



OPEN Desmoplakin CSM models unravel mechanisms regulating the binding to intermediate filaments and putative therapeutics for cardiocutaneous diseases

Cedric Badowski¹✉, Paula Benny^{1,2,3}, Chandra S. Verma⁴ & E. Birgitte Lane^{1,5}✉

Arrhythmogenic cardiomyopathy (AC) is a common cause of sudden cardiac arrest and death in young adults. It can be induced by different types of mutations throughout the desmoplakin gene including the R2834H mutation in the extreme carboxyterminus tail of desmoplakin (DP CT) which remains structurally uncharacterized and poorly understood. Here, we have created 3D models of DP CT which show the structural effects of AC-inducing mutations as well as the implications of post-translational modifications (PTMs). Our results suggest that, in absence of PTMs, positively charged wildtype DP CT likely folds back onto negatively-charged plectin repeat 14 of nearby plakin repeat domain C (PRD C) contributing to the recruitment of intermediate filaments (IFs). When phosphorylated and methylated, negatively-charged wildtype DP CT would then fold back onto positively-charged plectin repeat 17 of PRD C, promoting the repulsion of intermediate filaments. However, by preventing PTMs, the R2834H mutation would lead to the formation of a cytoplasmic mutant desmoplakin with a constitutively positive DP CT tail that would be aberrantly recruited by cytoplasmic IFs instead of desmosomes, potentially weakening cell-cell contacts and promoting AC. Virtual screening of FDA-approved drug libraries identified several promising drug candidates for the treatment of cardiocutaneous diseases through drug repurposing.

Keywords Desmoplakin carboxyterminal tail, Desmin intermediate filaments, Arrhythmogenic cardiomyopathy, Serine phosphorylation, Arginine methylation

Abbreviations

DP CT Tail	Desmoplakin carboxyterminal tail
PTMs	Posttranslational modifications
PRD	Plakin repeat domain

Desmosomes are large cell adhesion structures which ensure robust cell cohesion within cardiac and epithelial tissues by establishing resilient mechanical links between neighboring cells. While the extracellular domain of desmosomes is made of cadherins (e.g. desmocollin and desmoglein), the intracellular or cytoplasmic region is dominated by desmoplakin (DP), a protein of the plakin family. DP interacts with plakoglobin via its N-terminus, and with intermediate filaments (IF) such as keratins (in epithelial cells) and desmin (in cardiomyocytes) via its C-terminus^{1–5}. The DP-IF interaction is reportedly mediated by three consecutive plakin repeat domains (PRD) referred to as A, B and C. However, further studies have reported that other parts of DP are also involved in IF binding, including the linker between PRD domain B and PRD domain C as well as the DP carboxyterminus tail

¹Institute of Medical Biology (IMB), Agency for Science, Technology and Research (A*STAR), Singapore 138648, Singapore. ²Department of Obstetrics and Gynaecology, National University Hospital, National University of Singapore, 1E Kent Ridge Rd, Level 12 NUHS Tower Block, Singapore 119228, Singapore. ³NUS Bio-Echo Asia Centre of Reproductive Longevity and Equality, Yong Loo Lin School of Medicine, National University of Singapore, Immunos Building, 8A Biomedical Grove, Singapore, Singapore. ⁴Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore. ⁵Skin Research Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Immunos Building, 8A Biomedical Grove, Singapore 138648, Singapore. ✉email: cedricb2023@hotmail.com; birgit.lane@sris.a-star.edu.sg

(termed DP CT tail or DP C-tail by Albrecht et al.⁶) which corresponds to the segment 2822–2871 downstream PRD domain C⁷. Interestingly, recent reports based on net charge analyses have shown that, in desmoplakin, the role played by PRD A (net charge: +3) and PRD B (net charge: +4) would be minimal compared to PRD C (net charge: +7), in terms of binding affinity towards the acidic negatively-charged rod domains of intermediate filaments (IFs)⁸. Therefore, this study seems to suggest that the main element in desmoplakin that is promoting the binding to IFs appears to be PRD C. By extrapolation, it thus seems evident that, physiologically, any regulatory process aiming to disassemble DP-IF complexes would have to mainly involve PRD C (and not necessarily PRD A and PRD B) and possibly DP CT tail as well. Indeed, DP CT tail is known to host the key phosphorylation site Ser2849 (also referred to as Sc23 due to its location 23 residues away from the C-terminus⁹) which, when phosphorylated by kinases such as PKC, PKA or GSK3, has been shown to be sufficient and necessary to promote DP-IF disassembly, while its mutation to glycine reportedly increases IF-DP interactions^{2,6,7,9–13}. However, while these reports highlight the functional importance of the carboxyterminal part of desmoplakin, more studies are needed to better understand its implications in health and disease. In particular, while PRD C is reportedly the main structure promoting the binding to IFs and DP CT tail the main regulatory part of the protein, the interplay between DP CT tail and the nearby PRD C is yet to be established.

Arrhythmogenic cardiomyopathy is a common cause of sudden cardiac arrest and death in young adults^{14,15}. It's a disease that is extremely heterogeneous both genetically and clinically. It is commonly induced by a large variety of different types of mutations (including nonsense, missense, frameshift and splice site mutations) in the plakophilin-2 (*PKP2*) gene (e.g. R158K, Q211X, R413X, L419S, S615F, K654Q, A793D, C796R, V570fs576X, C693fsX741 or N852fsX930)^{16–18} or throughout the desmoplakin (*DSP*) gene (e.g. S299R, R315C, R415G, S442F, S507F, R1113X, A2019S, E2343K, I2347V, R2541S, T2595I, R2639Q, K2689T, D2757H, R2834H, or S1015fs1023X)^{8,16,17} with some noticeable differences in terms of clinical outcomes. For instance, some DP variants, especially non-missense mutations, have been shown to predominantly lead to a left ventricle form of the disease¹⁹. Of note, a significant portion of mutations inducing a left-sided ventricular myopathy can also trigger additional clinical features including cutaneous manifestations. For instance, the mutations G2056R, E2193K, Q2371K and G2375R (which interestingly all result in the incorporation of a positively charged residue) have been associated with palmo-plantar keratosis and woolly hair, which can aid in diagnosis and often lead to the use of the term cardiocutaneous diseases⁸. Conversely, a mutation leading to a removal of a positively charged residue (R2366C) has been shown to induce a skin fragility disorder⁸. Other variants reportedly induce a biventricular type of cardiomyopathy or specifically a right dominant form of the disease (arrhythmogenic right ventricular cardiomyopathy ARVC/AC also known as arrhythmogenic right ventricular dysplasia ARVD, MIM 607450). In particular, PKP2 variants are frequently found in patients with right dominant or biventricular forms of the disease^{18,20,21}.

Etiologically, the missense variants affecting the N-terminal part of desmoplakin (such as S299R, R315C, S442F, R415G, and S507F) are likely to affect the recruitment of key molecular partners by this region which could limit the incorporation of DP into desmosomes or promote the dissolution of desmosomes altogether^{16,17}. On the other hand, variants located in the C-terminal part of the protein could potentially affect the recruitment of intermediate filaments IFs to sites of cell-cell contact. The ARVC-inducing missense variants located in the carboxyterminal part of DP can reportedly affect all 3 PRD domains (e.g. A2019S in PRD A, E2343K in PRD B and R2639Q in PRD C) and often (but not always) involve the incorporation or removal of a positively charged residue⁸. The missense mutation R2834H in the DP CT Tail is particularly well known to be responsible for the development of arrhythmogenic cardiomyopathy, highlighting its critical role in DP function and its involvement in disease pathophysiology²². Ultimately, both N-terminal and C-terminal missense variants could potentially result in the same outcome, i.e. the absence of linkage between the plasma membrane of cardiomyocytes and the cytoskeleton of desmin intermediate filaments.

