ZEBRAFISH TO HUMANS: EVOLUTION OF THE $\alpha_3$-CHAIN OF TYPE IV COLLAGEN AND EMERGENCE OF THE AUTOIMMUNE EPITOPES ASSOCIATED WITH GOODPASTURE’S SYNDROME

Brian A. MacDonald¹, Malin Sund¹, Marianne A. Grant², Kathleen L. Pfaff³, Kathryn Holthaus¹, Leonard I. Zon³ and Raghu Kalluri¹, ⁴, ⁵, ⁶

¹Center for Matrix Biology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, ²Division of Molecular and Vascular Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA and ³Stem Cell Program and Division of Hematology and Oncology, Children’s Hospital, Dana Farber Cancer Institute, Howard Hughes Medical Institute and Harvard Medical School, Boston, MA and ⁴Division of Nephrology, Children’s Hospital and Harvard Medical School, Boston, MA and ⁵Harvard-MIT Division of Health Sciences and Technology and ⁶Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA

Author contributions: RK, BM, MS, MG: designed research
BM, MS, MG, KH: performed research
KP, LZ: contributed vital new reagents
RK, BM, MS: analyzed data
RK, BM, MS: wrote the paper

Running title: Emergence of the Goodpasture’s epitope

Word count: 3983 (text) and 200 (abstract)

Research grants and financial support: This work was supported by NIH grants DK55001, DK62987, and funds from the Center for Matrix Biology at the Beth Israel Deaconess Medical Center. MS is funded by the Sigrid Juselius Foundation, the Maud Kuistila Foundation, the Finnish Medical Society Duodecim and the Emil Aaltonen Foundation. MG is funded by AHA grant 05030348N. LZ is supported by the NIH and the HHMI. KP is supported by the Albert J. Ryan Fellowship.

Address correspondence to:
Dr. Raghu Kalluri
Associate Professor of Medicine
Harvard Medical School
Director, Center for Matrix Biology, DANA 514
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Boston MA 02215
Tel: 617-667-0445
Fax: 617-975-5663
E-mail: rkalluri@bidmc.harvard.edu
MacDonald et al.  Emergence of the Goodpasture’s epitope

ABSTRACT

Goodpasture’s syndrome is an autoimmune vascular disease associated with kidney and lung failure, with pathogenic circulating autoantibodies targeted to a set of discontinuous epitope sequences within the non-collagenous domain-1 (NC1) of the α3 chain of type IV collagen [α3(IV)NC1], the Goodpasture’s autoantigen. We demonstrate that basement membrane extracted NC1 domain preparations from C. elegans, Drosophila melanogaster and Danio rerio do not bind Goodpasture’s autoantibodies, while Xenopus laevis, chicken, mouse and human α3(IV)NC1 domains bind autoantibodies. The α3(IV)-chain is not present in C. elegans and Drosophila melanogaster, but is first detected in the Danio rerio. Interestingly, native Danio rerio α3(IV)NC1 does not bind Goodpasture’s autoantibodies. Next, we cloned, sequenced and generated recombinant Danio rerio α3(IV)NC1 domain. In contrast to recombinant human α3(IV)NC1 domain, there was complete absence of autoantibody binding to recombinant Danio rerio α3(IV)NC1. 3D molecular modeling from existing x-ray co-ordinates of human NC1 domain suggest that evolutionary alteration of electrostatic charge and polarity due to the emergence of critical serine, aspartic acid and lysine residues, accompanied by the loss of asparagine and glutamine contributes to the emergence of the two major Goodpasture’s epitopes on the human α3(IV)NC1 domain, as it evolved from the Danio rerio over 450 million years.
INTRODUCTION

Type IV collagen is the most abundant protein within all basement membranes (BM)\(^1\). This collagen is present in most multi-cellular organisms, and thus is an evolutionary constant\(^2,3\). In the BM, type IV collagen forms a network, which other basement membrane molecules attach to or become sequestered within\(^1,4\). Type IV collagen in humans and mice is composed of six distinct \(\alpha\)-chain gene products, \(\alpha_1(IV)\)-\(\alpha_6(IV)\). The \(\alpha_1(IV)\) and \(\alpha_2(IV)\)-chains are ubiquitously present in most organs, while the \(\alpha_3(IV)\)-\(\alpha_6(IV)\)-chains exhibit more restricted distributions and are present mostly in BMs with special functional roles, such as the kidney glomerular basement membrane (GBM)\(^1,5-7\).

