



Conservation and Identity Selection of Cationic Residues Flanking the Hydrophobic Regions in Intermediate Filament Superfamily

Wenbo Zhang^{1†}, Mingwei Liu^{1†}, Robert L. Dupont², Kai Huang³, Lanlan Yu¹, Shuli Liu⁴, Xiaoguang Wang^{2,5*} and Chenxuan Wang^{1*}

¹State Key Laboratory of Medical Molecular Biology, School of Basic Medicine Peking Union Medical College, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, Beijing, China, ²William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, United States, ³Shenzhen Bay Laboratory, Shenzhen, China, ⁴Department of Clinical Laboratory, Peking University Civil Aviation School of Clinical Medicine, Beijing, China, ⁵Sustainability Institute, The Ohio State University, Columbus, OH, United States

OPEN ACCESS

Edited by:

Tsuyoshi Minami,
Institute of Industrial Science, Japan

Reviewed by:

Vladimir Kubyskhin,
University of Manitoba, Canada
Peter Cragg,
University of Brighton,
United Kingdom

*Correspondence:

Xiaoguang Wang
wang.12206@osu.edu
Chenxuan Wang
wangcx@ibms.pumc.edu.cn

[†]These authors share first authorship

Specialty section:

This article was submitted to
Supramolecular Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 03 August 2021

Accepted: 23 August 2021

Published: 02 September 2021

Citation:

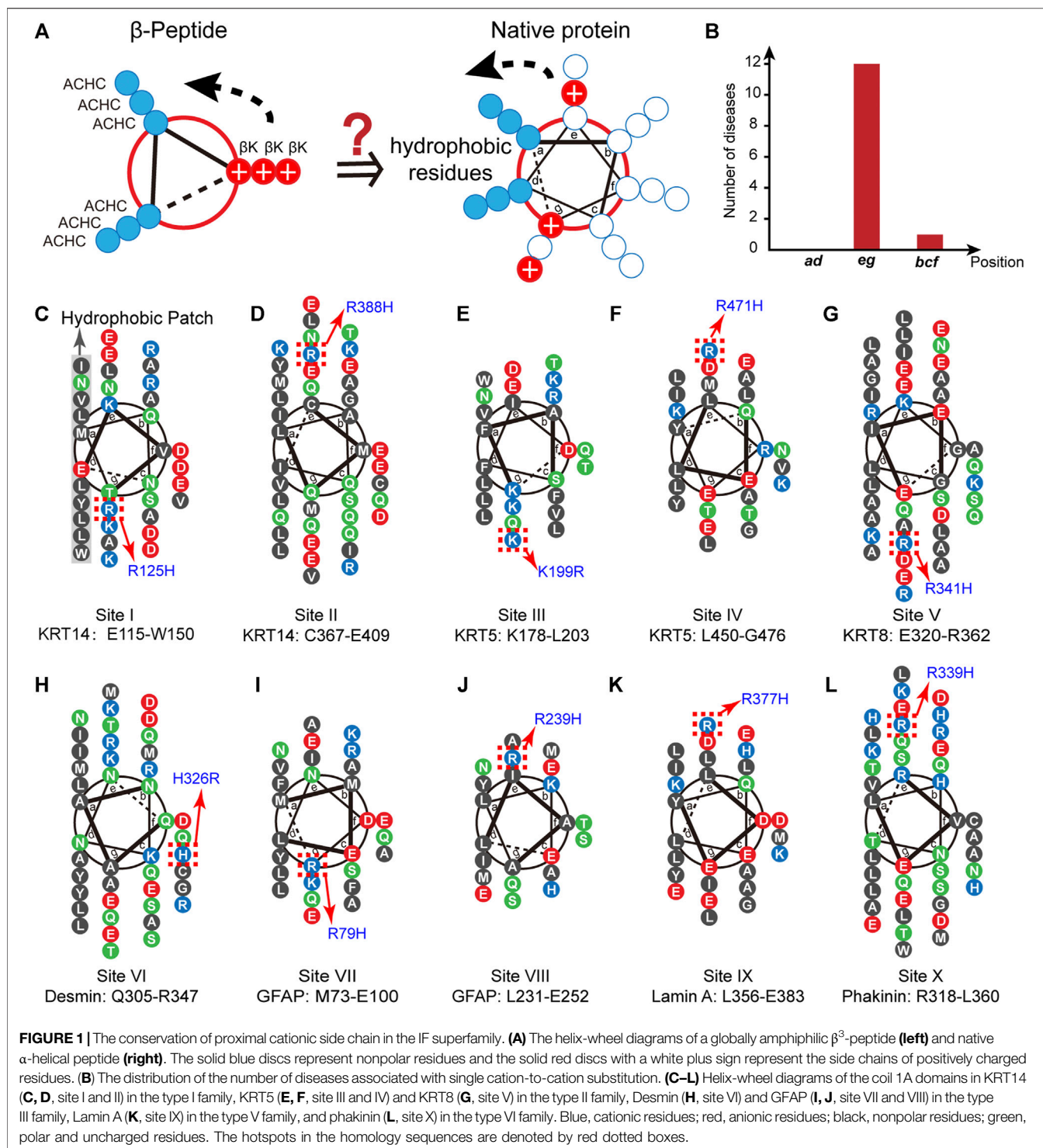
Zhang W, Liu M, Dupont RL, Huang K,
Yu L, Liu S, Wang X and Wang C
(2021) Conservation and Identity
Selection of Cationic Residues
Flanking the Hydrophobic Regions in
Intermediate Filament Superfamily.
Front. Chem. 9:752630.
doi: 10.3389/fchem.2021.752630