Such a loss of interactions could also be mediated by a dysregulated desmosomal protein degradation triggered by an imbalance in the ubiquitination-neddylolation equilibrium that particularly affects the protein desmoplakin, potentially constituting one of the most advanced molecular signatures ever reported for this disease^{16,17,23}. This is quite relevant especially considering that the carboxyterminal part of DP reportedly constitutes the main binding site for the Cops3 subunit of the de-neddylating COP9 signalosome (CSN)²³, while the N-terminus of DP has been shown to interact with the subunit 6 of the COP9 signalosome (CSN6)¹⁶. These findings are key to understand ARVC pathogenesis and more investigations are needed to elucidate the structural aspect of this molecular mechanism.

Truncating mutations associated with ARVC (e.g. R1113X, S1015fs1023X) can be extremely insightful since they can provide important clues on what is the minimum length of the desmoplakin protein that is required to maintain a certain degree of functionality in vivo while determining the part of the protein that has the most implications in terms of involvement in the pathogenesis of ARVC and other types of cardiomyopathies¹⁶. In particular, the Carvajal syndrome (MIM#605676, characterized by left-sided ventricular cardiomyopathy, palmoplantar keratoderma and woolly hair) which is reportedly induced by a truncation of desmoplakin resulting in the complete loss of PRD C and DP CT tail (while keeping PRD A and B entirely intact and fully functional) seems to further suggest that the end part of the desmoplakin protein plays a major role in the disease's development^{19,24}. Importantly, truncating cardiomyopathy-inducing frameshift variants p.(Ser2859fs) (rs727504909), p.(Tyr2862fs) (rs765683790) and p.(Ser2863fs) occurring in the DP CT tail of desmoplakin just a few residues away from the stop codon undoubtedly confirm the key role played by DP CT Tail in the pathogenesis of cardiomyopathies²⁵.

While arrhythmogenic cardiomyopathy is known to be associated with mutation of the desmoplakin protein, the understanding of the underlying DP-related molecular mechanisms that contribute to disease pathogenesis remains very limited, significantly reducing the prospect of discovering effective therapeutics in the near future. For instance, the molecular mechanisms mediating DP CT tail interactions with IF and their disassembly

upon Ser2849 phosphorylation remain poorly defined. Most importantly, the downstream molecular effects of Arg2834His mutation on DP structure and function, and their implications in arrhythmogenic cardiomyopathy are still currently unknown. Such a critical lack of knowledge and the absence of available treatment means there is an urgent unmet clinical need and highlights the importance of initiating new studies focusing on desmoplakin structure and functions in cardiomyocytes.

Here, we used cutting-edge bioinformatic tools to construct novel 3D structural models of DP CT tail that not only shed new light on its local structure and physicochemical properties, but also provide key insights into the mechanisms involving PTMs that regulate the reported interactions between DP CT tail and intermediate filaments. The models also suggest structure-based explanations for the deleterious effects of Arg2834His mutation in arrhythmogenic cardiomyopathy as well as promising therapeutic avenues based on predictions from virtual screening of FDA-approved drug libraries.

Results

Current status of scientific knowledge on the structure of DP CT tail

A detailed search for available crystal structures of desmoplakin in the UniProt database and RCSB PDB Protein Data Bank indicated that none of them include peptides within the DP CT tail (2824–2871) (Fig. 1). Indeed, the two original desmoplakin crystal structures (PDB 1LM7 and PDB 1LM5) refer to residues 2209–2456 and 2609–2822 respectively, and do not contain any amino acid beyond position 2822, that would be part of the DP CT tail⁴. The other two models (PDB 3R6N and PDB 5DZZ) refer to structures closer to the DP N-Terminus (positions 178–627 and positions 1960–2448, respectively)^{26,27}. In the absence of available crystal structures for DP CT tail, we then assessed whether DP CT tail would share some homology with known crystalized regions of related proteins that could be used to build a homology model. BLAST analyses revealed that DP CT tail (2825–2854) shares significant homology with the tail domain of plectin (4618–4647), another cell adhesion molecule involved in IF-binding and cell-matrix adhesion (Fig. 1). However, none of the crystal structures reported for plectin include residues beyond position 4606 according to UniProt database and RCSB PDB (Table 1), further supporting the absence of known structures for sequences that match the DP CT tail in homologous proteins. This critical lack of available data on the structural organization of DP CT tail and homologous regions in related proteins poses significant challenges for homology model building of DP CT tail, and thus supports the use of alternative strategies such as *de novo* 3D modeling. A search for Computed Structure Models (CSMs) in the RCSB PDB database revealed that, while there are indeed AlphaFold models such as AF-P15924-F9 that cover the segment 2824–2871 corresponding to DP CT tail, the structures proposed for this region show very low confidence (pLDDT < 50, Fig. 1) and, although DP CT has been hypothesized to remain unstructured or structurally random, this remains ultimately debatable given the known functional importance of this region, its high degree of conservation across species and the emergence of lethal diseases such as cardiomyopathies upon its mutation.

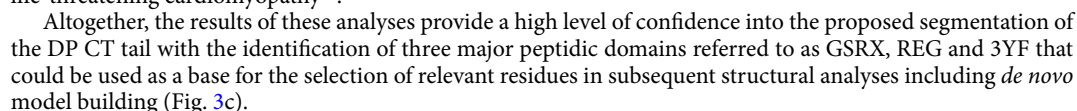
PEP-FOLD peptide modeling as a reliable method for *de novo* peptide secondary structure prediction in desmoplakin protein folding analyses

The lack of crystal structure clues led us to consider the use of alternative *de novo* modeling approaches which are specialized in the prediction of secondary structures of small peptidic domains (below 36 residues), such as the PEP-FOLD algorithm. The PEP-FOLD peptide modeling method has been successfully used for the accurate prediction of secondary structures and *de novo* building of 3D structural models primarily based on the peptide's amino-acid sequence^{28–33}. We first tested the ability of PEP-FOLD by benchmarking against the reported crystal structure of the IF-binding domain of desmoplakin (PRD C, 2613–2808, PDB ID 1LM5). The desmoplakin region 2613–2808 indeed contains at least 8 main α -helices which altogether constitute a 3D docking site for intermediate filaments. For each of these 8 α -helices, the amino-acid sequence was extracted and uploaded into the PEP-FOLD algorithm for *de novo* model building and secondary structure prediction. The results of such analyses are illustrated in Fig. 2. PEP-FOLD accurately identified an α -helical conformation for each of these peptidic sequences as seen in the desmoplakin crystal structure PDB 1LM5, with mild variations at the terminal regions.

Altogether, the literature reports^{28–33} and our results strongly support the use of PEP-FOLD-based modeling to predict the 3D structures of the uncharacterized desmoplakin carboxyterminus tail (DP CT Tail).

DP CT tail segmentation prior to PEP-FOLD *de novo* 3D modeling

PEP-FOLD can accurately predict the secondary structures of peptides of interest but only for a maximal length of 36 residues. To select the most relevant 36 amino-acids within the 48 residue-long DP CT tail (2824–2871), we segmented the DP CT tail based on the identification and demarcation of specific peptidic regions with potentially high functional importance. The selection process was essentially based on (i) the amino-acid composition of the regions and the presence of patterns such as repeats, (ii) the conservation of the regions across species, (iii) the homology between desmoplakin and plectin sequences, and (iv) the presence of functionally important residues such as known phosphorylation sites including Ser2849 (DP) and Ser4642 (PLEC)^{9,12,34}. This methodology led to the identification of three major peptidic domains. The first domain referred to as the “GSRX” region corresponds to the segment 2824–2843 in the DP CT tail and the segment 4617–4636 in the plectin tail. In desmoplakin, this region was defined by the extensive repetition of the GSRS motif which is highly conserved across species as shown by multiple sequence alignment (Fig. 3a). A similar pattern was observed in plectin but with a certain degree of variation in the repeated motif (e.g. GSRT, GSRA), justifying the overall GSRX labeling for this region (Fig. 3b). The second segment referred to as “REG” (for regulatory) corresponds to a highly conserved 11 residue-stretch spanning the positions 2844–2854 in the DP CT tail and 4637–4647 in the plectin sequence. These regions host the major phosphorylation sites Ser2849 (DP) and Ser4642 (PLEC)



PDB entry	Method	Resolution (Å)	Chain	Positions
1MB8	X-ray	2.15	A	175–403
2N03	NMR		A	4403–4606
2ODU	X-ray	2.30	A	410–640
2ODV	X-ray	2.05	A	410–640
3F7P	X-ray	2.75	A/B	175–403
3PDY	X-ray	2.22	A/B	653–858
3PE0	X-ray	2.95	A/B	750–1028
4GDO	X-ray	1.70	A/B/C/D/E/F	1492–1530
4Q58	X-ray	4.00	A/B	175–400
4Q59	X-ray	2.30	A/B	175–400
5J1F	X-ray	3.00	A/B	860–928
			A/B	999–1116
5J1G	X-ray	1.80	A/B	1114–1343
5J1H	X-ray	2.80	A/B	860–928
			A/B	999–1116
5J1I	X-ray	2.80	A/B	1114–1482

Table 1. Available crystal structures for plectin protein based on Uniprot protein database.