The \(\alpha_3\)-chain of type IV collagen is found in the kidney GBM, the alveolar BM of the lung, the basement membranes of the inner ear and the testis\(^1,8\) and is present in most mammals\(^9\). Mice deficient in the \(\alpha_3(IV)\)-chain die due to renal failure as this protein is required for the proper function of the GBM\(^10,11\). This phenotype resembles human Alport syndrome, a disease affecting kidneys and the inner ear, which results from mutations in the \(\alpha_3\)-chain, the \(\alpha_4\)-chain or the \(\alpha_5\)-chain of type IV collagen\(^8,10\). Pathogenic autoantibodies directed against the NC1 domain of the \(\alpha_3\)-chain of type IV collagen (Goodpasture’s autoantigen) are associated with human Goodpasture’s syndrome (GP), a disease characterized by rapidly progressive glomerulonephritis and lung hemorrhage\(^8,12-13\). These autoantibodies are targeted to specific amino acid sequences within the NC1 domain of the \(\alpha_3(IV)\)-chain\(^8,14\). By using mutagenesis analysis, it is known that a majority of the autoantibodies are directed against two distinct epitopes on the \(\alpha_3(IV)\)NC1 domain, namely epitopes \(E_A\) and \(E_B\)\(^15-18\). Epitope \(E_A\) appears to be the immunodominant epitope, as the majority of patients’ antibodies are directed against this site\(^18\). While the production of IgG1 and IgG3 subclasses of \(\alpha_3(IV)\) autoantibodies is considered to be the pathogenic feature of Goodpasture’s syndrome, there is also strong evidence now for the role of T-cells in the initiation
of the disease\textsuperscript{19-26}. Therefore, in the past few years, Goodpasture’s Syndrome has emerged as a classic autoimmune vascular disease mediated by B-cells\textsuperscript{27-29} and T-cells\textsuperscript{19-26}, and immunosuppression in conjunction with plasmapheresis remains the most effective therapy\textsuperscript{30}.

Since the vast majority of autoantibodies bind to the immunodominant epitope $E_A$ and epitope $E_B$, several studies have sought to elucidate the key residues in these sites. Such studies have been primarily focused on epitope $E_A$ because this site is capable of inducing disease in rodents (14). The method most commonly employed has been to recombinantly produce chimeric proteins\textsuperscript{15-18,31-32}. Many of these chimeric proteins are comprised of amino acids of the epitope $E_A$ from the $\alpha_3(IV)$ NC1 domain with site-specific substitutions of analogous residues from the $\alpha_1(IV)NC1$ domain\textsuperscript{17,31-32}. Such proteins were subsequently tested for their binding to autoantibodies from Goodpasture’s patients, thus identifying putative critical amino acids that comprise the Goodpasture’s epitope.

While others have employed site directed mutagenesis and chimeric protein analysis as a way to identify the critical amino acids within the Goodpasture’s epitopes, we hypothesized that evolutionary changes over 450 million years\textsuperscript{33} in the amino acids within the $\alpha_3(IV)NC1$ domains might offer insight into the unique amino acid residues that comprise the B-cell epitope for Goodpasture’s autoantibodies. Comparative type IV collagen NC1 sequence analysis between nine species from \textit{C. elegans} to humans, BM extraction, NC1 domain analysis from different species, cloning and recombinant protein production were used to identify the critical amino acids within the B-cell epitopes of the Goodpasture’s autoantigen. This study provides a strategy to identify critical amino acids responsible for Goodpasture’s autoantibody binding and offers new insights into potential therapeutic alternatives. A precise identification of amino acids responsible for autoimmunity will help design cell-mediated therapies to potentially impart systemic tolerance to pathogenic epitopes associated with this devastating disease.
MATERIALS AND METHODS

Animals and Tissues - Wild type C57BL/6 mouse kidney, wild type Danio rerio strain AB* kidney, and wild type domestic chicken kidney (Pel-Freez Clinical Systems) were used for all procedures. Wild type Xenopus laevis kidneys were the kind gift of Dr. Sergei Sokol at the Beth Israel Deaconess Medical Center/ Harvard Medical School. Wild type strains CS and W^{118} Drosophila melanogaster were the kind gifts of Dr. Mel Feany at the Brigham and Women’s Hospital/ Harvard Medical School. Wild type C. elegans, strain NL, were the kind gift of Dr. Monica Colaiacovo, Department of Genetics, Harvard Medical School.

Cloning, Primers, and Sequencing - Total RNA was isolated from 24 hours post-fertilization (hpf) Danio rerio embryos using TRIzol™ reagent (Invitrogen). A collection of cDNAs was created using Superscript II™ reverse transcriptase (Invitrogen) and oligo dT primers (Invitrogen). Polymerase chain reaction (PCR) amplification of Danio rerio α3(IV)NC1 domain was performed using AccuPrime™ Pfx DNA polymerase (Invitrogen) and the specific primers 5’ CACCAGGTGCAAAAGGTCCACAAC 3’ and 5’ ACGATTTTGTGTCAGGGTCAGAA 3’, designed from sequence data obtained from the Project Ensembl database and software system, maintained by the EMBL-EBI and the Sanger Institute. The resulting cDNA clones were subcloned into the pCR® II 4.0-TOPO® vector (Invitrogen), and sequenced by the Beth Israel Deaconess Medical Center sequencing facility using the M13(-20) forward and M13 reverse primers. Sequences were analyzed using the Mac Vector 6.0 (Oxford Molecular Group) and BLAST programs (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/). The cloned sequences were deposited in GenBank under the accession number: AY954909.1.
**Recombinant protein production** - The following PCR primers were designed to amplify a cDNA encoding the full-length Danio rerio α3(IV)NC1 which we have termed zα3(IV)NC1; 5' TCA CCA GGT GCA AAG CTT CCA CAA 3' and 5' TCT GTT CTC GAG TTT GGC GGT TTG 3'. The underlined sequence indicates the Hind III and Xho I restriction sites that were used for subsequent cloning of the generated cDNA fragment. The Hind III/Xho I fragment of the amplified product was cloned into the pcBFT expression vector containing the CMV promoter, the BM-40 signal peptide and a FLAG tag\(^3\). 293 human embryonic kidney cells were transfected with the plasmid using the 293 Fectin\(^\text{®}\)-system (Invitrogen) according to manufacturer’s protocol. Recombinant zα3(IV)NC1 was secreted into the culture supernatant as a fusion protein with the FLAG epitope. Supernatant containing protein was collected 24 hours post-transfection. Recombinant human α3(IV)NC1 production was described earlier\(^3\).

