGSEGTGQALASPGSCLEEFRASPFLECHGRGTCNYYSNSYSFWLASLN
PERMFRKPI PSTVKAGELEKI **I SRCQVCMGTG**FLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLA
GSCLPVFSTL PFAYCNIHQVCHYAQRNDRSYWLASAAPLPMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSI PP
CPQTRWRLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFFANKYSFWLTTVKADLQFSSA
PAPDTLKEQAQRQKI SRCQVCV**APG**FLLITRHSQTTDAPQCPQGTLOVYEGFSLLYVQGNKRAHGQDLGTAGSCLRR
FSTMPFMFCNINNVNCFASRNDYSYWLSTPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQG
WDSLWIGYSFMMHTSAGAEGSQALASPGSCLEEFRASPFIECHGRGTCNYYANSYSFWLATVDVSDMFSKQPSETL
KAGDLRTRISRCQVCMKRT*

>human_a543NC1

MRAWIFFLLCLAGRALAAPLA**DYKDDDDK**LAGTSSVAHGFLITRHSQTTDAPQCPQGTLOVYEGFSLLYVQGNKRAH
GQDLGTAGSCLRRFSTMPFMFCNINNVNCFASRNDYSYWLSTPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAV
HSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSQALASPGSCLEEFRASPFIECHGRGTCNYYANSYSFWLATVD
VSDMFSKQPSETLKAGDLRTRISRCQVCMGT**G**FLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLA
GSCLPVFSTL PFAYCNIHQVCHYAQRNDRSYWLASAAPLPMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSI PP
CPQTRWRLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFFANKYSFWLTTVKADLQFSSA
PAPDTLKEQAQRQKI SRCQVCV**APG**FVFTTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSLQR
FTTTPFLFCNVNDVCNFASRNDYSYWLSTPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHG
WISLWKGFSSFIMFTSAGSEGTGQALASPGSCLEEFRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPI PSTV
KAGELEKI **I SRCQVCMKRRH***

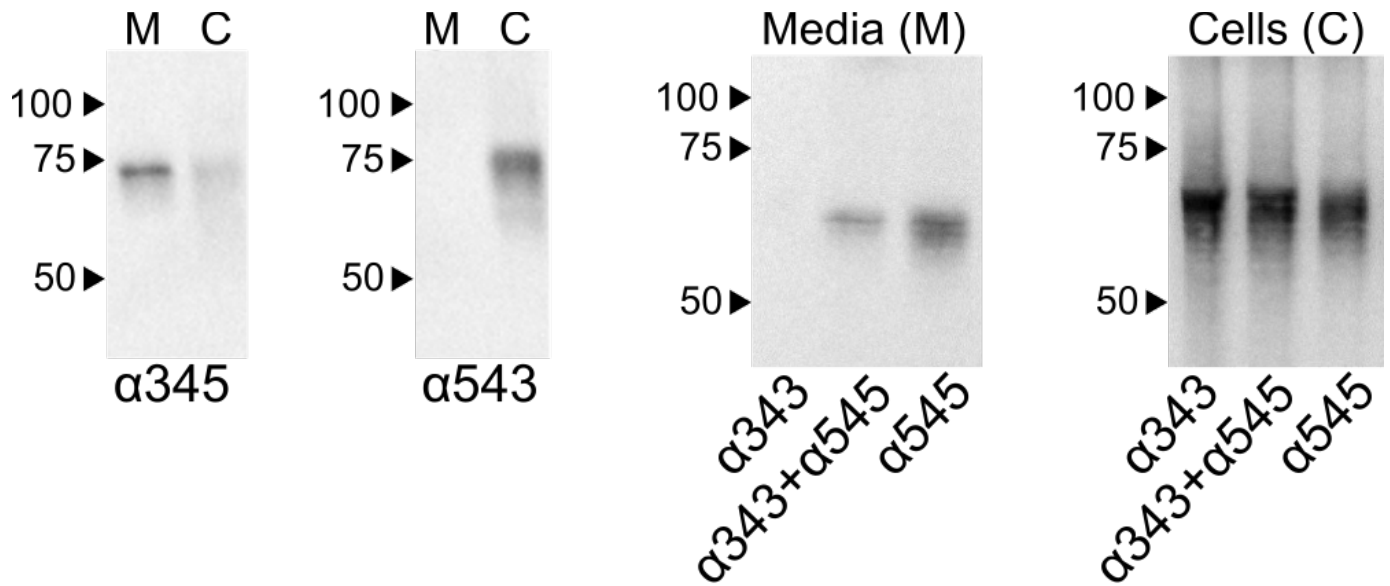
>human_a343NC1

MRAWIFFLLCLAGRALAAPLA**DYKDDDDK**LAGATWTTTRGFVFTTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAH
GQDLGTLGSLQRFTTTPFLFCNVNDVCNFASRNDYSYWLSTPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAV
HSQTTDIPPCPHGWISLWKGFSSFIMFTSAGSEGTGQALASPGSCLEEFRASPFLECHGRGTCNYYSNSYSFWLASLN
PERMFRKPI PSTVKAGELEKI **I SRCQVCMGTG**FLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLA
GSCLPVFSTL PFAYCNIHQVCHYAQRNDRSYWLASAAPLPMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSI PP
CPQTRWRLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFFANKYSFWLTTVKADLQFSSA
PAPDTLKEQAQRQKI SRCQVCV**APG**FVFTTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSLQR
FTTTPFLFCNVNDVCNFASRNDYSYWLSTPALMPMNMAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHG
WISLWKGFSSFIMFTSAGSEGTGQALASPGSCLEEFRASPFLECHGRGTCNYYSNSYSFWLASLNPERMFRKPI PSTV
KAGELEKI **I SRCQVCMKRRH***