PEP-FOLD *de novo* modeling predicts an antiparallel β -sheet secondary structure within DP CT tail leading to a “hairpin” type of conformation

Based on the DP CT tail segmentation described above, we selected the most relevant 36 residue-long sequence (that covers the full 3YF and REG regions and a portion of the GSRX region) and used PEP-FOLD for the prediction of the secondary structure and *de novo* 3d structural models (Fig. 4a, b). The results from PEP-FOLD modeling indicated the likely presence of a β -sheet conformation in the DP CT tail, with a small probability of an α -helical secondary structure (Fig. 4b, green areas: β -sheet; red patterns: α -helix). The antiparallel β -sheet identified by PEP-FOLD in the DP CT tail would in fact result in the formation of a characteristic hairpin-like structure, essentially made by the overlapping of the GSRX and 3YF regions (Fig. 4c, d and, e). Further analyses of the PEP-FOLD 3D model with 3D Structure Viewer iCn3D (NIH) revealed that the hairpin associated with the antiparallel β -sheet would be extensively stabilized by a large number of hydrogen bonds as well as 3 π -cation interactions and a salt bridge (Fig. 4f). The H-bonds (threshold set at 3.8Å) were detected especially between serine residues of the GSRX region (Ser2831, Ser2833, and Ser2835) and those of the 3YF region (Ser2861, Ser2859, and Ser2857) with the following pairings and distances: Ser2831-Ser2861 (2.7Å), Ser2833-Ser2859 (2.9Å), Ser2835-Ser2857 (2.8Å) (Fig. 4g). The hairpin structure would also be maintained by 3 π -cation interactions (threshold set at 6Å) between Tyr2860 and Arg2830 (4.8Å), Tyr2862 and Arg2830 (5.6Å), and especially between Tyr2858 and Arg2834 (4Å) which would provide further stability to the β -sheet (Fig. 4h and i). Altogether, the three reported π -cation interactions suggest that all three tyrosine residues of the 3YF region may actively be involved in stabilizing interactions that would contribute to the hairpin structure, attributing unreported functional importance to this region as supported by its high degree of conservation across species. The fact that Arg2834 is predicted to interact with Tyr2858 via a π -cation interaction likely to stabilize the hairpin conformation may explain why mutation of this residue is known to lead to severe diseases such as arrhythmogenic cardiomyopathy²². A salt bridge (threshold set at 6Å) was also detected between Arg2842 and Asp2851 (3.9Å) (Fig. 4j and k). While this ionic interaction is not physically part of the β -sheet, it may nonetheless contribute to the hairpin structure by stabilizing a U-shaped curvature in the protein where the peptidic sequence changes direction. The turning point where such a drastic change in spatial orientation occurs (similar to a “ β -turn”) partially corresponds to the location of the REG domain and contains the key phosphorylation site Ser2849 (Fig. 4l). 3D molecular surface visualization notably revealed the presence in the β -turn area of several bulky and/or charged residues in close proximity, including Arg2842 (that is involved in the salt bridge), Phe2850, as well as two consecutive arginine residues -Arg2846 and Arg2847- which would contribute to the β -turn by the combined effect of steric hindrance and charge repulsion (Fig. 4l). Since this part of the protein appears to be rich in positively charged residues (including Arg2842, Arg2846 and Arg2847), the introduction of the negative charge upon phosphorylation of Ser2849 is expected to trigger a major structural reorganization of the DP CT tail. This is particularly relevant considering that DelPhi equipotential mapping (with potential contour set at 2kT/e -25.6mV at 298 K- and a salt concentration set at 0.15 M, Fig. 4m) as well as titration curve analyses (based on amino-acid composition, Fig. 4n) revealed that overall DP CT tail presents a strong net positive charge that is likely to promote the interaction with negatively charged peptides or protein domains, something that can potentially be affected by PTMs such as serine phosphorylation.

To assess whether our DP CT tail model is biologically relevant and structurally similar to known protein domains found in human proteins, we performed further analyses based on the comparison of our original model with known experimentally-determined PDB structures using the Structure Similarity Search module of RCSB PDB database. The results showed that the predicted structure of DP CT is relevant and apparently fairly common since it has been detected in 49 different proteins (Figure S1a). Interestingly, many of the proteins harboring a structure similar to our DP CT model seems to be related directly or indirectly to the proteasome