**Histology and Immunofluorescence** - Adult *Danio rerio* were embedded in Tissue-Tek\(^\text{®}\) O.C.T. compound (Sakura Finetek) and snap frozen on liquid nitrogen, as well as fixed in formalin and embedded in paraffin for further histological analysis. Paraffin sections (5 μm) were stained with hematoxylin and eosin according to routine histological practice. Frozen sections (5 μm) were fixed with acetone, blocked with 1% BSA in PBS, incubated with the anti-T7 antibody (raised in rabbit against the T7 peptide of human α3(IV)NC1 domain)\(^3\) at 1:75 dilution, and the anti-human α3(IV)-36mer antibody at 1:200 dilution\(^3\), followed by an anti-rabbit FITC conjugated secondary antibody at 1:100 dilution (Jackson Immunoresearch). The sections were mounted using the Vectashield mounting media containing DAPI for nuclear staining (Vector Laboratories). Histological images were captured using a Zeiss brightfield and fluorescence microscope (Zeiss).
Isolation of Native Type IV Collagen NC1 Domains - Kidneys dissected from mouse, chicken, *Xenopus laevis* and *Danio rerio* were snap frozen in liquid nitrogen followed by homogenization by mortar and pestle. In the case of *Drosophila* and *C. elegans*, the whole organisms were subjected to homogenization. Tissue homogenates were centrifuged and resuspended in phosphate-buffered-saline (PBS) containing Complete Protease Inhibitor Cocktail® (Roche), thoroughly homogenized again, centrifuged, and resuspended in 1M NaCl supplemented with DNase I (80 μg/ml). Following a third centrifugation, samples were resuspended in 2% deoxycholate/ protease inhibitor solution and centrifuged a final time. Samples were resuspended in bacterial collagenase solution (1 unit/μl) (Worthington Biochemical) to digest proteins with collagenous regions, in order to liberate soluble NC1 domains of type IV collagen.

Western Blotting and antibodies - All western blots were performed on either 10% or 15% SDS polyacrylamide gels (SDS-PAGE) (Bio-Rad) under reducing or non-reducing conditions. Gels were transferred to nitrocellulose and incubated with sera, primary and secondary antibodies at the appropriate dilutions. The following primary antibodies were used; anti-α3(IV)-36mer antibody at 1:7500 dilution32, two separate α3(IV)NC1 antibodies27, three confirmed Goodpasture’s patient sera at 1:20 dilution36 and anti-FLAG M2 antibody at 1:1000 dilution (Sigma). In addition, two normal human sera were used as negative controls at the same dilutions as the Goodpasture’s sera (data not shown). Anti-human IgG and anti-rabbit IgG HRP (Sigma) conjugated secondary antibodies were used at 1:1000 dilutions. ECL reagent for peroxidase substrate (Amersham Biosciences) was used to visualize antibody binding by exposure to x-ray film (Denville).

Molecular Graphics – The structures of the [(α1)2α2]3 NC1 hexamers from human placenta basement membranes (Protein Data Bank ID 1LI1)37 were rendered using the molecular graphics
visualization program YASARA (YASARA Biosciences) and analyzed to generate modeled structures using the molecular graphics software InsightII (Accelrys, Inc.).
RESULTS AND DISCUSSION

The emergence of the Goodpasture’s epitope - To study the binding of Goodpasture’s autoantibodies, we extracted NC1 domains from BM extracts using bacterial collagenase from the following species: mouse, chicken, *Xenopus laevis*, *Danio rerio*, *Drosophila* and *C. elegans*. For the higher vertebrate organisms, kidney tissue was used as a source of BM, whereas for *Drosophila* and *C. elegans* whole organism BM extracts were utilized. Type IV collagen NC1 domains were successfully extracted from all organisms as shown by SDS-PAGE analysis and Western blot evaluation under reducing conditions (Fig. 1A-B). When analyzing Goodpasture’s autoantibody reactivity to the isolated BM extracts under non-reducing conditions, positive binding of sera to α3(IV)-chain dimers and monomers could be observed in mouse, chicken and *Xenopus* (Fig. 1C). Interestingly, binding to dimers was stronger than to monomers, as others have described9. (Fig. 1C). The Goodpasture’s epitope is considered to be sensitive to conformational changes and thus complete reduction of disulfide bonds β-mercaptoethanol leads to a total loss of antibody binding (Fig. 1E), while incomplete reduction with leads to stronger binding to α3(IV)NC1 monomers in all species (Fig. 1D). Reactivity with Goodpasture’s sera could not be observed with NC1 domains extracted from *Danio rerio*, *C. elegans* and *Drosophila* (Fig. 1C). In *Drosophila* and *C. elegans* both α1(IV)-like and α2(IV)-like genes have been described38-40, but our genomic database analysis and cloning attempts suggest that an α3(IV) gene does not appear to be present. Interestingly, our database search indicates that *Danio rerio* has six type IV collagen α-chains similar to higher organisms (MacDonald, B. et al.: unpublished data). However, while the *Danio rerio* α3(IV) gene was identified in our database search, the native NC1 domain extracts do not bind to Goodpasture’s autoantibodies (Fig. 1C).
Cloning, sequencing and recombinant expression of the Danio rerio α3(IV)-chain NC1 domain - Our results suggest that while the α3(IV)-chain is present in the Danio rerio (by inference from genomic database sequences), the NC1 domain extract from BM preparations of Danio rerio renal tissue does not bind to Goodpasture’s autoantibodies. To further evaluate the capacity of the Danio rerio α3(IV)NC1 domain to bind these antibodies, we attempted to clone and generate recombinant protein.