>human_a545NC1

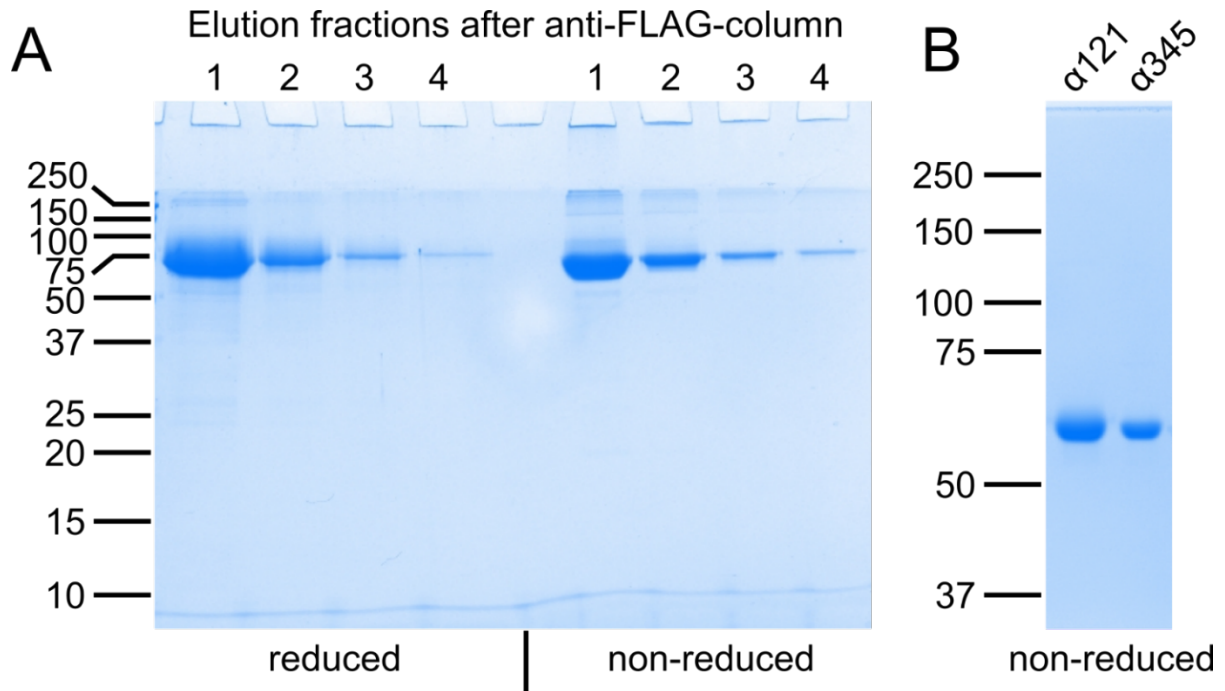
MRAWIFFLLCLAGRALAAPLA**DYKDDDDK**LAGTSSVAHGFLITRHSQTTDAPQCPQGTLOVYEGFSLLYVQGNKRAH
GQDLGTAGSCLRRFSTMPFMFCNINNVNCFASRNDYSYWLSTPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAV
HSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSQALASPGSCLEEFRASPFIECHGRGTCNYYANSYSFWLATVD
VSDMFSKQPSETLKAGDLRTRISRCQVCMGT**G**FLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLA
GSCLPVFSTL PFAYCNIHQVCHYAQRNDRSYWLASAAPLPMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSI PP
CPQTRWRLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFFANKYSFWLTTVKADLQFSSA
PAPDTLKEQAQRQKI SRCQVCV**APG**FLLITRHSQTTDAPQCPQGTLOVYEGFSLLYVQGNKRAHGQDLGTAGSCLRR
FSTMPFMFCNINNVNCFASRNDYSYWLSTPEPMPMSMQPLKGQSIQPFISRCAVCEAPAVVIAVHSQTIQIPHCPQG
WDSLWIGYSFMMHTSAGAEGSQALASPGSCLEEFRASPFIECHGRGTCNYYANSYSFWLATVDVSDMFSKQPSETL
KAGDLRTRISRCQVCMKRT*

Supplementary Fig. 1. Amino acid sequences of single-chain NC1 trimers of human collagen IV composed from α 3, α 4, and α 5 NC1 domains that were used in this study. BM40 signal peptide sequences are underlined and FLAG-tag sequences are highlighted. Artificially introduced linkers GTG and APG between chains are shown in bold. NC1 domain coloring is red for α 3, blue for α 4, and green for α 5.



Supplementary Fig. 2. Protein secretion analysis of different single-chain NC1 trimers using transient expression in ExpiCHO cells. Media (M) and cell lysate (C) were analyzed for protein presence using Western blotting with JK2 antibody. Only the single-chain $\alpha 345$ NC1 trimer was efficiently secreted into the medium. The $\alpha 543$ and $\alpha 345$ were not detected in the medium indicating potential folding problem. The $\alpha 545$ showed only limited secretion to the medium, also indicating potential folding problem. Co-expression of $\alpha 343$ and $\alpha 545$ did not resolve the folding problem, ultimately excluding these combinations as native forms. Finally, only the $\alpha 345$ trimer present a native form.