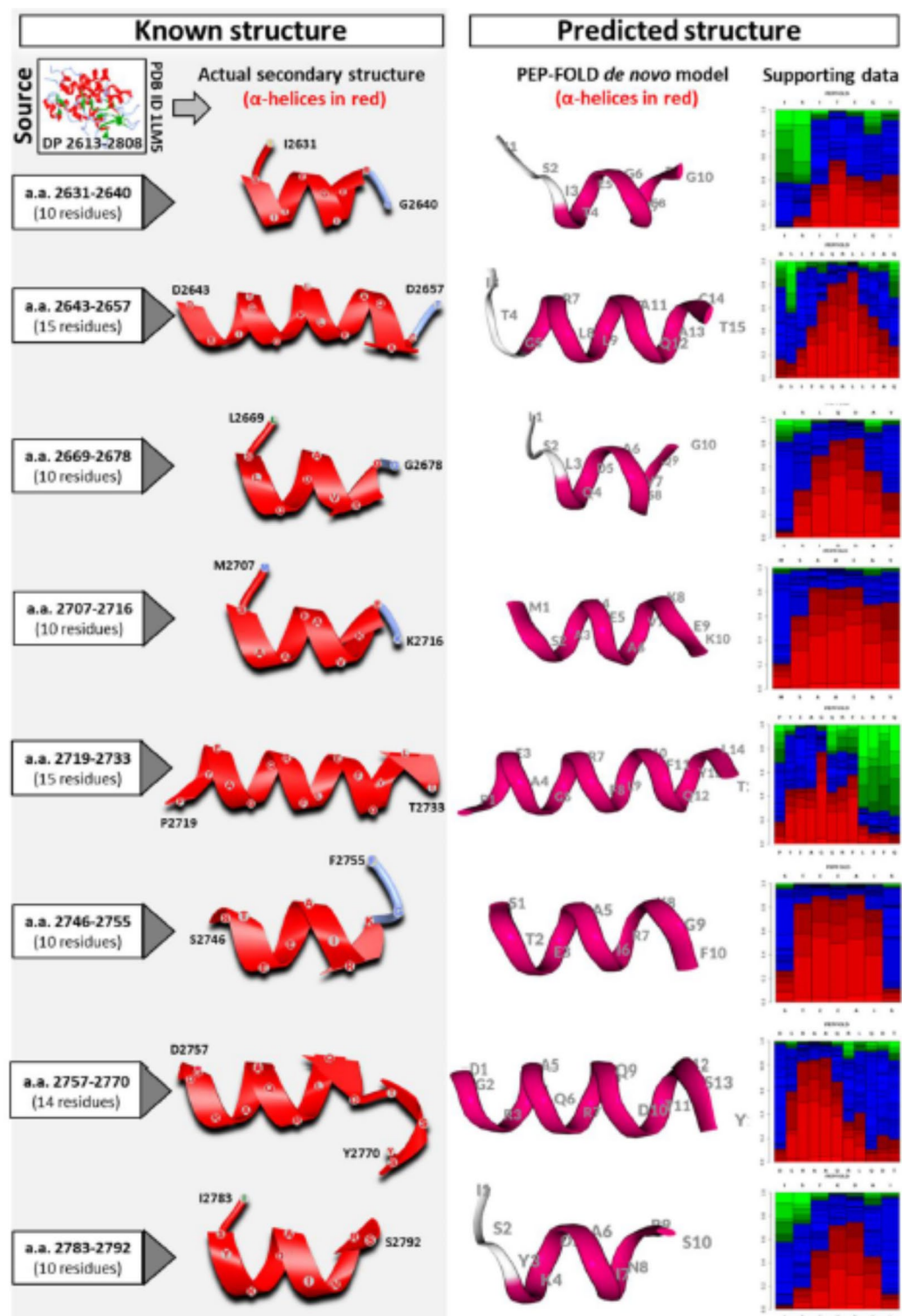


Fig. 2. Validation of PEP-FOLD peptide modeling algorithm for the prediction of secondary structures in desmoplakin protein. Comparison between known secondary structures extracted from the reported IF-binding PRD domain C of desmoplakin (2613–2808, PDB ID 1LM5) and *de novo* 3D models provided by PEP-FOLD algorithm (red: α-helix). The PEP-FOLD graphs on the right side illustrate the respective probabilities for each type of secondary structure (red: α-helix; green: β-sheet).

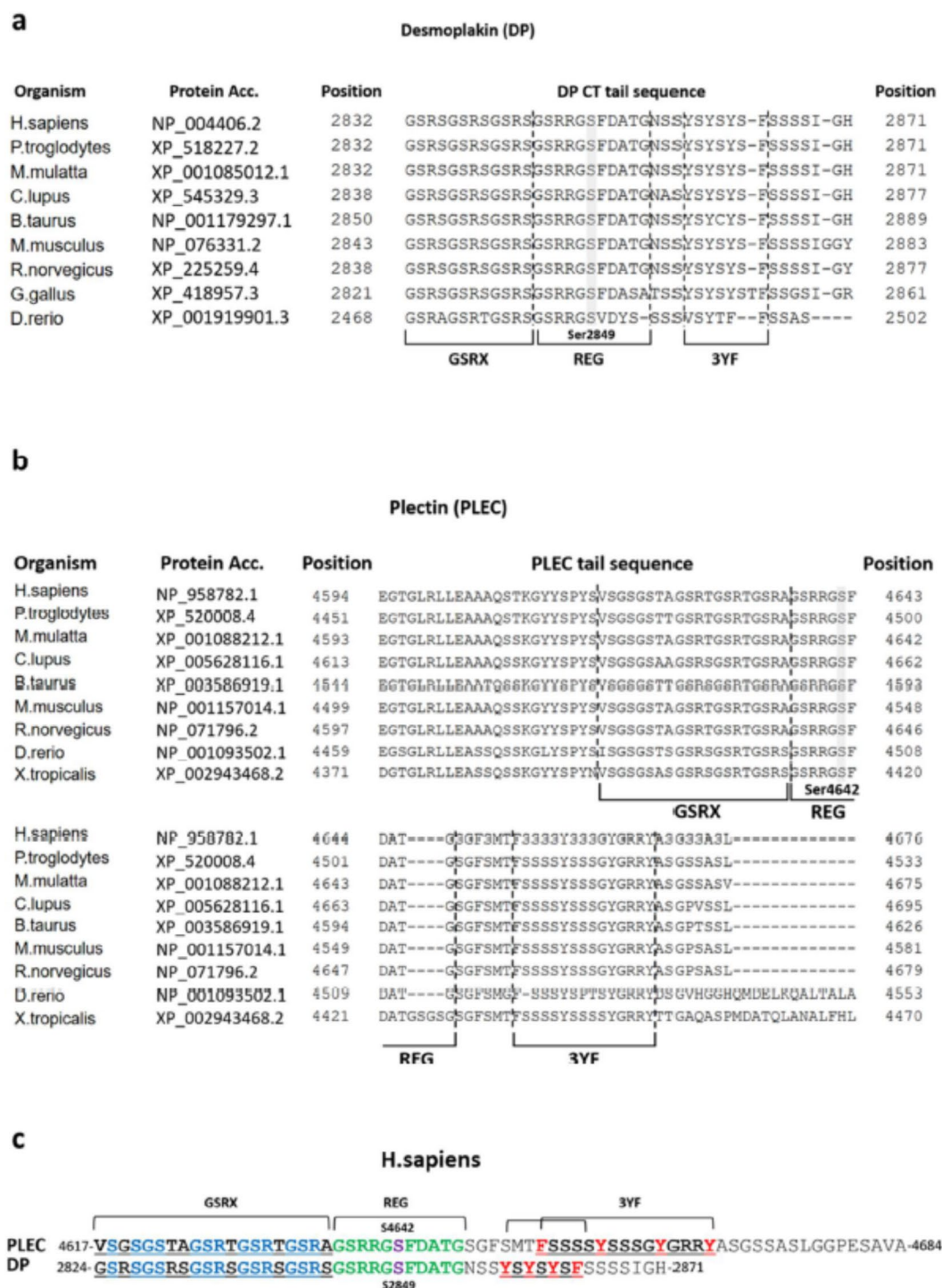


Fig. 3. DP CT tail segmentation based on multiple sequence alignment comparing desmoplakin and plectin sequences across species. **(a)** Multiple sequence alignment assessing DP CT tail sequence homology across species. Three main regions are distinguished and highly conserved: the GSX region for the repetition of the GSX motif, the REG domain (for regulatory) that contains the key phosphorylation site Ser2849 and the 3YF region that is characterized by the systematic presence of 3 tyrosine residues and 1 phenylalanine residue across species. **(b)** Multiple sequence alignment assessing plectin tail sequence homology across species. Three main regions are distinguished and highly conserved: the GSX region for the repetition of the GSX motif, the REG domain (for regulatory) that contains the key phosphorylation site Ser4642 and the 3YF region that is characterized by the systematic presence of 3 tyrosine residues and 1 phenylalanine residue across species. **(c)** Alignment of human DP CT tail (2824–2871) and human plectin tail (4617–4684) showing similarities in domain segmentation including the GSX, REG and 3YF regions. Blue: residues common to DP and PLEC in the GSX region. Green: residues common to DP and PLEC in the REG domain. Red: aromatic residues in the 3YF region of DP CT and plectin tails.