In order to clone the Danio rerio α3(IV)NC1 domain (zα3(IV)NC1), we isolated total RNA from Danio rerio embryos and generated a collection of cDNAs by reverse transcription (RT), using oligo dTs as primers for the reaction. These cDNAs were subsequently used as a template to obtain the zα3(IV)NC1 domain by PCR. Gene specific PCR primers were designed using sequence data information obtained from the project Ensembl database. We successfully generated a 821 bp cDNA fragment, and sequencing verified that the cloned cDNA fragment contained 23 bp of the collagenous region, 702 bp containing the entire NC1 domain, and 96 bp of non-coding sequence. The nucleotide and amino acid sequence of zα3(IV)NC1 is shown in Fig. 2A. The homology of the five exons encoding the overall nucleotide and amino acid sequence of the zα3(IV)NC1, when compared to human α3(IV)NC1, is 60% and 74%, respectively. The twelve cysteines known to be essential for the appropriate conformational folding of the type IV collagen NC1 structure41 are all conserved in zα3(IV)NC1 (Fig. 2A). Interestingly, a methionine and a lysine residue, potentially involved in a novel non-disulfide covalent crosslink37 are also conserved in the zα3(IV)NC1 (Fig. 2A).

The cloned cDNA fragment was subsequently transferred to the pcBFT expression vector and transfected into 293 human embryonic kidney cells to produce recombinant zα3(IV)NC1 using a BM-40 signal peptide to enable extracellular secretion. Recombinant FLAG-tagged zα3(IV)NC1 fusion protein in the culture supernatants was analyzed by SDS-PAGE. Western blot analysis using antibodies directed against the FLAG-tag and the human α3(IV)NC1, verified the
recombinant expression of zα3(IV)NC1 monomers at the predicted molecular weight of approximately 29.4 kDa (zα3(IV)NC1+FLAG-tag) (Fig. 2B). However, zα3(IV)NC1 did not bind to Goodpasture’s autoantibodies from three different patients with clinical disease (Fig. 2C), despite prolonged incubation with high concentrations of the autoantibodies to detect even weak binding. However, as shown earlier, strong binding of anti-human α3(IV)NC1 polyclonal antibodies to isolated native zNC1 domain monomers, dimers and trimers from BM extracts were observed (Fig. 2C). Under reducing conditions the higher molecular weight trimers were lost and monomers of Danio rerio α3(IV)NC1 revealed stronger binding (Fig. 2C). However, binding with the Goodpasture’s sera to the isolated Danio rerio NC1 domains could be not observed, as shown earlier (Fig. 2D). Together, these results demonstrate that although Danio rerio contains the encoding gene and the α3-chain of type IV collagen protein, the protein does not possess the capacity to bind Goodpasture’s autoantibodies.

Expression of the α 3(IV)-chain in Danio rerio tissues - Since this report is the first description of the α3-chain of type IV collagen in the Danio rerio, and human α3(IV)-chain is localized to the GBM of the kidney and the alveolar BMs of the lung, we performed immunohistochemistry to evaluate the expression of the zα3(IV)NC1 in gills and kidney tissue of Danio rerio. Immunohistochemical localization of the zα3(IV)-chain in tissues was performed using two separate antibodies directed against the human α3(IV)NC1 domain. Identical results were obtained with both antibodies (Fig. 3 and data not shown). The α3(IV)-chain in Danio rerio is strongly expressed in the gills and the kidney (Fig. 3B-D). Gills are the fish equivalent of lungs, and the lungs and kidney are the sites of high expression of the α3(IV)-chain in mouse and humans1,5-7. On the other hand, we could not observe any expression of the α3(IV)-chain in Danio rerio testis (data not shown), although in mammals it is a site with high α3(IV) expression1,5-7,42. In the gills the α3(IV)-chain appeared to be expressed in the BM underlying the
epithelial cell layer (Fig. 3A-B). In the *Danio rerio*, the kidney morphology is distinct from that of higher organisms as it is also the site of hematopoiesis, in addition to its function in ultrafiltration (Fig. 3C). In *Danio rerio* kidney the α3(IV)-chain is expressed in both the BM of the glomeruli as well as some tubules (Fig. 3D). Interestingly, we could also observe strong staining of the α3(IV)-chain in the hematopoietic tissue (Fig. 3D). In higher organisms, the α3(IV)-chain is an important component of the GBM in the kidney and is also expressed in some tubular BMs 6. However, the potential role of this molecule in hematopoiesis is unknown. As observed in humans and other mammals, we were unable to detect αα3(IV)NC1 in liver, and stomach basement membranes.