Supplementary Section 2: Crystal structure of the α 345 hexamer



Supplementary Fig. 3. (A) Purification of the single-chain α 345 NC1 trimer on anti-FLAG-column showing serial elution fractions with FLAG peptide. Molecular mass of non-reduced α 345 NC1 corresponds to a trimer, demonstrating that all disulfides are formed within the trimer. (B) Purified samples of single-chain α 121 and α 345 NC1 trimers after size-exclusion chromatography demonstrate consistent molecular masses under non-reducing condition.

Supplementary Table 1. Data collection and refinement statistics

Data collection statistics	
Wavelength (Å)	0.9786
Resolution range (Å) ^a	68.60-1.76 (1.79-1.76)
Space group	P4 ₁ 2 ₁ 2
Unit cell dimensions (Å)	a=b=128.4, c=104.71
Unit cell angles (°)	α=β=γ=90
Total reflections	711,994 (36,982)
Unique reflections	86,926 (4,534)
Redundancy	8.2 (8.2)
Completeness (%)	100.0 (100.0)
R _{meas} (%)	8.9 (52.1)
I/σI	13.1 (3.4)
Refinement statistics	
Resolution range (Å) ^a	45.40-1.76 (1.79-1.76)
Unique reflections used in refinement ^b	86,814 (5,993 ^c)
R _{work} (%)	13.9 (19.7)
R _{free} (%) ^d	16.2 (23.0)
Bond distance (Å) ^e	0.0059
Bond angle (°) ^e	0.8962
Chiral center (Å ³) ^e	0.0570
Planar group (Å) ^e	0.0061
Dihedral angle (°) ^e	18.7408
Number of atoms (excluding hydrogens)	6,429
Protein atoms	5,389
Chloride ions	6
Polyethylene glycol molecules/atoms	87/684
Water molecules	364
Mean B-value (Å ²)	30.5
Mean protein B-value (Å ²)	25.7
Mean chloride ion B-value (Å ²)	19.5
Mean polyethylene glycol molecule B-value (Å ²)	64.7
Mean water B-value (Å ²)	36.5
Ramachandran statistics	
Favored (%)	97.5
Additionally allowed (%)	2.5
Outliers (%)	0

^a Data for highest resolution shell are given in parentheses

^b Data cutoff $\sigma F=1.34$

^c In working set

^d 2.3% (2,000 reflections) of data excluded from refinement and used for R_{free}

^e Root mean square deviation

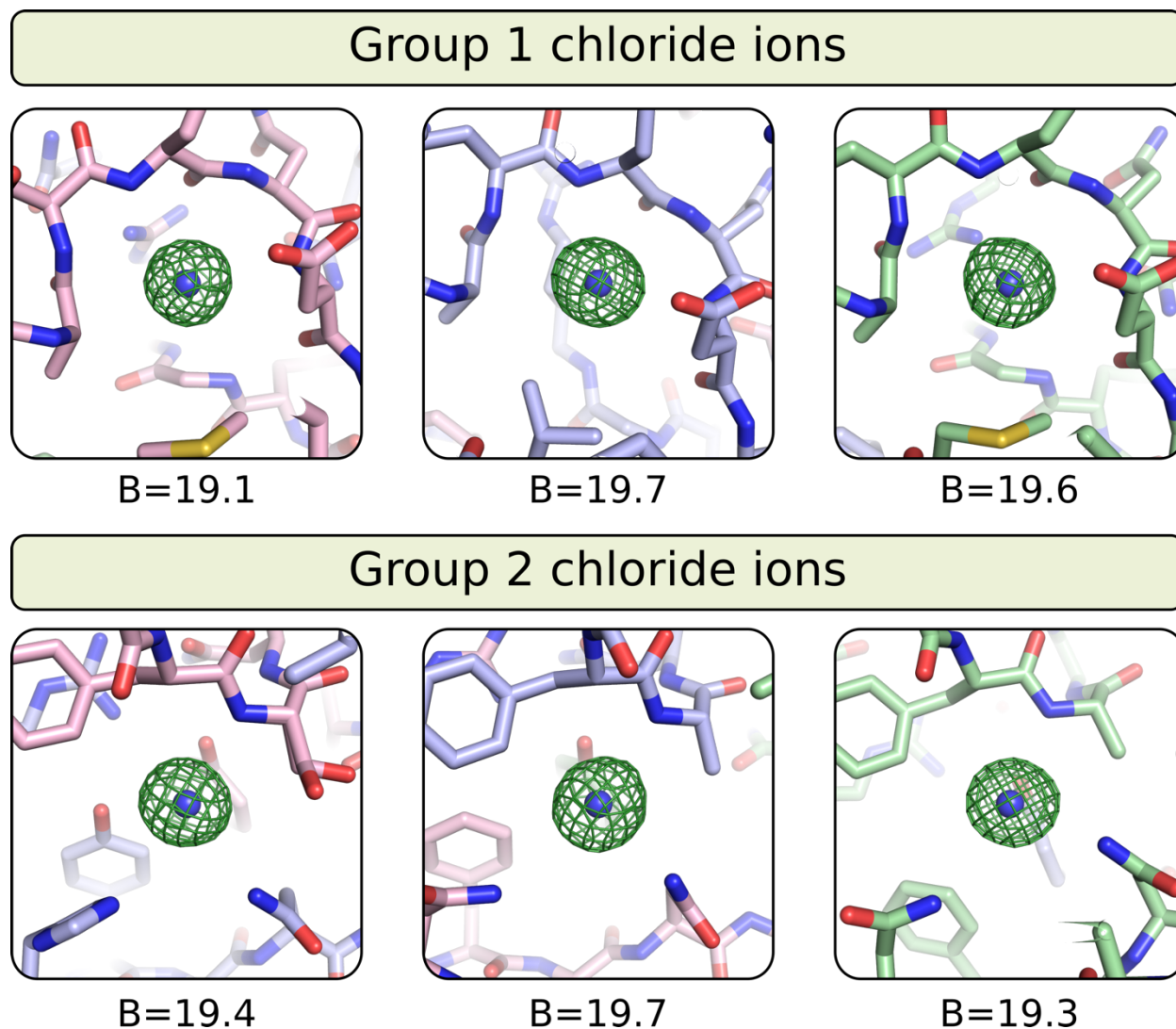
Supplementary Table 2. Structural comparison of α chains of the $\alpha 345$ crystal structure (present report, PDB ID 6wku) with the corresponding chains in the $\alpha 121$ structure (PDB ID 6mpx). Superimpositions were performed for C_{α} atoms. *-first $\alpha 1$ domain; #-last $\alpha 1$ domain