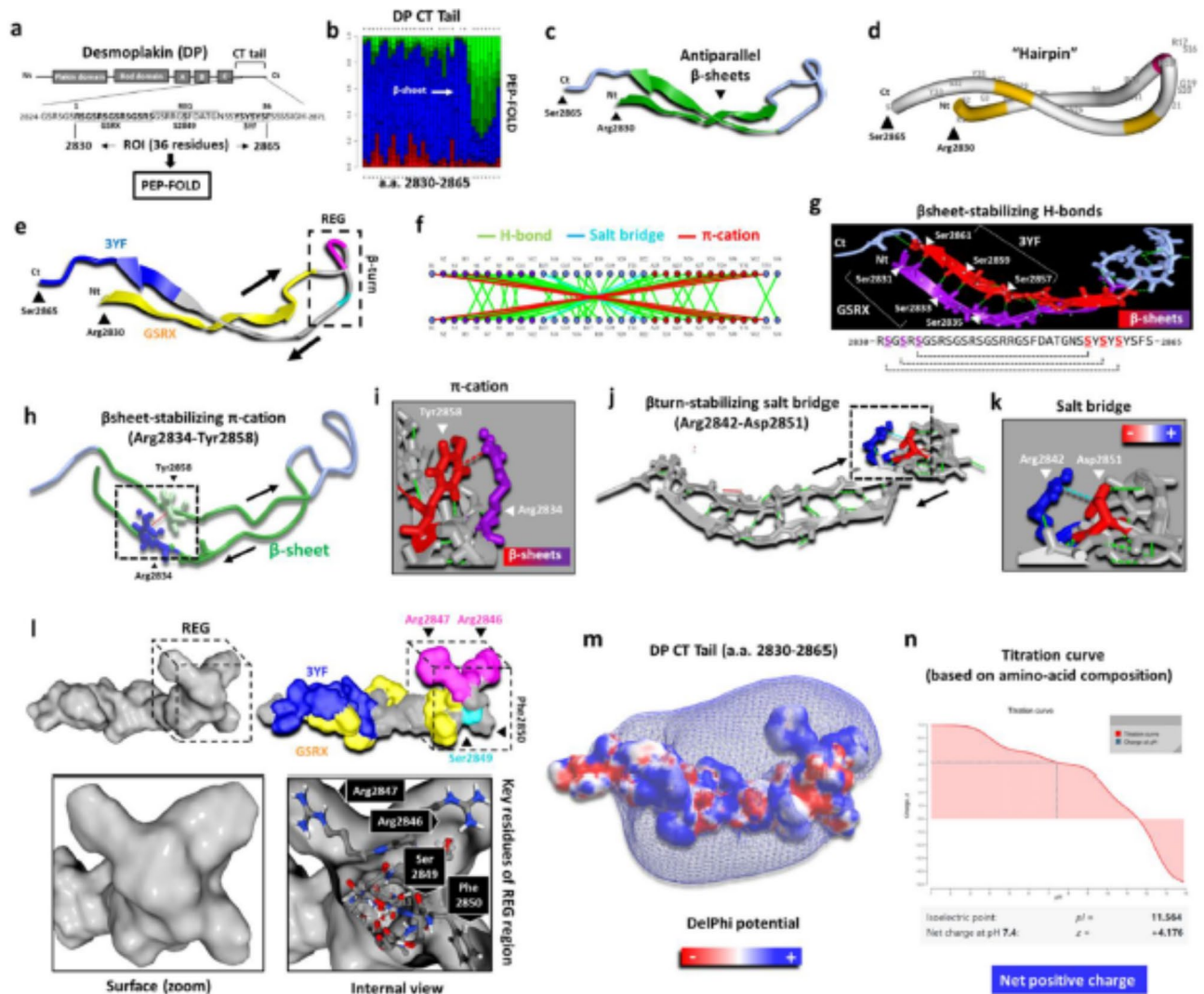


Fig. 4. PEP-FOLD-based modeling of DP CT Tail predicts an antiparallel β -sheet stabilized by multiple hydrogen bonds, π -cation interactions and a salt bridge, with a net positive charge. **(a)** 36 residue-long region of interest selected for uploading into the PEP-FOLD algorithm. **(b)** PEP-FOLD graph showing the respective probabilities for each type of secondary structure (red: α -helix; green: β -sheet). **(c)** Ribbon representation of the antiparallel β -sheet (green) predicted by the PEP-FOLD algorithm and displayed by iCn3D Structure Viewer (NIH). **(d)** Rope representation of the β -sheet-associated hairpin organization of DP CT tail as predicted by PEP-FOLD. **(e)** Ribbon representation of the antiparallel β -sheet showing the GSRX region (yellow) overlapping with the 3YF region (blue). **(f)** 2D interaction map showing the involvement of H-bonds, π -cation interactions and a salt bridge in the β -sheet-associated hairpin structure of DP CT tail. **(g)** Major hydrogen bonds (H-bonds, green dashed lines) between serine residues of the GSRX and 3YF regions stabilizing the antiparallel β -sheet (spectrum color display: one strand in red, the other in purple). H-bond threshold set at 3.8 Å (iCn3D, NIH). White arrows highlight serine residues involved in H-bonds in the antiparallel β -sheet (not exhaustive). H-bond pairing and distances: Ser2831-Ser2861 (2.7 Å), Ser2833-Ser2859 (2.9 Å), Ser2835-Ser2857 (2.8 Å). **(h)** Localization of the Tyr2858-Arg2834 π -cation interaction which is predicted to be a key element stabilizing the anti-parallel β -sheet (Arg2834 in blue, Tyr2858 in light green). **(i)** Detail of the π -cation interaction (red dashed line) between Tyr2858 (from one strand of the β -sheet, red) and Arg2834 (from the other strand of the β -sheet, purple) (distance: 4 Å). π -cation threshold set at 6 Å. The other 2 π -cation interactions identified by PEP-FOLD (Tyr2860-Arg2830, 4.8 Å, and Tyr2862-Arg2830, 5.6 Å) are not represented. **(j)** Inset highlights the precise localization of the salt bridge between Arg2842 and Asp2851 which is predicted to stabilize the β -turn of the hairpin. **(k)** Detail of the salt bridge (cyan dashed line) between Arg2842 (blue, positively charged) and Asp2851 (red, negatively charged) (distance 3.9 Å). Salt bridge threshold set at 6 Å. **(l)** Surface display (grey) of DP CT Tail (PEP-FOLD). REG: regulatory region. Yellow: GSRX region. Blue: 3YF region. Magenta: Arg2846 and Arg2847. Cyan: Ser2849. The 3D square highlights the predicted location of the β -turn. The zooms below highlight the positions of key residues within the β -turn. **(m)** DelPhi equipotential map and molecular surface with potential of DP CT tail. Potential contour set at 2kT/e (25.6mV at 298 K) with salt concentration at 0.15 M. Grid size: 65 (iCn3D, NIH) **(n)** Titration curve of DP CT tail indicating a net positive charge of +4.176 at pH 7.4 (based on amino-acid composition) (Prot pi, <https://www.protpi.ch/Calculator/ProteinTool#Results>).

(PDB: 5VHS) and ubiquitination (E3 protein ligase: PDBs: 8DWL, 8DWK, 8U5G), which is reportedly part of a finely tuned balance with neddylation²³, a critical protein modification that particularly affects desmosomal proteins in ARVC^{16,17}. Other proteins whose activity is known to be regulated by ubiquitination have also been highlighted by our search (Fig. S1a). Of note, systematic quantitative proteomics have shown that desmoplakin can physically interact with NEDD8³⁵ (Fig. S1b). Overall, while this search does not directly confirm our model, it does provide key valuable supporting information such as: (1) the type of DP CT structure proposed by our model does physically exist and is present in many other proteins; (2) this structure seems to be related to post-translational modifications (e.g. ubiquitination, neddylation) which are known to affect desmosomal proteins and play an important role in ARVC pathogenesis; (3) our findings supporting a structure-function link between our DP CT tail model and ubiquitination-neddylation modifications are consistent with reports showing that the carboxyterminal part of DP constitute the main binding site for the Cops3 subunit of the de-neddylating COP9 signalosome (CSN)²³. Our model is also supported by ESMFold predictions (iCn3D, NIH) which indicated a similar hairpin structure bringing the Nt and Ct of DP CT tail spatially closer to each other and which also placed important regulatory residues such as Ser2849 closer to the point of curvature, similarly to what PEP-FOLD predicted (but no actual anti-parallel β -sheet was confirmed by ESMFold) (Fig. S2).

Altogether, the data provided by PEP-FOLD, ESMFold, DelPhi equipotential maps, titration curve analyses and the Structure Similarity Search module of RCSB PDB suggest that the DP CT tail would assume a hairpin conformation with a strong net positive charge at cellular pH, a structure that also appears to be reminiscent of proteins involved in the ubiquitination-neddylation balance which is known to be dysregulated in ARVC. However, the DP CT tail does not contain any lysine (K) residues, suggesting the potential involvement of allosteric effects with neighboring protein domains.