**Comparison of the Goodpasture’s epitope sequences within the α3(IV)NC1 among different organisms** - Goodpasture’s autoantibodies bind to human and chicken α3(IV)NC1 domains, but do not bind to the *Danio rerio* α3(IV)NC1 domain. In order to explain observed differences in Goodpasture’s autoantibody binding, we compared amino acid sequences at the two putative B-cell Goodpasture’s epitope regions, EA and EB, among different species (Fig. 4A-B). The *Danio rerio* and chicken sequence analysis reveals that while most amino acids are conserved, some critical changes can be noticed, which are further preserved in the human sequence. Our results demonstrate that Goodpasture’s antibody binding at epitope EA on the human α3(IV)NC1 domain requires conservation of threonine-17 (T17) and the presence of serine-21 (S21). Binding at epitope EB requires threonine-127 (T127), loss of asparagine-128 (N128), substitution of proline or alanine for glutamine-131(Q131), and lysine-141 (K141). This analysis is the first to show that T127 and K141 are important residues for antibody binding at Goodpasture’s epitope EB.

Interestingly, while many species are positive for Goodpasture’s autoantibody binding, they do not contain all of the amino acids previously determined to be critical for such activity by
various chimeric protein and site directed mutagenesis studies. Amino acids thought to be crucial for binding at the epitope $E_A$ are T17, A18, I19, P20, V27, P28 and S31. Amongst all of the amino acids determined to be critical utilizing a chimeric protein approach, our results are in agreement with T17 of epitope $E_A$. Previous work demonstrates that although necessary, this threonine is not sufficient for autoantibody recognition. Previous work also reinforces the importance of S21. By altering this residue in a chimeric construct, autoantibody recognition by 100% of patient serum samples decreases to 25%. At epitope $E_B$, similar mutational analysis has not been described.

All of the $\alpha 3(IV)$ sequences, whether positive or negative for Goodpasture’s sera binding, share the following sequence homology in epitope $E_A$: TxxPxCPxGTxxLYx, (Fig. 4A and 5A). Cross-species conservation indicates that this sequence motif is evolutionarily preserved for structure or function. A similar conserved sequence motif, TxxPxCPxGWxSLWx (Fig. 4B), exists in epitope $E_B$, regardless of Goodpasture’s antibody binding. The only exceptions to this sequence conservation are isoleucine (I) substituted for the N-terminal threonine (T) in *Danio rerio*, and aspartate (D) substituted for the middle glycine (G) in mouse. Consistent with epitope $E_A$, conservation of the $E_B$ sequence motif indicates evolutionary preservation of the sequence for NC1 domain structure or function, and underscores the importance of the intervening non-conserved residues (x) within these conserved sequence motifs for antigenic recognition by Goodpasture’s autoantibodies directed against epitope $E_A$ and $E_B$. The conserved $E_B$ sequence is distinguished from $E_A$ by tryptophan (W) residues (W136 and W140), rather than threonine (T) and tyrosine (Y), and one additionally conserved C-terminal serine residue (S138).

**Molecular modeling for 3D analysis of the human, chicken and *Danio rerio* $\alpha 3(IV)$NC1 domains** - In order to spatially evaluate charge, hydropathy and side-chain polarity of the epitope $E_A$ and $E_B$ sequence in the context of the entire NC1 domain, we utilized molecular modeling software and the existing human type IV collagen NC1 domain crystal structure to generate
models (Fig. 5A-G). From crystallographic data it has been shown that the α1(IV)NC1 and α2(IV)NC1-chains are folded virtually identically, and their topologies very closely resemble one another[37]. Based on the high sequence identity between all type IV collagen NC1 domains, the α3(IV)NC1 domain is expected to fold with the same topology as the α1(IV)NC1[37]. Therefore, by aligning the α3(IV)NC1 and α1(IV)NC1 sequences and using the known [(α1)2α2]2 NC1 hexamer structure (Protein Data Base ID 1LI1), we have modeled amino acids that are divergent in the Danio rerio, chicken and human E_A and E_B epitope regions on a backbone homology α3(IV)NC1 structure (Fig. 5B-D). Amino acid differences in epitope E_A (blue) and E_B (green) regions in chicken compared to Danio rerio are indicated by side chain rendering in yellow, while further amino acid differences in human compared to chicken are shown by side chain rendering in red (Fig. 5E-G).

We found that in the context of the three-dimensional α3(IV) NC1 domain homology structures, residues that we identified as important in Goodpasture’s autoantibody binding at epitopes E_A and E_B are localized to the outer loops of the epitopes (Fig. 5E-G). In addition, many of these residue changes alter electrostatic potentials in these solvent exposed epitope regions. For example, S21 and K141, described above as important for epitope E_A and E_B binding, respectively, are glutamate residues in the immuno-negative Danio rerio.