α chains	$\alpha 345$ vs $\alpha 121$	$\alpha 3$ vs $\alpha 1^*$	$\alpha 4$ vs $\alpha 2$	$\alpha 5$ vs $\alpha 1^{\#}$
r.m.s.d., Å	0.67	0.55	2.03	0.54

Supplementary Table 3. Structural comparison of NC1 α chains in homo-oligomeric complexes(1) with the corresponding structures in the $\alpha 345$ hexamer (present report, PDB ID 6wku). Superimposition were performed for C_{α} atoms.

α chain (PDB id)	$\alpha 3$ hexamer (5nb0)								$\alpha 4$ octamer (5nb1)						$\alpha 5$ hexamer (5naz)
chain ID	A	B	C	D	E	F	G	H	A	B	C	D	E	F	A
r.m.s.d., Å	0.58	0.60	0.59	0.58	0.57	0.57	0.58	0.55	4.73	4.71	4.50	4.66	4.71	4.74	0.58

Supplementary Section 3: The chloride ions at the NC1 α 345 trimer-trimer interface



Supplementary Figure 4. Electron density map of chloride ions. All six Cl^- ions (twelve per hexamer) have comparable densities and atom displacement values (B values, \AA^2). Ions are shown as **blue spheres**. The **dark green** meshwork represents the $F_o - F_c$ electron density map calculated before fitting chloride ions (contoured at $\sigma=3$). Carbon atom coloring in NC1 chains is **light red** for $\alpha 3$, **light blue** for $\alpha 4$, and **light green** for $\alpha 5$.

Supplementary Table 4. Coordination distances for two groups of chloride ions. Shown are distances between Cl⁻ ions and atoms involved in hydrogen bonding or ionic interactions (**bold**). Atoms that belong to the opposite trimer are marked with *. Water molecules coordinated by the opposite trimer are also marked with *, whereas water and polyethylene glycol molecules coordinated by both trimers are marked with #.