DP CT tail likely acts as a molecular adaptor that folds back onto plectin repeat 14 of PRD domain C and enhances electrostatic interactions with desmin rod domains

To understand the mechanism of action of DP CT tail, we investigated the neighboring region including the nearby PRD domain C (Fig. 5a), for which a crystal structure has been successfully obtained (PDB 1LM5)⁴ (Fig. 5b). Interestingly, it seems that the four plectin repeats that constitute PRD C do not display similar net charges. For instance, while K-rich plectin repeat 17 (PLEC17) presents a net positive charge (+2.035) at pH 7.4 (Fig. 5b), plectin repeat 14 (PLEC14) is highly anionic (net charge: -7.772 at pH 7.4) as indicated by the red DelPhi equipotential map around PLEC14 (Fig. 5c, potential contour set at 2kT/e -25.6mV at 298 K- and a salt concentration set at 0.15 M). This may appear quite surprising since the rod domain of desmin (model AF-P17661-F1) is extremely negatively charged as well, and PRD domain C (which comprises PLEC14) of desmoplakin is well known to bind the rod domain of intermediate filaments. Novel 3D models then proposed that due to its large net positive charge (in the absence of PTMs) as well as its expected relative flexibility (high glycine content), unmodified DP CT tail may potentially have the ability to fold back onto negatively charged PLEC14. This would mask exposed negatively charged regions and eventually promote stabilizing electrostatic interactions with the rod domains of desmin intermediate filaments, which are required to ensure the critical linkage between membrane-bound desmoplakin-rich desmosomes and the cytoskeleton of cardiomyocytes (Fig. 5d). This is supported by PyRx-based docking analyses which showed that our DP CT Tail model is indeed capable of binding PLEC14 of PRD C (PDB 1LM5) with relatively high affinity (Fig. 5e).

The PTM switch: a major molecular mechanism mediating a loss of interaction between DP CT tail and the rod domains of intermediate filaments

While DP CT tail is relatively small in length, it is nonetheless the object of two major cascades of post-translational modifications which are intricately linked. Both eventually lead to a major shift in the net charge of the DP CT tail, in particular from a highly positive to a highly negative net charge, which may have significant structural consequences (Fig. 6a). First, a serine phosphorylation cascade, which is reportedly initiated by phosphorylation of Ser2849, progressively leads to the phosphorylation of a multitude of serine residues in a very specific sequential order (Ser2845, Ser2843, Ser2841, etc.). However, in vitro studies have shown that this phosphorylation cascade cannot proceed without a second cascade of PTMs consisting in the methylation of key arginine residues such as Arg2834 which is mutated in arrhythmogenic cardiomyopathy⁶ (Fig. 6a). To assess the structural rearrangements induced by serine phosphorylation and arginine methylation as well as their impact on local and global charges, we designed novel models of DP CT tail that mimic such PTMs by performing virtual site-directed mutagenesis. Site-directed mutagenesis is often carried out in vitro to address the downstream functional effects of PTMs, and involves mimicking phosphorylation by replacing a serine residue with a negatively charged aspartic acid (S $\rightarrow$ D) (hereby reflecting the covalent binding of negatively charged phosphate groups), and mimicking methylation by replacing arginine with a phenylalanine residue (R $\rightarrow$ F)³⁶ (Fig. 6b). By using available information from UniProt database as well as in vitro data from K. Green's Lab⁶ to know the exact number and location of all known PTMs, we generated a virtual fully methylated, fully phosphorylated (FM/FP) form of DP CT tail which corresponds to the end product of both PTM cascades. All these models were then analyzed through DelPhi equipotential mapping as well as titration curve analyses to better understand the effects of PTMs on the local and net charges DP CT tail (Fig. 6c). Our results indicated that the progressive adding of PTMs incrementally affects the charge of DP CT tail which gradually becomes more and more negative (with a rapid and substantial decrease in the size of the -blue- positive equipotential map), either directly through the covalent binding of negative phosphate groups onto serine residues or indirectly by removing positive charges through the methylation of arginine residues (Fig. 6c). In the last stage, the FM/FP form of DP CT tail is predicted to harbor a strong net negative charge of -8.819 at pH 7.4 with a massive (red) negative equipotential map surrounding it. Since PLEC14 of PRD domain C is expected to remain negatively charged as well, it becomes clear that the strong negative charge of FM/FP DP CT tail would represent a major

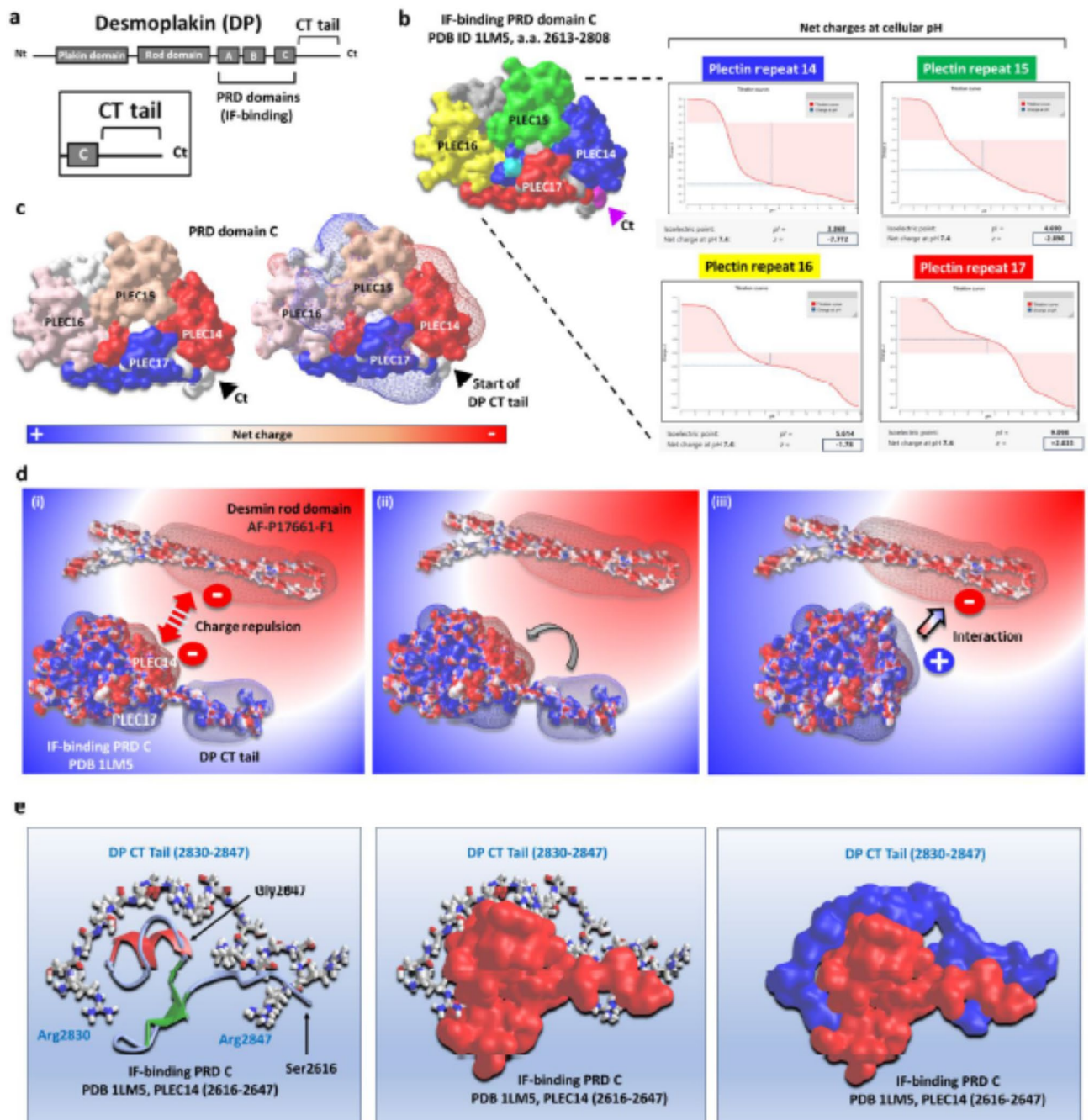
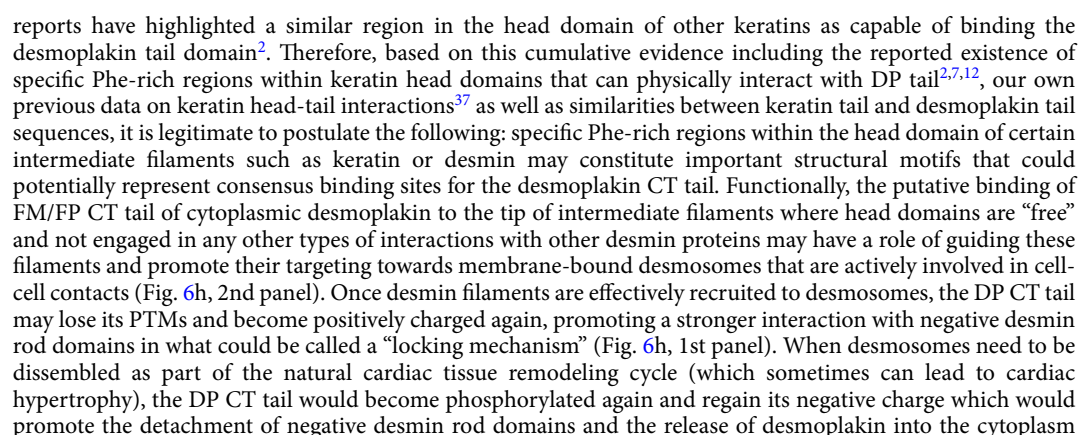