The interplay between the hydrophobic interactions generated by the nonpolar region and the proximal functional groups within nanometers of the nonpolar region offers a promising strategy to manipulate the intermolecular hydrophobic attractions in an artificial molecule system, but the outcomes of such modulations in the building of a native protein architecture remain unclear. Here we focus on the intermediate filament (IF) coiled-coil superfamily to assess the conservation of positively charged residue identity via a biostatistical approach. By screening the disease-correlated mutations throughout the IF superfamily, 10 distinct hotspots where a cation-to-cation substitution is associated with a pathogenic syndrome have been identified. The analysis of the local chemical context surrounding the hotspots revealed that the cationic diversity depends on their separation distance to the hydrophobic domain. The nearby cationic residues flanking the hydrophobic domain of a helix (separation <1 nm) are relatively conserved in evolution. In contrast, the cationic residues that are not adjacent to the hydrophobic domain (separation >1 nm) tolerate higher levels of variation and replaceability. We attribute this bias in the conservation degree of the cationic residue identity to reflect the interplay between the proximal cations and the hydrophobic interactions.

Keywords: protein assembly, coiled-coil, self-assembly, hydrophobic interactions, charge-related interactions

INTRODUCTION

The complex networks of weak noncovalent interactions, *i.e.*, hydrophobic interactions, electrostatic interactions, van der Waals interactions, and hydrogen bonding, engender the highly specific folding and intermolecular recognition of proteins (Pace et al., 2014; Zhang et al., 2020a; Zhang et al., 2020b). To date, however, a *de novo* prediction of the tertiary and quaternary structures from an ensemble of amino acid scratches via a free energy summation approach remains challenging (Rego et al., 2019; Zhang et al., 2020a). This difficulty is further exacerbated by the nonadditive interplay among the various noncovalent interactions. Additionally, conceptual models regarding the different types of noncovalent interactions as divided elements are typically inadequate (Giovambattista et al., 2007; Huang et al., 2015; Ma et al., 2015; Wang et al., 2017). Recently, we explored how the identity of an



adjacent cation modulates the strength of hydrophobic interactions generated by a nonpolar domain in an artificial system, specifically in synthetic amphiphilic β -amino acid oligomers (Ma et al., 2015; Wang et al., 2017). The amphiphilic β -amino acid oligomers adopt a 14-helical conformation and flank the positively charged residues of either β^3 -homolysine or β^3 -homoarginine in the proximity of a

nonpolar domain formed by six cyclohexyl groups (from *trans*-2-aminocyclohexanecarboxylic acid (ACHC) residues) (**Figure 1A**) (Ma et al., 2015; Wang et al., 2017). A single-molecule force experiment revealed that the side chains of homolysine and homoarginine display divergent effects on the magnitude of the hydrophobic interactions mediated by the ACHC-rich domain. Specifically, the protonation of β^3 -homolysine