The 3D homology structures enabled us to assess conserved α3(IV) NC1 residues immediately proximal to epitope E_A. Glutamine 57 (Q57) is conserved in all immuno-positive proteins, but absent from all immuno-negative proteins, including Danio rerio α3(IV) NC1. Surprisingly, although 26 residues away from epitope E_A in the tertiary NC1 domain structure, Q57 is immediately adjacent to epitope E_A (Fig. 5F,G; rendered in magenta) and provides an additional strongly hydrophobic residue. Previous work demonstrates Q57 is as important as S21 for autoantibody binding to E_A. Specifically, alteration of Q57 results in 80% loss of patient sera binding[17] to this discontinuous and conformational immunodominant epitope.
Our combined findings prompted us to map the amino acid character and molecular surface of species-variable residues (non-conserved) onto the modeled *Danio rerio*, chicken and human E_A and E_B epitope regions of our backbone homology α3(IV)NC1 structures (Fig. 5B-D). Using color-coded amino acid character maps, it is readily apparent that in both E_A and E_B electrostatic charge and polarity changes are important in Goodpasture’s autoantibody binding. A significant loss of electrostatic potential (positive/basic residue character) accompanies autoantibody recognition in the E_A epitope region. The adjacent K27, R28 residue pair in *Danio rerio* is either a hydrophilic/polar residue as in chicken or hydrophobic/non-polar residue as in human. Overall, the immuno-positive chicken and human E_A domains have reduced electrostatic potential and greater hydrophilic character, as compared to the homologous region in *Danio rerio* (Fig. 5B-D). In contrast, the gain of a positive/basic residue in the epitope E_B C-terminus is important for autoantibody binding (Fig. 5A-G). K141 is one of two residue changes in epitope E_B in chicken compared to *Danio rerio* that is present in human. The other residue change in epitope E_B in chicken compared to *Danio rerio* that is also in human is T127. This substitution decreases the non-polar character of the epitope E_B N-terminus. In summary, these comparisons demonstrate that evolutionary alterations of electrostatic charge and polarity contribute to Goodpasture’s autoantibody binding in the α3(IV)NC1 E_A and E_B epitopes.

Interestingly, we found that the high hydrophobic character of both the E_A and E_B domains is largely conserved from *Danio rerio* to chicken to human. In previous studies, both V27 and P28 were shown to be important for antibody binding and critical residues of the E_A autoepitope, through mutagenic and chimeric analysis between α1(IV)NC1 and α3(IV)NC1 sequences (in human α1(IV) these residues are lysine and isoleucine). Examination of the corresponding residues in chicken, serine and glutamine, lessens the idea that V27 and P28 are important in epitope E_A for hydrophobic contributions, and suggests that the reduced antibody binding observed in mutagenic/chimeric analyses possibly reflects the effects of basic charge...
insertion and side-chain steric clash (Fig. 5A-D). Of note, one variable (non-conserved) hydrophobic residue at position 131 appears important for the E_B antibody recognition site (Fig. 5A-D). In *Danio rerio* a glutamine occurs at this position, while there is a transition to alanine in chicken and proline in human. Interestingly, a non-polar/hydrophobic residue type was conserved at position 137 (isoleucine/leucine/valine) in epitope E_B between *Danio rerio*, chicken and human. The immuno-negative sequences of bovine α1(IV)NC1, and *Drosophila* and *C. elegans* α1(IV)NC1 and α2(IV)NC1 contain glutamate or serine at position 137. This finding suggests that a non-polar/hydrophobic residue at position 137 may be involved in autoantibody binding, but is not sufficient in the context of the *Danio rerio* E_B region to impart antibody binding.

Our results demonstrate that there has been an evolutionary pressure to obtain the α3-chain of type IV collagen in the *Danio rerio*. This could be linked to the required homeostasis and physiology of vertebrates to exercise greater regulation of fluid and gas exchanges. Interestingly, the Goodpasture’s epitope is not present within the α3(IV)NC1 domain of *Danio rerio*. The emergence of this epitope (in frogs) coincides with the transfer from aquatic to terrestrial living. Lastly, our study demonstrates the importance of *Danio rerio* in elucidating the evolutionary changes within the α3(IV)NC1 domain, being the only species to possess an α3 chain of type IV collagen not recognized by Goodpasture’s autoantibodies. This insight into the critical amino acids required for binding autoantibodies will potentially lead to cell-mediated therapies to impart systemic tolerance to the pathogenic epitopes which cause this devastating disease. Future studies will hopefully shed more light on the physiological need for such modification in the α3-chain of type IV collagen.
REFERENCES


ACKNOWLEDGEMENTS

We would like to thank Dr. Ramani Ramchandran for his initial help in searching the *Danio rerio* genomic databases for type IV collagen genes.
FIGURE LEGENDS