Fig. 5. Molecular mechanism involving positively charged unmodified wildtype DP CT tail as a molecular adaptor that enhances the recruitment of desmin rod domains to PRD domain C by masking negatively charged plectin repeat 14. **(a)** Diagram depicting the close physical proximity between PRD domain C (a.a. 2613–2808) and DP CT tail (a.a. 2824–2871). PRD domain C is one of three plakin repeat domains (A, B and C) which are known to bind intermediate filaments in vitro. **(b)** Illustration of the crystal structure of PRD domain C (PDB ID 1LM5) and its four plectin repeats: plectin repeat 14 (PLEC14, blue), plectin repeat 15 (PLEC15, green), plectin repeat 16 (PLEC16, yellow) and plectin repeat 17 (PLEC17, red). The net charge of each of these plectin repeats at pH 7.4 is indicated below the respective titration curves (Prot pi, <https://www.protpi.ch/Calculator/ProteinTool#Results>). **(c)** Charge-based color coding providing a representation of the net charge of each of the plectin repeats of PRD domain C (blue: positive, red: negative) as well as corresponding DelPhi equipotential maps. Potential contour set at $2kT/e$ (25.6mV at 298 K) with salt concentration at 0.15 M. Grid size: 65 (iCn3D, NIH). **(d)** Sequence of events (i–iii) predicted by electrostatic interactions between positively charged wildtype DP CT tail (PEPFOLD model) and negatively charged PLEC14 of neighboring PRD domain C (PDB 1LM5) resulting in the recruitment of rod domains of desmin intermediate filaments (AlphaFold model AF-P17661-F1). DelPhi equipotential maps from iCn3D (NIH). Potential contour set at $2kT/e$ (25.6mV at 298 K) with salt concentration at 0.15 M. Grid size: 65. **(e)** PyRx-based docking analyses confirming interactions between DP CT Tail and PLEC14 of PRD C (PDB 1LM5).

destabilizing factor for the DP CT/PLEC14 complex. In fact, the two concomitant PTM cascades are likely to trigger a “PTM switch” whereby the PTM-rich DP CT tail is progressively pushed away from the negatively charged PLEC14 towards the positively charged PLEC17 (Fig. 6d), which is located on the opposite side of PRD domain C (as is seen in the crystal structure PDB 1LM5; Fig. 5c). By doing so, the FM/FP DP CT tail would not only expose the negatively charged PLEC14 but also mask the positively charged PLEC17, resulting in the overall exposure of a large amount of negative charges on the surface of the PRD domain C (Fig. 6d, 4th panel). Consequently, such a negative PRD domain C may lose its ability to effectively recruit negatively charged desmin rod domains, potentially leading to the dissociation of desmin/DP complexes and the complete disassembly of desmosomes which is known to occur *in vitro* upon extensive serine phosphorylation. Of note, the high content in glycine residues (from the GSR repeats) suggests a substantial degree of flexibility which would allow the dynamic polymodal DP CT tail to flip back and forth between PLEC14 and PLEC17 of PRD C depending on its net charge determined by its PTM status. This is also supported by the fact that DelPhi equipotential maps of the experimentally-determined structure of PRD C (PDB 1LM5) indicated that the beginning of the CT tail (i.e. last amino-acid of PRD C) is precisely located between the negative (red) and positive (blue) electrostatic zones of PRD C (corresponding to PLEC14 and PLEC17, respectively) which represents an adequate position that is conducive to bilateral movements (Fig. 5c). This crucial feature was also observed for the CT tail of the plectin protein suggesting that this molecular mechanism may be common to the members of the plakin family (Figs. S3, S4). Interestingly, further analyses revealed that while desmin rod domains are mostly negatively charged, the head domains are in fact mainly positively charged with an expected net charge of +5.211 at pH 7.4 (Fig. 6e). These positive charges essentially carried by arginine and lysine residues appear to be mostly located on the two flanks of an aromatic-rich region which in fact strikingly resembles the 3YF region of the DP CT tail (which is itself highly conserved across species and thus suggests a major evolutionary role and key biological function). Indeed, amino-acid sequence analyses revealed that, like the 3YF region of DP CT tail, the head domain of desmin is also rich in aromatic residues with a highly hydrophobic region (residues 9–44) containing at least 4 Phe amino-acids clustered in a sequence of 35 residues in length flanked by positively charged residues on both sides (Fig. 6f). If we consider that the two flanks of the 3YF region are expected to become highly negatively charged due to the extensive serine phosphorylation and arginine methylation from the two PTM cascades, it then appears plausible that the head domain of desmin (a.a. 9–44) and the tail domain of desmoplakin (a.a. 2843–2871, which comprises the 3YF region) might be able to interact with each other as supported by polarity plots (Zimmerman scale) indicating that the DP CT tail and desmin head display similar hydropathy patterns (i.e. hydrophilic-hydrophobic-hydrophilic) which is suitable for the establishment of intermolecular interactions (Fig. 6f). These would notably include electrostatic interactions on the two flanks of a central highly hydrophobic region which may involve intermolecular aromatic Pi-Pi interactions as previously described for keratin 1 and keratin 10 terminal domains³⁷ (Fig. 6g, Fig. S5a). These predicted interactions between the very beginning of desmin head domain and the very end of desmoplakin tail domain are consistent with *in vitro* data showing that truncated DP CT tail on its own is capable of binding intermediate filaments^{2,7,12}. Moreover, our finding that the DP CT tail is likely to interact directly with desmin head domains via an alternative mode of interaction (intermolecular aromatic interactions) independently from PRD domain C (which reportedly can bind IFs as well), may explain *in vitro* data which until now remained elusive. For instance, it was never clearly understood how DP CT tail can effectively bind IFs *in vitro* without the presence of an actual plakin repeat domain. The π - π intermolecular aromatic interactions proposed here between the DP CT tail and the desmin head domain may provide a potential explanation for such an observation and could indeed explain why a small region like DP CT tail which is lacking any IF-binding PRD was nonetheless shown to be able to bind intermediate filaments *in vitro*, while its deletion did not effectively prevent truncated DP constructs without a DP CT tail from interacting with IF proteins since PRD C and other PRD domains can bind IFs as well⁷. On the other hand, single point mutation of residues within the DP CT tail such as Ser2849 or Arg2834 have been shown to significantly alter IF-binding properties, suggesting that the DP CT tail may eventually have more of a regulatory role than a structural role *per se*^{7,12,22}. Still, the ability of the DP CT tail to bind IF is undeniable and proven experimentally *in vitro*. It may actually stem from the fact that the DP CT tail does resemble the tail domain of an intermediate filament to a certain degree, while the head and tail domains of many IFs are known to interact with each other to promote polymerization and filament extension. Indeed, sequence homology analyses revealed that the 3YF region of the DP CT tail partially resembles the tail domain of intermediate filaments such as keratins which are naturally rich in tyrosine and serine residues (Table 2). Early work on modeling of intermediate filaments such as keratins dating from the 1990s proposed that the numerous tyrosine residues of keratin tail domains may self-arrange into tyrosine stacks via π -stacking, with interspaced folded flexible glycine loops. These would enable the compaction of keratin terminal domains while promoting their extension under mechanical stress to provide flexibility and resilience to the skin tissue^{38,39}. To ease the visualization of such aromatic residue stacking and enhance the similarities between the tail domains of DP and keratin intermediate filaments, we displayed these regions following a novel nomenclature referred to as the Steinert representation (in memory of the late scientist who first introduced the concept of tyrosine residue stacking and glycine loop packing in keratin tail domains)^{38,39} (Table 3). We previously developed this model further by proposing that the numerous tyrosine residues of keratin tail domains may interact via a succession of intermolecular π -stacking interactions with the phenylalanine residues of keratin head domains (as an “aromatic zipper”) to mediate physical interactions between keratin tetramers while providing mechanical flexibility (through a “molecular spring”) to the epithelial tissue (Fig. S5a)³⁷. In the same modeling analyses, we noticed in the head domain of some keratins such as K1 the presence of a short peptidic region (46–62 in K1) which, despite its high content in Phe residues, was not physically part of the intermolecular stacking with the Tyr residues of keratin tails, suggesting potential interactions with alternative protein partners such as desmoplakin. Moreover, Meng et al. have shown by two-hybrid analysis that the region 59–79 in the K1 head domain was able to interact with the DP tail¹² (Fig. S5b). Interestingly, other