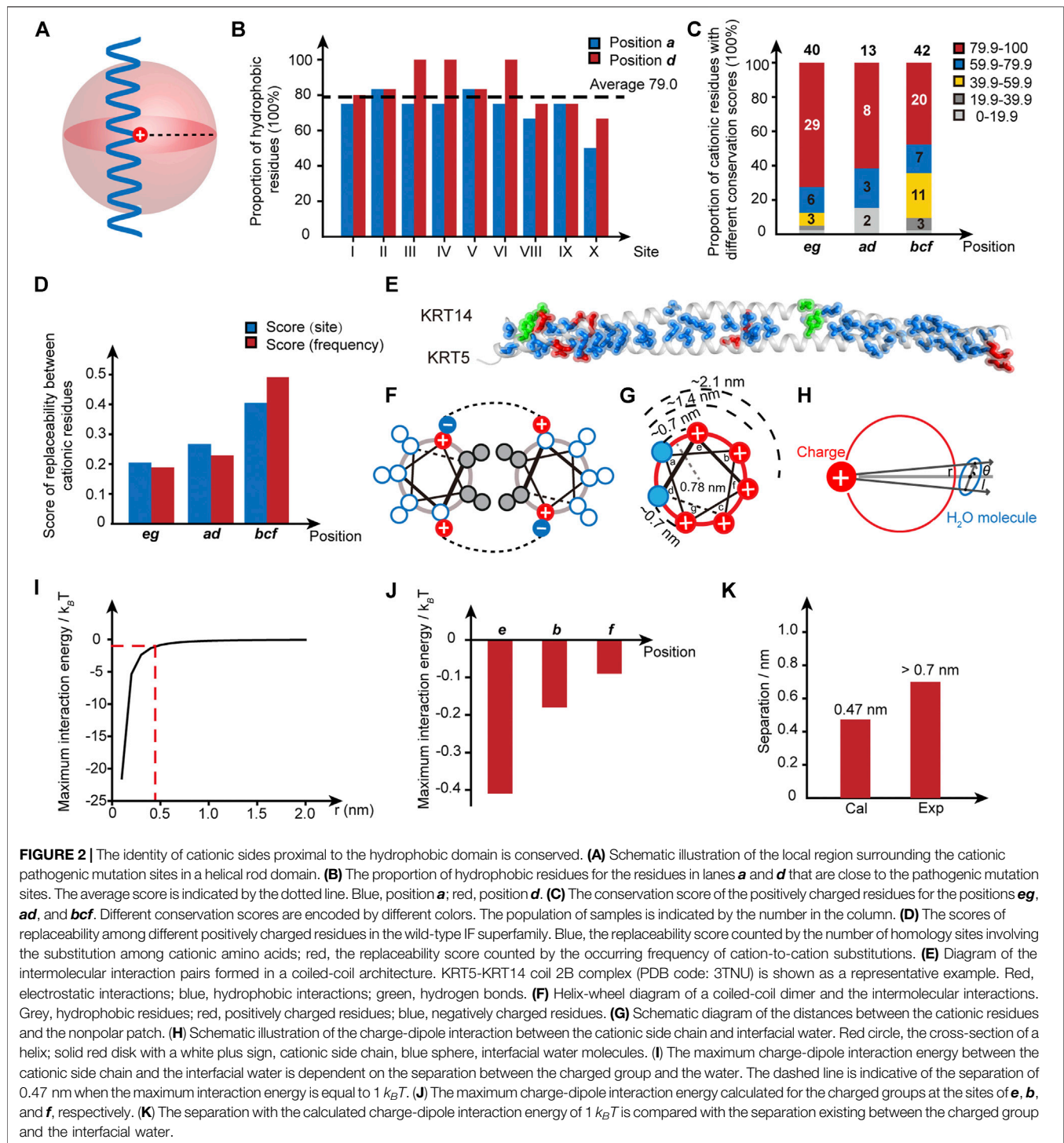
strengthens hydrophobic interactions, whereas the protonated β^3 -homoarginine diminishes hydrophobic interactions (Ma et al., 2015). This pronounced modulation is attributed to the distinct impacts of the charged groups on the local density and organizational structure of interfacial water (Wang et al., 2017). A key question raised from this study is whether or not this striking outcome affects a biologically relevant system. Considering the prevalence of cationic-hydrophobic heterozygotes in nature, such as cell-penetrating peptides (e.g., melittin and gramicidin) and coiled-coil architectures, we are motivated to investigate whether the bias of the neighboring cation identity at the boundary of a nonpolar domain is a broadly distributed evolutionary selection tool that exists in native protein systems.

We chose the superfamily of intermediate filament (IF)-forming proteins as a model native system to evaluate the conservation of positively charged residues in the proximity of nonpolar patches in the IF members via a biostatistical approach. The IF superfamily comprises six multigene families ranging from type I to type VI (Omary et al., 2004). All IF-forming proteins feature a central α -helical rod domain comprising four heptads of repeating helical segments (1A, 1B, 2A, and 2B) that present a hydrophobic central core that provides the driving force for the formation of twisted coiled-coils that further build up the structural skeletons accounting for cell mechanics (Omary et al., 2004; Jacob et al., 2018). This hydrophobic core is typically flanked by a population of positively charged residues (Figure 1A). To date, thousands of single-site substitutions within the human IF protein sequences correlating with pathogenic syndromes or non-pathogenic phenotypes have been identified. The abundant single-site variation records provide us with an enriched database to test the selection and conservation of positively charged residues flanking the nonpolar domain.

Inspection of the IF database leads to an important observation that the occurrence of diseases associated with single cation-to-cation substitution, that is the substitutions among arginine (R), lysine (K), and histidine (H), exhibits a strong correlation with the positions flanking the nonpolar patch, i.e., *e* and *g* (Figure 1B, Supplementary Table S1). We began with the point mutation at the “hotspot” R125 located in the α -helical 1A domain of keratin 14 (KRT14, type I IF family), where substitution of R125 by histidine leads to the occurrence of a monogenic disease, specifically the Dowling-Meara type of epidermolysis bullosa simplex (EBS) (Coulombe et al., 1991) (Figure 1C). This fact suggests that the identity of the positively charged residue at 125 in KRT14 is conserved to be arginine. The interchange between arginine and its positively charged homolog results in a disordered syndrome. We searched in the Human Intermediate Filament Database (<http://www.interfil.org/index.php>) (Szeverenyi et al., 2008) and screened the 28 members in the type I family (i.e., KRT9, 10, 12–20, 23–28, and 31–40). As revealed by the multiple-sequence alignment (MSA) algorithm created using ClustalW (Larkin et al., 2007) seen in Supplementary Figures S1A, 5 proteins (KRT14, 10, 13, 16, and 17) among the 28 type I members have been documented to carry a hotspot cationic residue at the