FIGURE 1. Extraction of basement membrane, bacterial collagenase solubilization and the analysis of native type IV collagen NC1 domains from different species. A. Native protein (60μg per lane) from the six different species and 250 ng of recombinant human α3(IV)NC1, (rh-α3(IV)NC1)\(^{35}\) was run on a 15% SDS-PAGE under extreme reducing (10% β-mercaptoethanol) conditions, and incubated with the anti-36mer antibody\(^{23}\). Monomer forms (M) of NC1s are detected in all lanes. Although subjected to extreme reduction, significant amounts of the isolated NC1 remain as dimers (D). B. *C. elegans* NC1 in dimeric form (D) was visualized after increased film exposure. C. The same samples were run on a 15% SDS-PAGE under non-reducing conditions and incubated with three different Goodpasture’s sera (1:20 dilution). A representative blot is shown. As expected, sera from Goodpasture’s patients bound recombinant human α3(IV)NC1 (lane 1). Mouse, chicken, and frog kidney BM extracts show strong reactivity with Goodpasture’s sera (lanes 2-4). NC1 in monomer form (M) could only be detected for chicken without the use of a reducing agent, whereas for the other species the Goodpasture’s sera detected dimers (D) under non-reducing conditions. Interestingly, no reactivity to *Danio rerio*, *Drosophila* and *C. elegans* NC1 was found with any of the analyzed Goodpasture’s sera (lanes 5-7). D. Reduction of the samples eliminates the conformational structure of the Goodpasture’s epitope. With slight reduction (1.5% β-mercaptoethanol) some monomer (M) from each of the three positive species native NC1s was recognized by the Goodpasture’s sera. E. Under extreme reducing conditions (10% β-mercaptoethanol), binding to all monomer and dimer forms of NC1 by Goodpasture’s sera was eliminated, as expected and previously described\(^{17}\).

FIGURE 2. Cloning and sequencing of the *Danio rerio* α3(IV)NC1 and characterization of recombinant *Danio rerio* α3(IV)NC1 domain. A. Nucleotide and peptide sequence of the NC1
domain of *Danio rerio* type IV collagen α3-chain, (zα3(IV)NC1). The two regions homologous to Goodpasture’s epitope, $E_A$ and $E_B$, are boxed. The conserved methionine and lysine (diamonds) residues, potentially involved in a conserved novel non-disulfide covalent crosslink$^{27}$, are shown. The circles indicate the twelve conserved cysteines. **B.** FLAG-tagged *Danio rerio* α3(IV)NC1 was recombinantly produced and an expected band at 29.4 kDa is recognized by the anti-FLAG antibody (lane 1), as well as with the anti-human α3(IV) NC1 antibody (lane 2). However, no binding with Goodpasture’s sera is observed (lane 3), despite prolonged exposure of film. **C.** Native *Danio rerio* kidney extract (30μg per lane) digested with bacterial collagenase was run on 15% SDS-PAGE under both reducing and non-reducing (NR) conditions. Under non-reducing conditions, only dimers (D) and trimers (T) can be observed with the anti-human α3(IV) NC1 antibody antibody (lane 1). Under reducing conditions, native α3(IV)NC1 monomers (M) are observed at the expected molecular weight (lane 3). 250 ng of FLAG-tagged recombinant human α3(IV)NC1 (rh-α3(IV)NC1) was used as positive control (lane 2)$^{35}$. **D.** No reactivity with Goodpasture’s sera with native *Danio rerio* kidney extract can be seen (lane 1). FLAG-tagged rh-α3(IV)NC1 was used as positive control (lane 2).

**FIGURE 3.** Expression of type IV collagen α3-chain in the *Danio rerio*. Adult *Danio rerio* tissues were stained with anti-human α3(IV) antibodies. **A-B.** In *Danio rerio* the gills are the equivalent of the lung. Strong expression of α3(IV) can be seen in the *Danio rerio* gills (green). DAPI staining is used to stain the nuclei of cells (blue). **C-D.** In *Danio rerio* the kidney is the main hematopoietic organ and thus morphologically distinct from that of higher organisms (**C**). Strong expression of α3(IV) can be seen in both the GBM (G), the tubular basement membranes (T) as well as in the hematopoietic tissue (HP) can be seen (**D**).
FIGURE 4. Sequence comparison of the putative Goodpasture’s epitopes E_A and E_B in different species. A. Amino acid sequence comparison of epitope E_A among various species. Numbers at the bottom denote human α3(IV)NC1 residue positions. All organisms have a threonine at position 17 (T17) in the α3(IV)NC1, and this residue is an isoleucine in the human α1(IV)-chain, which does not bind the Goodpasture’s autoantibodies. A serine at position five (S21) is similarly shared by all the species positive (+) for Goodpasture’s sera binding but is missing from the immuno-negative (-) Danio rerio, Drosophila and C.elegans. B. Sequence comparison of epitope E_B among the various species. Numbers at the bottom denote human α3(IV)NC1 residue positions. Threonine 127 (T127) is speculated to be necessary for immune reactivity at this epitope. In addition, most positive organisms share a structure with the very inflexible amino acid proline at position 131 (P131), as well as a terminal lysine (K141). These amino acids could therefore be critical for the epitope structure. Interestingly, Danio rerio is missing all three of these amino acids shared by most of the species positive for Goodpasture’s sera binding.