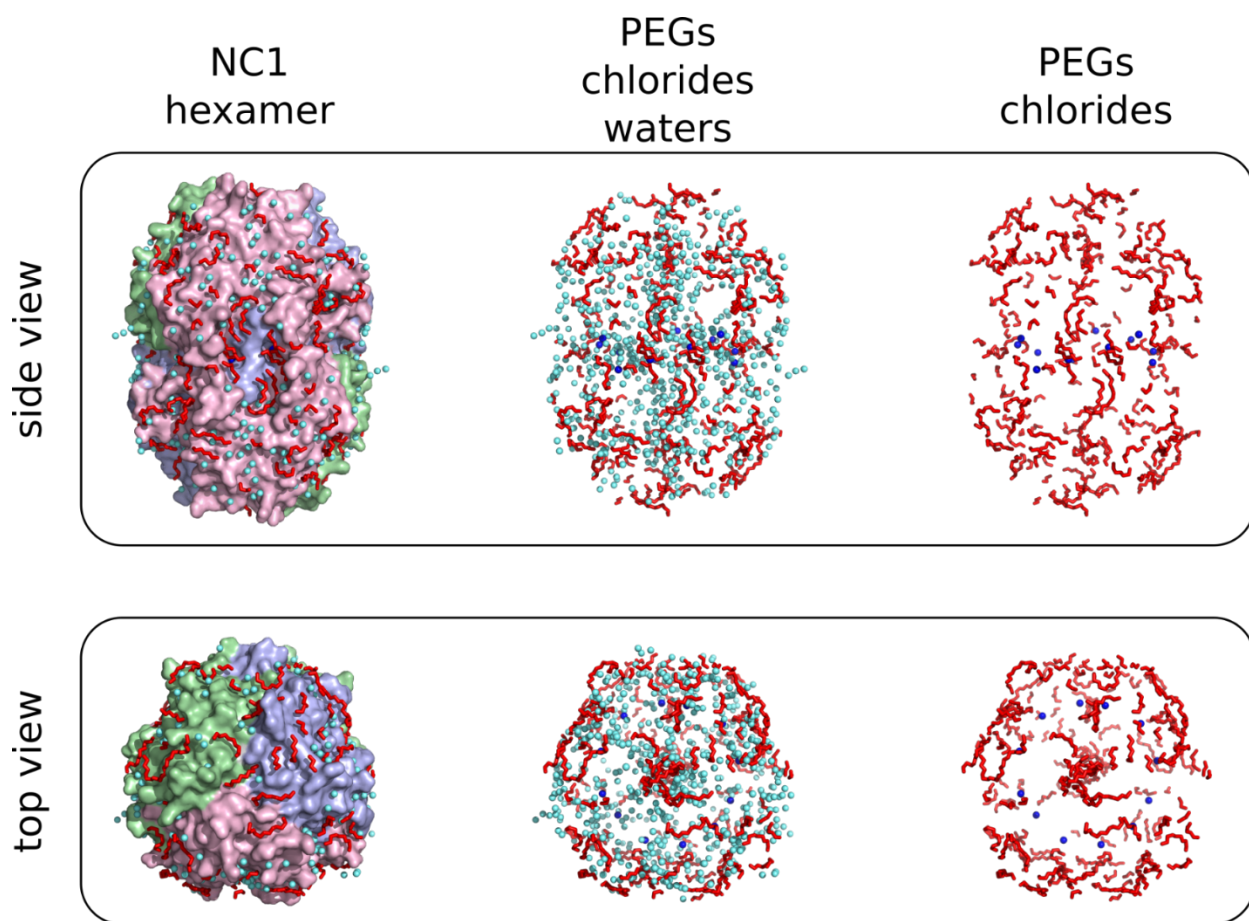
Group 1

Cl ⁻ #1		Cl ⁻ #2		Cl ⁻ #3	
A74, C α	3.7 Å	A298, C α	3.9 Å	A523, C α	3.6 Å
R76, N	3.4 Å	R300, N	3.2 Å	R525, N	3.5 Å
D78, N	3.3 Å	D302, N	3.4 Å	D527, N	3.3 Å
G178, C α	3.7 Å	L284, C δ 2	3.7 Å	G627, C α	3.7 Å
R179, Nη2	3.6 Å	R401, C γ	3.8 Å	*R628, Nη1	3.4 Å
		R401, C δ	3.7 Å		
H ₂ O, O	3.1 Å	H ₂ O, O	3.0 Å	H ₂ O, O	3.0 Å
#H ₂ O, O	3.4 Å	H ₂ O, O	3.1 Å		