analogous positions of R125 in KRT14, where a single site substitution among different cationic homologs (lysine, arginine, and histidine) occurring at this position leads to a pathogenic syndrome (Supplementary Figure S1B). Specifically, the pathogenic mutations include the R156H in KRT10 corresponding to epidermolytic hyperkeratosis (bullous congenital ichthyosiform erythroderma) (Cheng et al., 1992; Rothnagel et al., 1992), R114H in KRT13 corresponding to oral white sponge naevus (Nishizawa et al., 2008), R127H in KRT16 corresponding to pachyonychia congenital-K16 (Wilson et al., 2014), and R94H in KRT17 corresponding to pachyonychia congenital type 2 (Smith et al., 1997). Collective structural biology evidence illustrated the coiled-coil characteristic features of these IF proteins (Fuchs and Weber, 1994; Herrmann and Aebi, 2004; Parry, 2005; Kim and Coulombe, 2007). α -Helical heptads can be depicted as repeats of seven residues (*a-b-c-d-e-f-g*). As demonstrated by the helix-wheel diagrams in Supplementary Figure S1B, the five helical wheels of KRT 14, 10, 13, 16, and 17 exhibit the same hydrophobic regions (L, I, V, Y, W or M) at the positions *a* and *d*, whereas the conserved arginine residues are located at the flanking position *g*. This preliminary result conveys a message that the identity of proximal positive charged residues, R125 and its analogous sites, close to the hydrophobic domain are not interchangeable in the IF superfamily.

Inspired by the discovery of the conserved R125 in KRT 14 (named site I) and analogues, we screened the 4,999 disease-correlated mutations throughout the IF superfamily in the Human Intermediate Filament Database and identified an additional nine distinct sites, besides site I, in the six IF families bearing variations among positively charged residues which are related to human diseases (Figures 1D–L): site II (R388H in KRT14) is associated with epidermolysis bullosa simplex Weber-Cockayne type (EBS-WC) (Abu Sa'd et al., 2006); site III (K199R in KRT5) is associated with EBS-WC (Abu Sa'd et al., 2006); site IV (R471H in KRT5) is associated with EBS-WC (Wertheim-Tysarowska et al., 2016); site V (R341H in KRT8) is associated with liver disease, inflammatory bowel disease, and primary biliary cirrhosis (Ku et al., 2005); site VI (H326R in desmin) is associated with sudden cardiac death related disease (Brodehl et al., 2013); site VII (R79H in glial fibrillary acid protein, GFAP) is associated with Alexander disease (Brenner et al., 2001; Rodriguez et al., 2001); site VIII (R239H in GFAP) is associated with Alexander disease (Brenner et al., 2001; Rodriguez et al., 2001); site IX (R377H in lamin A) is associated with limb-girdle muscular dystrophy type 1B, dilated cardiomyopathy 1A, and Emery-Dreifuss muscular dystrophy (Muchir et al., 2000); and site X (R339H in phakinin) is associated with cataract 12 (Ma et al., 2008). These specific pathogenic mutations are located in the helical rod domain of the corresponding proteins. As illustrated by the helix-wheel diagrams in Figure 1, most pathogenic mutations (I–V and VII–X) occur at positions *e* or *g* in proximity to the hydrophobic regions at *a* and *d*. A single exception, the H326R mutation in desmin (site VI), which barely impacts the folding and assembly structure of desmin, is at position *f* far from the continuous hydrophobic regions at *a* and *d* (Figure 1H) (Brodehl et al., 2013).



The similarity in the structural patterns exhibited by the disease-correlated sites (**Figure 1**) motivates us to analyze the pertinent properties of the local region surrounding the hotspot sites. Two conclusions can be drawn from the analysis. First, a hydrophobic patch close to the pathogenic mutation sites is consensual. For example, take the E115-K146 segment in KRT14 to analyze the hydrophobicity of the microenvironment surrounding site I, R125.

In lane **a**, there are four residues (M119, L126, V133, and N140) located within a separation distance of three sets of heptad repeats (21 residues, ~ 3 nm, **Figures 1C, 2A**) from the conserved site R125 in the α -helical domain. Thus, the proportion of hydrophobic residues at this position is 75% (**Figure 2B** and **Supplementary Table S2**). Similarly, the proportion of hydrophobic residues in lane **d** is 80%. As shown in **Figure 2B**, hydrophobic residues

dominate the proximity of 10 pathogenic mutation sites where the average proportion of hydrophobic residues in lanes **a** and **d** is 79%. Second, the conservation of the cationic residue identity displays a dependence on the relative position to the hydrophobic patch. To reflect the diversity of a specific cationic residue in the homology sequences within a wild-type human IF family, we defined a conservation score (C) as:

$$C = \frac{H}{T} \times 100\% \quad (1)$$

where H is the number of homology sequences carrying a conserved cationic residue at the analogous site, and T is the total number of homology sequences within the single-family in the MSA analysis. For instance, the analogous site of R125 in KRT14 is conserved through the members of the type I family in the MSA analysis (human KRT9-10, 12-20, 23-28, 31-32, 33A, 33 B, and 34) (**Supplementary Figure S2A**), and thus the conservation score of R125 is 100%. Whereas for the analogous site of K132 in KRT14, there is a total of 22 homologous members in the MSA analysis in which 16 members carry a lysine at this site. Correspondingly, the conservation score of K132 is $16/22 \approx 72.7\%$. We calculated the conservation scores for all cationic residues in the mutation-located helical rod domains of each disease-causing sequence in the IF superfamily (**Supplementary Table S3**). Intriguingly, the percentage of positively charged residue with a high conservation score ($>79.9\%$) decreases from 72.5% (29/40) at the positions **e** and **g** (proximal to the hydrophobic patch), to 61.5% (8/13) at the positions **a** and **d** (located in the hydrophobic patch), and to 47.6% (20/42) at the positions **b**, **c** and **f** (far from the hydrophobic patch) (**Figure 2C**). The average conservation score of positively charged residue throughout the IF superfamily decreases from 84.3% (**e**, **g**), to 80.5% (**a**, **d**), and to 69.7% (**b**, **c**, and **f**) (**Supplementary Table S3**). This result suggests there exists an identity selection of the adjacent cationic residues flanking the hydrophobic domain of a helix and that these cationic residues are conserved in evolution. In contrast, the cationic residues that are not adjacent to the hydrophobic domains, i.e., the cations at the positions **b**, **c**, and **f**, tolerate higher levels of variation and replaceability among different types of cationic side chains (**Figure 2D** and **Supplementary Table S4**).

We have shown that the identity of the flanking cationic residues around the hydrophobic patch is delicately selected by nature, i.e., the interchange among cationic residues at this position is rare in nature and is usually correlated with the occurrence of pathogenic syndromes. An important question emerging from this finding is why these conserved cations are located at the boundary of the hydrophobic domain. Considering there exist dozens of intermolecular interaction pairs (**Figure 2E**) between IF helices that maintain a coiled-coil, the fluctuation associated with the single-site substitution to the localized salt bridge formed between the cationic side chain on one helix and the anion side chain on another helix within a coiled-coil pair (**Figures 2E,F**) appears unlikely to account for a delocalized and long-distance perturbation in the architectures of IF proteins.

Alternatively, the structural perturbations generated by the single-site mutation reflect the interplay between the proximal cationic side chains and the hydrophobic interactions mediated by the nonpolar patch. This aligns with our previous observation of the globally amphiphilic β -amino acid oligomer system (Ma et al., 2015; Wang et al., 2017; Wang et al., 2021). Specifically, the substitution among different types of cationic side chains perturbs the local structure and density of the interfacial water around the mutation site, affecting the hydration shell around the nonpolar domain and resulting in a fluctuation in the magnitude of hydrophobic interactions.

It is necessary to comment that whether the intermolecular network presented by IF proteins would adopt a similar modulation effect as the β -peptide is not obvious, considering the complexity of the chemical pattern and noncovalent network present on the native protein surface relative to the synthetic β -peptide. If such a mechanism does exist in the α -amino acid system, it coincides with the previous experimental and theoretical results that the effective length scale of the interfacial charge modulation effect on the interfacial water shell is around 1 nm (Chandler, 2005; Acharya et al., 2010; Wang et al., 2017). The cationic residues at positions **e** and **g** are ~ 0.7 nm away from the nonpolar patch and thus are conserved to preserve the charge-related interactions and hydrophobic interactions of the structures (**Figure 2G**). In contrast, the cationic residues at positions **b**, **f**, and **c** are too distant (the distance between **a** and **b** is ~ 1.4 nm, and the distance between **a** and **f** is ~ 2.1 nm) to impact the hydrophobic interactions mediated by the nonpolar domain, and thus they display a higher level of mutation tolerance (the conservation degree of cationic residues at positions **a** and **d** is discussed in the SI). Two mechanisms might be useful in the understanding of the effective length scale. The first mechanism is the charge-dipole interaction occurring between the cationic side chain and the interfacial water molecules near the nonpolar patch. The second mechanism is proposed to originate from the unique organizational structure and density of the interfacial water molecules surrounding the mutation site, as the perturbation of the organized water structure caused by the interface usually decays within 1 nm of the interface (Bjorneholm et al., 2016).