FIGURE 5. Species-specific sequence changes in the emergent E_A and E_B epitope regions of the α3(IV)NC1 domain. A. Alignment of epitopes and E_A and E_B of Danio rerio, chicken, and human. The six variable (non-scaffold) residues of epitope E_A and the five variable (non-scaffold) residues of E_B are numbered. The color-coding of residues (which is identical to that in figure 5B-D) denotes the chemical properties of residues, and is explained in the colored key at right. The numbered keys delineate the emergence of these epitopes as they arose from immuno-negative Danio rerio to immuno-positive chicken and human by defining the evolutionarily induced changes in chemical properties and their effects on Goodpasture’s autoantibody binding. B-G. Using the known [(α1)₂α2]₂ NC1 hexamer crystal structure (PDB ID 1LI1), we have modeled the species-specific amino acids in the E_A and E_B regions for Danio rerio (B and E), chicken (C and
and human (D and G) onto a α3(IV)NC1 backbone-homology structure. B-D. The molecular surface and amino acid character of non-conserved residues in both the E_A and E_B epitope regions for Danio rerio (B), chicken (C) and human (D) are rendered in a color-coded, semitransparent surface representation: neutral (grey), strongly hydrophilic (green), moderately hydrophilic (cyan), basic (blue) and acidic (red) amino acid character. The remainder of the α3(IV)NC1 domain backbone is shown in ribbon representation (grey). E-G. The peptide backbone from residues 17-31 (using human α3(IV) numbering) in the α3(IV)-chain corresponding to the E_A epitope region is colored blue, and from 125-139 corresponding to the E_B epitope region is colored green. Non-conserved residues in E_A and E_B are shown with rendered side chains, and identified by single letter residue code. Sequence differences between the Danio rerio and chicken E_A and E_B epitopes are highlighted in yellow on the chicken homology α3(IV)NC1 domain structure. Additionally, sequence differences between the chicken and human E_A and E_B epitope regions are highlighted further in red on the human homology α3(IV)NC1 domain structure. The Q57 critical in forming the conformational, discontinuous epitope E_A is colored magenta.
FIGURE 2

A

Zebrafish Collagen Type IV α3 NC1

B

anti-FLAG

anti-FLAG NC1

GP serum

C

NR

REDUCED

D

NR

28
FIGURE 3

H&E

anti-α3(IV) NC1

A

Gills

B

C

D

Kidney

T

HP

G
### FIGURE 4

#### A

<table>
<thead>
<tr>
<th>Species/chain</th>
<th>GP +/-</th>
<th>Epitope EA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human α1 NC1</td>
<td>-</td>
<td>IDDPQCPGKTILYH</td>
</tr>
<tr>
<td>Human α3 NC1</td>
<td>+</td>
<td>TAIPSCPGETVPLLYS</td>
</tr>
<tr>
<td>Bovine α3 NC1</td>
<td>+</td>
<td>TASI EEP S</td>
</tr>
<tr>
<td>Rattus α3 NC1</td>
<td>+</td>
<td>TANS EQP S</td>
</tr>
<tr>
<td>Mus α3 NC1</td>
<td>+</td>
<td>TASI EQP S</td>
</tr>
<tr>
<td>Gallus α3 NC1</td>
<td>+</td>
<td>TKISSQSQF</td>
</tr>
<tr>
<td>Xenopus α3 NC1</td>
<td>+</td>
<td>TVIAAKR T</td>
</tr>
<tr>
<td>Danio rerio α3 NC1</td>
<td>-</td>
<td>EVASTEHWT</td>
</tr>
<tr>
<td>Drosophila α1 NC1</td>
<td>-</td>
<td>ETVAQP SQWE</td>
</tr>
<tr>
<td>C. elegans α1 NC1</td>
<td>-</td>
<td>TAVQPSQWE</td>
</tr>
<tr>
<td>Residue number</td>
<td>17 18 19 20 21 22 23 24 25 26 27 28 29 30 31</td>
<td></td>
</tr>
</tbody>
</table>

#### B

<table>
<thead>
<tr>
<th>Species/chain</th>
<th>GP +/-</th>
<th>Epitope EB Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human α1 NC1</td>
<td>-</td>
<td>IQIPCPSGWSSLWI</td>
</tr>
<tr>
<td>Human α3 NC1</td>
<td>+</td>
<td>TDIPCPHGWISLWK</td>
</tr>
<tr>
<td>Bovine α3 NC1</td>
<td>+</td>
<td>TDIAIK</td>
</tr>
<tr>
<td>Rattus α3 NC1</td>
<td>+</td>
<td>TAVQK</td>
</tr>
<tr>
<td>Mus α3 NC1</td>
<td>+</td>
<td>TAVQK</td>
</tr>
<tr>
<td>Gallus α3 NC1</td>
<td>+</td>
<td>TAVAGIK</td>
</tr>
<tr>
<td>Xenopus α3 NC1</td>
<td>+</td>
<td>INQVLE</td>
</tr>
<tr>
<td>Danio rerio α3 NC1</td>
<td>-</td>
<td>IEVDNEGI</td>
</tr>
<tr>
<td>Drosophila α1 NC1</td>
<td>-</td>
<td>IEVDNENEGI</td>
</tr>
<tr>
<td>C. elegans α1 NC1</td>
<td>-</td>
<td>TSVPQSQMT</td>
</tr>
<tr>
<td>Residue number</td>
<td>127 128 129 130 131 132 133 134 135 136 137 138 139 140 141</td>
<td></td>
</tr>
</tbody>
</table>

MacDonald et al. Emergence of the Goodpasture's epitope
FIGURE 5
Zebrafish to humans: evolution of the α3-chain of type IV collagen and emergence of the autoimmune epitopes associated with Goodpasture’s Syndrome

Brian A MacDonald, Malin Sund, Marianne A Grant, Kathleen L Pfaff, Kathryn Holthaus, Leonard I Zon and Raghu Kalluri