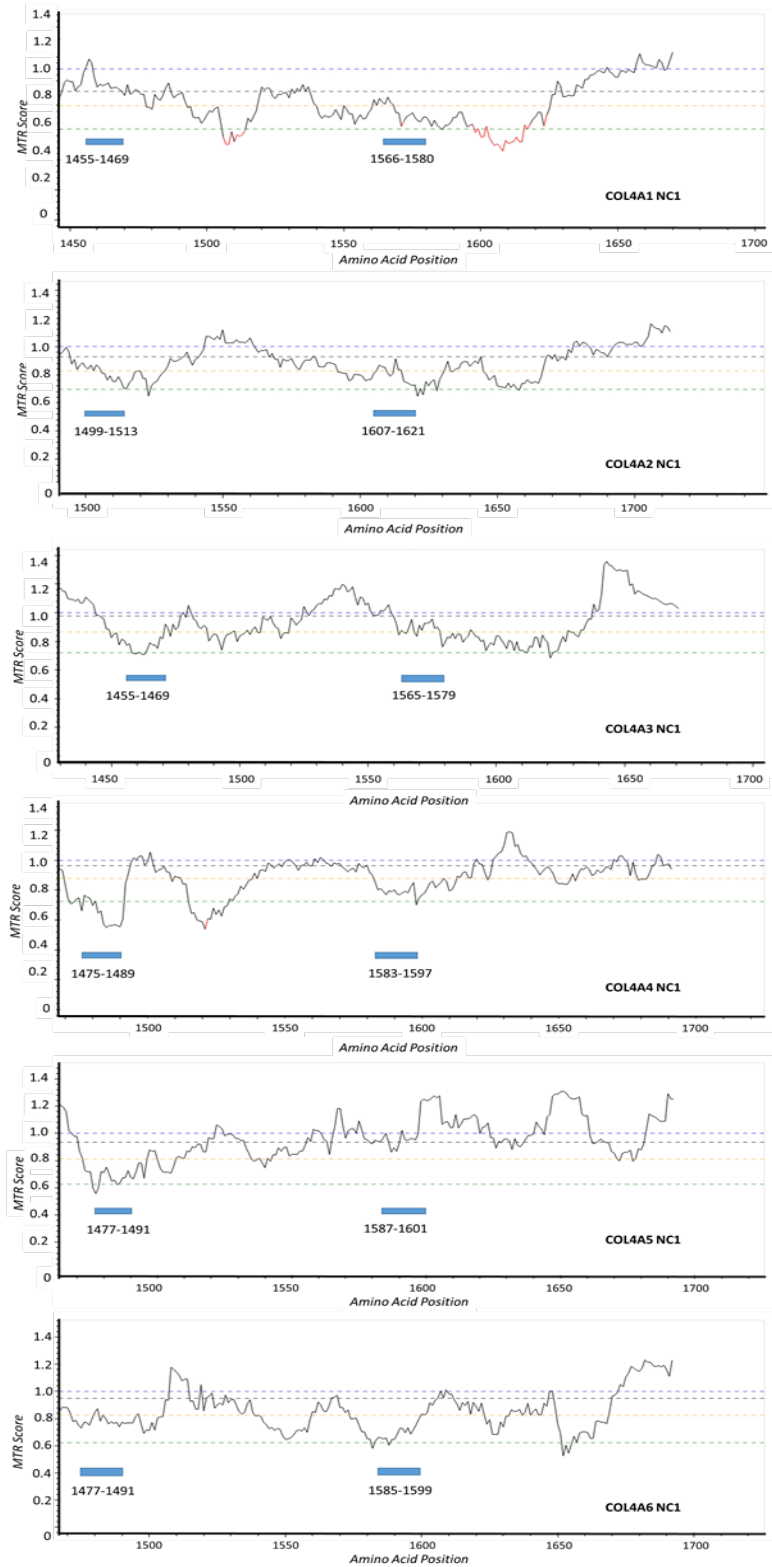
Group 2

Cl ⁻ #4		Cl ⁻ #5		Cl ⁻ #6	
Y185, C δ 1	4.0 Å	F408, C δ 1	3.6 Å	Y634, C δ 1	3.8 Å
Y185, C β	4.0 Å	F408, C β	3.9 Å	Y634, C α	3.8 Å
S186, N	3.2 Å	A409, N	3.2 Å	A635, N	3.2 Å
S186, C β	3.5 Å	A409, C β	3.6 Å	A635, C β	3.8 Å
*Y189, O η	3.2 Å	*Y638, O η	3.1 Å	*Y412, O η	3.2 Å
*Y288, O η	3.1 Å	*F64, C ϵ	3.8 Å	*F513, C ϵ	3.9 Å
*N290, N δ 2	3.4 Å	*N66, N δ 2	3.2 Å	*N515, N δ 2	3.3 Å
		*N66, C β	3.9 Å		
#PG4, C2	3.5 Å	#PGE, C1	3.6 Å	#PGE, C5	3.8 Å
#PG4, C1	3.7 Å	#PGE, C2	3.8 Å	#PGE, C6	4.0 Å
		*H ₂ O, O	3.3 Å	*H ₂ O, O	3.3 Å
				#H ₂ O, O	3.4 Å

Supplementary Section 4: Crevices, pockets, inner cavities, loops and surface-exposed reactive residues in the α 345 hexamer