To provide insight into the mechanism corresponding to the conservation of cationic residues at positions **e** and **g**, we calculated the interaction energy $W(r, \theta)$ for a charged side chain interacting with water as (**Figure 2H**):

$$W(r, \theta) = -\frac{Qq}{4\pi\epsilon_0\epsilon} \left[\frac{1}{r - \frac{1}{2}l\cos\theta} - \frac{1}{r + \frac{1}{2}l\cos\theta} \right] \quad (2)$$

where Q is the point-charge carried by the cationic side chain, q is the dipole charge carried by the water, r is the separation between the water and the charged group, l is the water dipole length, ϵ_0 is the vacuum permittivity, ϵ is the permittivity of a helix (approximately equal to the value for oil at ~ 10), and θ is the included angle between the molecular axis of water and the line connecting the charged side chain and the water. When θ is equal to 0, the dependence of the maximum interaction energy, $W(r, \theta = 0)$, at 300 K on the separation between the

cationic side chain and the water can be calculated (Figure 2I). As revealed by this calculation, the magnitudes of the charge-dipole interactions for the cationic residues at positions *e*, *b*, and *f* are determined to be less than $1 k_B T$, which is too small to resist thermal motion (Figure 2J). The divergence between the separation distance (0.47 nm) corresponding to the calculated charge-dipole interaction of $1 k_B T$ and the effective separation distance (>0.7 nm) measured from the protein database (Figure 2K) hints that the high conservation degree of cationic residues at positions *e* and *g* within the IF superfamily is correlated with the perturbation of interfacial water molecules by the cationic side chain instead of charge-dipole interactions.

We end this paper by making two additional comments. First, as an alternative pathogenic mechanism, post-translational modifications of cationic side chains potentially play a role in the disease development associated with the single cation-to-cation substitution within the IF superfamily. For example, K residues, instead of R residues, in the keratins can be modified by acetylation, ubiquitination, SUMOylation, and transamidation (Jacob et al., 2018). These post-translational modifications presumably provide a hierarchical regulation of the keratin self-association (Jacob et al., 2018). Second, the pKa of H side chains in a protein is typically below or near 7 (Grimsley et al., 2009), resulting in a partial deprotonation of the H side chain to some extent in a normal physiological environment, which is in contrast to the full protonation of K and R side chains in a protein (Kubyskhin, 2021).

As a summary, we traversed the disease-correlated mutations occurring in the coiled-coil IF superfamily and identified 10 distinct hotspots that carry a pathogenic cation-to-cation substitution. The chemical diversity of these hotspots is positively dependent on the separation distance between the cationic site and the nonpolar patch. Specifically, the cationic residues which flank the nonpolar patch are relatively conserved compared to those that are distant from the nonpolar patch. Considering the prevalence of the coiled-coil supramolecular structure, involved in 5.8% of the protein structures documented in the protein data bank (PDB), the importance of our results is not limited to the understanding of IF protein assembly structures but extends to a broad perspective for revealing the interplay among the pairwise noncovalent interactions in biologically relevant systems (Hadley et al., 2008; Wang et al., 2017). In the future, more experimental evidence and theoretical calculations may be needed to elucidate the effects of the flanking charged group on the

hydrophobically driven folding and assembly in native protein systems. This effort will greatly leverage our understanding of design principles that can harness these intermolecular interactions and may offer a promising, rational design method for diverse supramolecular systems. In particular, the discovery of the strong impact that the nanopatterning of hydrophobic and cationic groups exerts on the self-assembly of artificial and natural systems could have an enormous impact on a variety of research fields relevant to supramolecular chemistry, such as the *de novo* calculation of protein folding and association, drug delivery platforms, and the design of stimuli-responsive materials.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

CW and WZ, designed and initiated the project. WZ and ML screened the database, analyzed the data, and prepared all illustrations. RD, KH, LY, SL, XW helped analyzed the data. All authors listed participated in the article writing and approved it for publication.

FUNDING

This work was funded by the National Natural Science Foundation of China (31901007), CAMS Innovation Fund for Medical Sciences (2018-I2M-3-006), China Postdoctoral Science Foundation (2020T130006ZX), the Open Project Fund provided by Key Laboratory for Biomedical Effects of Nanomaterials and Nanosafety, CAS (NSKF202019), and State Key Laboratory Special Fund 2060204.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.752630/full#supplementary-material>

REFERENCES

- Abu Sa'd, J., Indelman, M., Pfendner, E., Falik-Zaccari, T. C., Mizrahi-Koren, M., Shalev, S., et al. (2006). Molecular Epidemiology of Hereditary Epidermolysis Bullosa in a Middle Eastern Population. *J. Invest. Dermatol.* 126 (4), 777–781. doi:10.1038/sj.jid.5700163
- Acharya, H., Vembanur, S., Jamadagni, S. N., and Garde, S. (2010). Mapping Hydrophobicity at the Nanoscale: Applications to Heterogeneous Surfaces and Proteins. *Faraday Discuss.* 146, 353–365. discussion 367–393, 395–401. doi:10.1039/b927019a
- Björneholm, O., Hansen, M. H., Hodgson, A., Liu, L.-M., Limmer, D. T., Michaelides, A., et al. (2016). Water at Interfaces. *Chem. Rev.* 116 (13), 7698–7726. doi:10.1021/acs.chemrev.6b00045
- Brenner, M., Johnson, A. B., Boespflug-Tanguy, O., Rodriguez, D., Goldman, J. E., and Messing, A. (2001). Mutations in GFAP, Encoding Glial Fibrillary Acidic Protein, Are Associated with Alexander Disease. *Nat. Genet.* 27 (1), 117–120. doi:10.1038/83679
- Brodehl, A., Dieding, M., Klauke, B., Dec, E., Madaan, S., Huang, T., et al. (2013). The Novel Desmin Mutant p.A120D Impairs Filament Formation, Prevents Intercalated Disk Localization, and Causes Sudden Cardiac Death. *Circ. Cardiovasc. Genet.* 6 (6), 615–623. doi:10.1161/CIRCGENETICS.113.000103