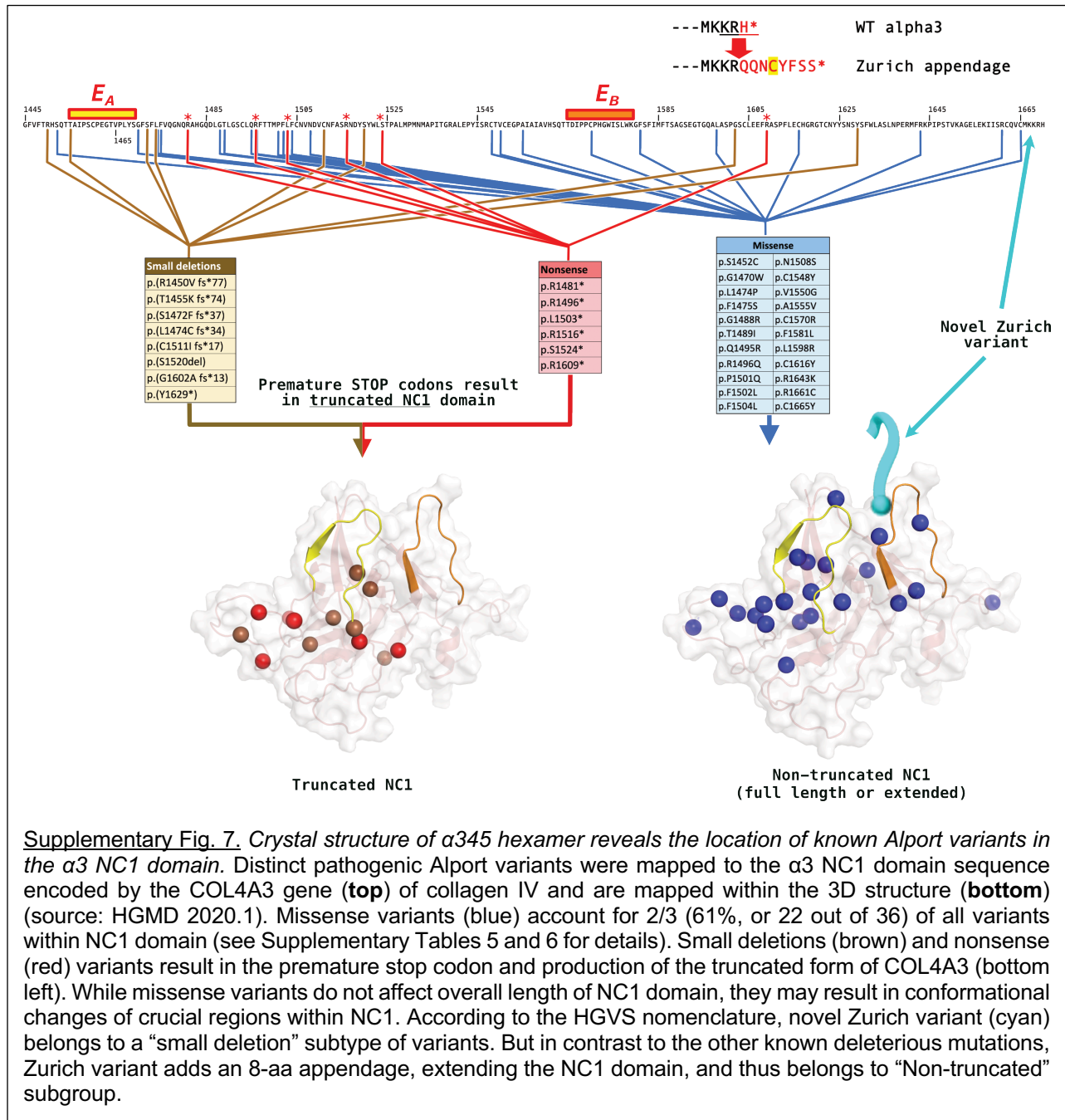


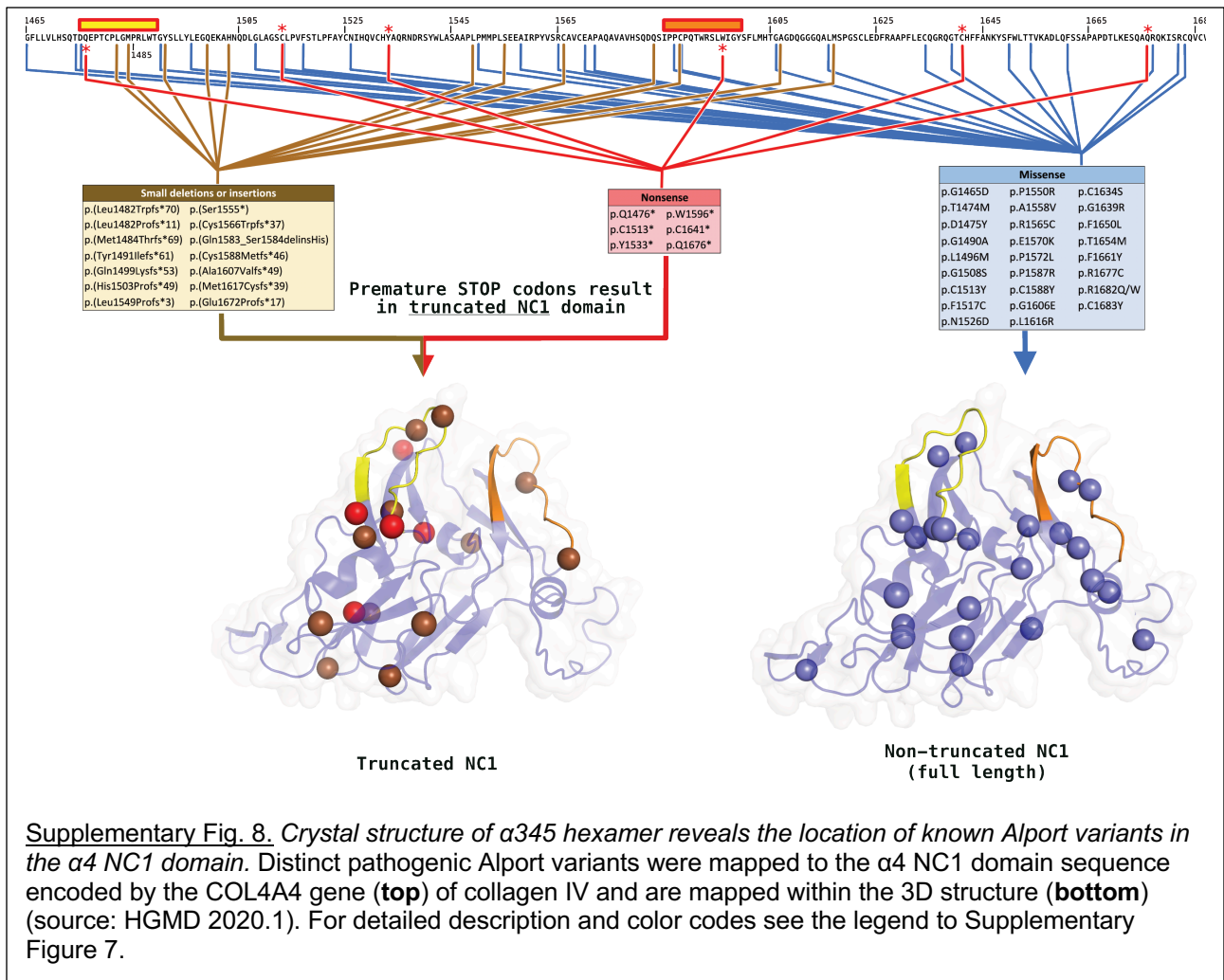
Supplementary Fig. 5. Structured polyethylene glycol and water molecules on the surface of the α 345 hexamer. Coloring of NC1 chains is **light red** for α 3, **light blue** for α 4, and **light green** for α 5. Chloride ions of both group 1 and group 2 are shown as **blue** spheres. Polyethylene glycol molecules are shown as **red** wireframes. Water molecules are shown as **cyan** spheres.