- Chandler, D. (2005). Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* 437 (7059), 640–647. doi:10.1038/nature04162
- Cheng, J., Syder, A. J., Yu, Q.-C., Letal, A., Paller, A. S., and Fuchs, E. (1992). The Genetic Basis of Epidermolytic Hyperkeratosis: A Disorder of Differentiation-specific Epidermal Keratin Genes. *Cell* 70 (5), 811–819. doi:10.1016/0092-8674(92)90314-3
- Coulombe, P. A., Hutton, M. E., Letal, A., Hebert, A., Paller, A. S., and Fuchs, E. (1991). Point Mutations in Human Keratin 14 Genes of Epidermolysis Bullosa Simplex Patients: Genetic and Functional Analyses. *Cell* 66 (6), 1301–1311. doi:10.1016/0092-8674(91)90051-y
- Fuchs, E., and Weber, K. (1994). Intermediate Filaments: Structure, Dynamics, Function and Disease. *Annu. Rev. Biochem.* 63, 345–382. doi:10.1146/annurev.bi.63.070194.002021
- Giovambattista, N., DeBenedetti, P. G., and Rosicky, P. J. (2007). Effect of Surface Polarity on Water Contact Angle and Interfacial Hydration Structure. *J. Phys. Chem. B* 111 (32), 9581–9587. doi:10.1021/jp071957s
- Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2008). A Summary of the Measured pK values of the Ionizable Groups in Folded Proteins. *Protein Sci.* 18 (1), NA. doi:10.1002/pro.19
- Hadley, E. B., Testa, O. D., Woolfson, D. N., and Gellman, S. H. (2008). Preferred Side-Chain Constellations at Antiparallel Coiled-Coil Interfaces. *Proc. Natl. Acad. Sci.* 105 (2), 530–535. doi:10.1073/pnas.0709068105
- Herrmann, H., and Aebi, U. (2004). Intermediate Filaments: Molecular Structure, Assembly Mechanism, and Integration into Functionally Distinct Intracellular Scaffolds. *Annu. Rev. Biochem.* 73, 749–789. doi:10.1146/annurev.biochem.73.011303.073823
- Huang, K., Gast, S., Ma, C. D., Abbott, N. L., and Szlufarska, I. (2015). Comparison between Free and Immobilized Ion Effects on Hydrophobic Interactions: A Molecular Dynamics Study. *J. Phys. Chem. B* 119 (41), 13152–13159. doi:10.1021/acs.jpcc.5b05220
- Israelachvili, J. N. (2011). *Intermolecular and Surface Forces*. San Diego: Academic Press.
- Jacob, J. T., Coulombe, P. A., Kwan, R., and Omary, M. B. (2018). Types I and II Keratin Intermediate Filaments. *Cold Spring Harb Perspect. Biol.* 10 (4), a018275. doi:10.1101/cshperspect.a018275
- Kim, S., and Coulombe, P. A. (2007). Intermediate Filament Scaffolds Fulfill Mechanical, Organizational, and Signaling Functions in the Cytoplasm. *Genes Dev.* 21 (13), 1581–1597. doi:10.1101/gad.1552107
- Ku, N. O., Lim, J. K., Krams, S. M., Esquivel, C. O., Keeffe, E. B., Wright, T. L., et al. (2005). Keratins as Susceptibility Genes for End-Stage Liver Disease. *Gastroenterology* 129 (3), 885–893. doi:10.1053/j.gastro.2005.06.065
- Kubyskhin, V. (2021). Experimental Lipophilicity Scale for Coded and Noncoded Amino Acid Residues. *Org. Biomol. Chem.* 19 (32), 7031–7040. doi:10.1039/D1OB01213D
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics* 23 (21), 2947–2948. doi:10.1093/bioinformatics/btm404
- Ma, C. D., Wang, C., Acevedo-Vélez, C., Gellman, S. H., and Abbott, N. L. (2015). Modulation of Hydrophobic Interactions by Proximally Immobilized Ions. *Nature* 517 (7534), 347–350. doi:10.1038/nature14018
- Ma, X., Li, F. F., Wang, S. Z., Gao, C., Zhang, M., and Zhu, S. Q. (2008). A New Mutation in BFSP2 (G1091A) Causes Autosomal Dominant Congenital Lamellar Cataracts. *Mol. Vis.* 14, 1906–1911.
- Muchir, A., Bonne, G., van der Kooij, A. J., van Meegen, M., Baas, F., Bolhuis, P. A., et al. (2000). Identification of Mutations in the Gene Encoding Lamins A/C in Autosomal Dominant Limb Girdle Muscular Dystrophy with Atrioventricular Conduction Disturbances (LGMD1B). *Hum. Mol. Genet.* 9 (9), 1453–1459. doi:10.1093/hmg/9.9.1453
- Nick Pace, C., Scholtz, J. M., and Grimsley, G. R. (2014). Forces Stabilizing Proteins. *Febs Lett.* 588 (14), 2177–2184. doi:10.1016/j.febslet.2014.05.006
- Nishizawa, A., Nakajima, R., Nakano, H., Sawamura, D., Takayama, K., Satoh, T., et al. (2008). A de Novo missense Mutation in the Keratin 13 Gene in Oral white Sponge Naevus. *Br. J. Dermatol.* 159 (4), 974–975. doi:10.1111/j.1365-2133.2008.08716.x
- Omary, M. B., Coulombe, P. A., and McLean, W. H. I. (2004). Intermediate Filament Proteins and Their Associated Diseases. *N. Engl. J. Med.* 351 (20), 2087–2100. doi:10.1056/NEJMra040319
- Parry, D. A. D. (2005). Microdissection of the Sequence and Structure of Intermediate Filament Chains. *Adv. Protein Chem.* 70, 113–142. doi:10.1016/S0065-3233(05)70005-X
- Rego, N. B., Xi, E., and Patel, A. J. (2019). Protein Hydration Waters Are Susceptible to Unfavorable Perturbations. *J. Am. Chem. Soc.* 141 (5), 2080–2086. doi:10.1021/jacs.8b11448
- Rodriguez, D., Gauthier, F., Bertini, E., Bugiani, M., Brenner, M., N'Guyen, S., et al. (2001). Infantile Alexander Disease: Spectrum of GFAP Mutations and Genotype-Phenotype Correlation. *Am. J. Hum. Genet.* 69 (5), 1134–1140. doi:10.1086/323799
- Rothnagel, J. A., Dominey, A. M., Dempsey, L. D., Longley, M. A., Greenhalgh, D. A., Gagne, T. A., et al. (1992). Mutations in the Rod Domains of Keratins 1 and 10 in Epidermolytic Hyperkeratosis. *Science* 257 (5073), 1128–1130. doi:10.1126/science.257.5073.1128
- Smith, F. J. D., Corden, L. D., Rugg, E. L., Ratnavel, R., Leigh, I. M., Moss, C., et al. (1997). Missense Mutations in Keratin 17 Cause Either Pachonychia Congenita Type 2 or a Phenotype Resembling Steatocystoma Multiplex. *J. Invest. Dermatol.* 108 (2), 220–223. doi:10.1111/1523-1747.ep12335315
- Szeverenyi, I., Cassidy, A. J., Chung, C. W., Lee, B. T. K., Common, J. E. A., Ogg, S. C., et al. (2008). The Human Intermediate Filament Database: Comprehensive Information on a Gene Family Involved in many Human Diseases. *Hum. Mutat.* 29 (3), 351–360. doi:10.1002/humu.20652
- Wang, C., Biok, N. A., Nayani, K., Wang, X., Yeon, H., Derek Ma, C.-K., et al. (2021). Cationic Side Chain Identity Directs the Hydrophobically Driven Self-Assembly of Amphiphilic β -Peptides in Aqueous Solution. *Langmuir* 37 (11), 3288–3298. doi:10.1021/acs.langmuir.0c03255
- Wang, C., Ma, C.-K. D., Yeon, H., Wang, X., Gellman, S. H., and Abbott, N. L. (2017). Nonadditive Interactions Mediated by Water at Chemically Heterogeneous Surfaces: Nonionic Polar Groups and Hydrophobic Interactions. *J. Am. Chem. Soc.* 139 (51), 18536–18544. doi:10.1021/jacs.7b08367
- Wertheim-Tysarowska, K., Oldak, M., Giza, A., Kutkowska-Kaźmierczak, A., Sota, J., Przybylska, D., et al. (2016). Novel Sporadic and Recurrent Mutations in KRT5 and KRT14 Genes in Polish Epidermolysis Bullosa Simplex Patients: Further Insights into Epidemiology and Genotype-Phenotype Correlation. *J. Appl. Genet.* 57 (2), 175–181. doi:10.1007/s13353-015-0310-9
- Wilson, N. J., O'Toole, E. A., Milstone, L. M., Hansen, C. D., Shepherd, A. A., Al-Asadi, E., et al. (2014). The Molecular Genetic Analysis of the Expanding Pachonychia Congenita Case Collection. *Br. J. Dermatol.* 171 (2), 343–355. doi:10.1111/bjd.12958
- Zhang, W., Mo, S., Liu, M., Liu, L., Yu, L., and Wang, C. (2020a). Rationally Designed Protein Building Blocks for Programmable Hierarchical Architectures. *Front. Chem.* 8, 587975. doi:10.3389/fchem.2020.587975
- Zhang, W., Yu, L., Ji, T., and Wang, C. (2020b). Tumor Microenvironment-Responsive Peptide-Based Supramolecular Drug Delivery System. *Front. Chem.* 8, 549. doi:10.3389/fchem.2020.00549

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Liu, Dupont, Huang, Yu, Liu, Wang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.