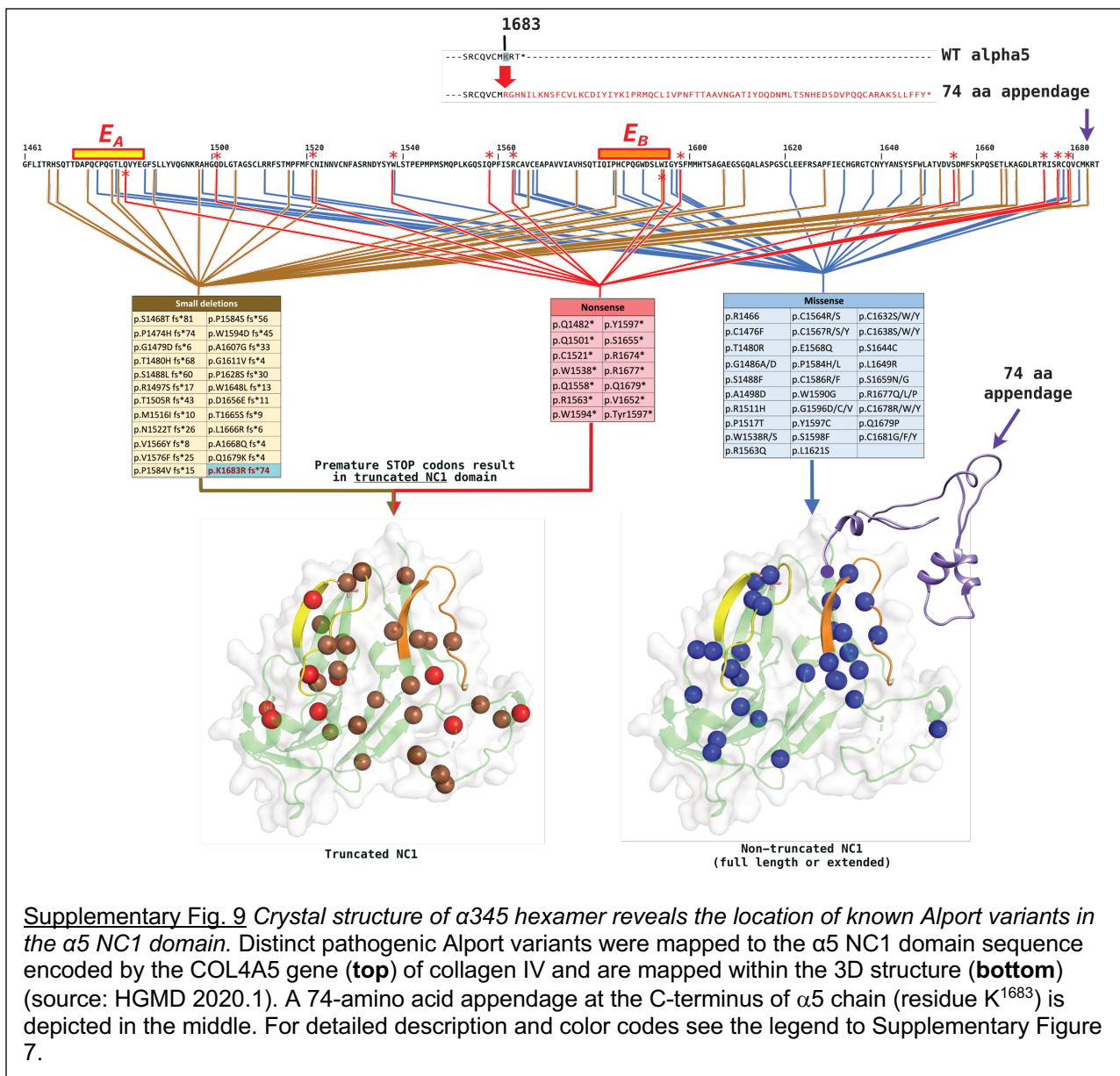


Supplementary Fig. 6. Missense Tolerance Ratio (MTR) plots for NC1 domains of α 1-6 chains of collagen IV generated with MTR-Viewer (21-codon window)(2). Horizontal dotted lines correspond, from top to bottom, to MTR=1 (no selective pressure) and gene-specific 50th, 25th and 5th MTR percentiles. Amino acid positions for E_A (left) and E_B (right) loops are indicated by blue boxes.

Supplementary Section 5. Analysis of known Alport variants in COL4A3, COL4A4, and COL4A5.







Supplementary Fig. 9 Crystal structure of $\alpha 345$ hexamer reveals the location of known Alport variants in the $\alpha 5$ NC1 domain. Distinct pathogenic Alport variants were mapped to the $\alpha 5$ NC1 domain sequence encoded by the COL4A5 gene (**top**) of collagen IV and are mapped within the 3D structure (**bottom**) (source: HGMD 2020.1). A 74-amino acid appendage at the C-terminus of $\alpha 5$ chain (residue K¹⁶⁸³) is depicted in the middle. For detailed description and color codes see the legend to Supplementary Figure 7.

Supplementary Table 5. Pathogenic variants of various types in the coding regions of COL4A3, COL4A4, and COL4A5 genes. Source: HGMD 2020.1

HGMD 2020.1

Mutation type	COL4A3	COL4A4	COL4A5	TOTAL
Missense/nonsense	180	164	557	901
Splicing substitutions	41	30	206	277
Regulatory substitutions	0	1	0	1
Small deletions	41	43	179	263
Small insertions/duplications	14	12	65	91
Small indels	2	2	11	15
Gross deletions	11	8	139	158
Gross insertions/duplications	3	2	6	11
Complex rearrangements	0	0	8	8
Repeat variations	0	0	0	0
TOTAL	292	262	1171	1725

Supplementary Table 6. Pathogenic variants in NC1 domains of COL4A3, COL4A4, and COL4A5. Source: HGMD 2020.1. Missense variants do not change overall length of the NC1 domains.

NC1 Domains			
Variant type	COL4A3	COL4A4	COL4A5
Missense/nonsense	22/6	29/6	49/14
Small deletions	8	11	18
Small insertions/duplications	-	3	7
Small indels	-	-	1
Terminal codon deletion (Zurich- and Zurich-like variants)	1	-	1
TOTAL	37	49	90

Supplementary References

1. Casino, P., Gozalbo-Rovira, R., Rodriguez-Diaz, J., Banerjee, S., Boutaud, A., Rubio, V., Hudson, B. G., Saus, J., Cervera, J., and Marina, A. (2018) Structures of collagen IV globular domains: insight into associated pathologies, folding and network assembly. *IUCrJ* **5**, 765-779
2. Silk, M., Petrovski, S., and Ascher, D. B. (2019) MTR-Viewer: identifying regions within genes under purifying selection. *Nucleic Acids Res* **47**, W121-W126