◀**Fig. 6.** The “PTM switch”: effects of post-translational modifications (PTMs) on the structure and net charge of wildtype (WT) and mutant DP CT tail, and their implications in terms of binding to desmin rod and head domains in pathophysiological contexts. **(a)** Diagram illustrating the two interconnected cascades of PTMs known to occur within WT DP CT tail, including a phosphorylation (P) cascade of key serine residues (green) starting with S2849 phosphorylation (Start) and a methylation cascade (Me) of key arginine residues (red) such as R2834 (*) that is mutated in arrhythmogenic cardiomyopathy. Total number, exact location and sequential order of PTMs were extracted from UniProt database and in vitro data from Albrecht et al.⁹. The numbers (1), (2) and (3) refer to the first three steps of the interconnected PTM cascades known to occur in DP CT tail⁹. **(b)** Phospho-mimetic and methyl-mimetic virtual constructs of WT DP CT tail that mimic the phosphorylation of S2849 (S2849D) (first step) or the methylation of R2834 in combination with S2849 phosphorylation (S2849D + R2834F, intermediate step) or the phosphorylation and methylation of the entire sets of serine and arginine residues known to receive PTMs as indicated in **(a)** (Full Met./Full Phos., final step of both PTM cascades). **(c)** Net charge and charge distribution in phospho-mimetic and methyl-mimetic virtual models of WT DP CT tail generated by PEP-FOLD. DelPhi equipotential maps and surfaces with potential from iCn3D (NIH). Potential contour set at 2kT/e (25.6mV at 298 K) with salt concentration at 0.15 M. Titration curves and net charges at pH7.4 from Prot pi (<https://www.protpi.ch/Calculator/ProteinTool#Results>). **(d)** Sequence of events (i–iv) representing the PTM-induced shift in the net charge of WT DP CT tail (“PTM switch”) which electrostatically would force negatively charged PTM-rich DP CT tail to move away from negative PLEC14 towards positively charged PLEC17 (PDB 1LM5), exposing large areas with negative charges from both PLEC14 and PTM-rich DP CT tail resulting in the complete disengagement of negative desmin rod domains (AlphaFold model AF-P17661-F1). **(e)** Desmin head domain extracted from AF-P17661-F1 model and represented with DelPhi equipotential map and surface with potential (iCn3D, NIH). The titration curve below reveals a net positive charge at cellular pH. **(f)** Hydrophobicity plots (Hopp & Woods scale) (<http://www.pepcalc.com/>) showing that desmin head domain (a.a. 9–44) contains an aromatic-rich hydrophobic sequence flanked by arginine-rich positive regions. Similar graph show that PTM-rich DP CT tail (a.a. 2843–2871) would also contain an aromatic-rich hydrophobic sequence (including 3YF region) flanked by negative regions made of phospho-serine and methyl-arginine residues. Polarity plots (Zimmerman scale) (<https://web.expasy.org/protscale/>) suggest similar polarity patterns **(g)** Juxtaposition of similar regions of desmin head domain (a.a. 9–44) and PTM-rich DP CT tail (a.a. 2843–2871) suggesting potential intermolecular aromatic Pi-Pi interactions (center) and electrostatic bounds (flanks). **(h)** Expected effects of PTM switch at the cellular level and the absence of it in disease.

Keratins	Aromatic- and serine-rich sequence in protein C-terminus
KRT9	485-YGRGSRGGSGGSYGGGGSGGGYGGGSGSRGGSGGSYGGGSGSGGGSGGGYGGGSGGGHSGGSGGGHSGGSGGN YGGGSGSGGGSGGGYGGGSGSRGGSGGSHGGGSGFGGESGGSYGGGEEASGSGGGY-613
KRT2	532-YSSGSSSYGSGGRQSGSRGGSGGGGSISSGGYSGGGGSGGRYSGGGGSKGGSISGGYSGGGGKHSSGGGSRGGSSS GGGY-612
KRT1	526-YGSGGSSSYGSGGGSYGSGGGGGGGGRGSYSGGSSSYSGGGGSGSGGGGGHGSYSGSSSSGGY-588
KRT10	473-YGGGSSGGGSSGGGHGGGHGGSSGGYGGGSSGGGSSGGGYGGGSSGGHGGSSSGYGGGSSGGGGGGYGGG SSGGGSSSGGGY-557
KRT8	419-YAGGLSSAYGGLTSPGLSYSLGSFSGAGSSSF-452
KRT5	502-YGSGSGYGGGLGGGLGGGLAGGSSGSYY-533
KRT6	515-YSYSGSLGVGGGF-527
KRT16	438-YSSREVF-444
PLEC	4654-FSSSSYSYSGYGRRY-4668
DP	2858-YSYSYSF-2864

Table 2. Comparison of aromatic- and serine-rich sequences in the tail domain of several keratins and in desmoplakin (DP) and plectin (PLEC). Bold: aromatic residues (mostly tyrosine). Italic: serine residues.

where FM/FP DP CT tail would be not able to bind cytoplasmic desmin IFs (except the “free” head domains) due to its numerous PTMs (Fig. 6h, 2nd panel). However, in the case of the R2834H mutation, it is highly likely that the CT tail of cytoplasmic mutant desmoplakin proteins, especially if newly translated, remains positively charged (the phosphorylation cascade reportedly requires R2834 methylation), which would promote the spontaneous binding of mutant desmoplakin to negatively charged desmin rod domains in the cytoplasm instead of to desmosomes (Fig. 6h, 3rd panel), which has been clearly reported by in vitro studies⁶. As a result, the sequestration of mutant DP into the cytoplasm and its subsequent poor localization to desmosomes may affect the strength of cell-cell contacts and prevent the formation of a contiguous and resilient cardiac tissue. This may affect the effectiveness of cardiac contractility and potentially contribute to arrhythmia.

Keratins	Steinert representation of aromatic residue stacking and glycine loop packing in keratin tail domains
KRT9	485-Y _{GSR} Y _{GS} Y _{GSR} Y _{GS} Y _{GSHN} Y _{GS} Y _{GSRH} F _{GSE} Y _{GSEA} Y-613
KRT2	532-Y _{GS} Y _{GSRQI} Y _{GSR} Y _{GSKI} Y _{GSKHR} Y-612
KRT1	526-Y _{GS} Y _{GS} Y _{GSR} Y _{GS} Y _{GS} Y _{GSH} Y _{GS} Y-588
KRT10	473-Y _{GSH} Y _{GS} Y _{GSH} Y _{GS} Y _{GS} Y-557
KRT8	419-Y _{GSLA} Y _{GSTPL} Y _{GSL} F _{GSA} F-452
KRT5	502-Y _{GS} Y _{GSLA} YY-533
KRT6	515-Y _S Y _{GSLV} F-527
KRT16	438-Y _{SREV} F-444
PLEC	4654-F _S Y _{GS} Y _{GR} Y-4668
DP	2858-Y _S Y _S Y _S F-2864

Table 3. Comparison of aromatic- and serine- rich sequences in the tail domain of several keratins and in desmoplakin (DP) and plectin (PLEC) displayed by Steinert representation. The Steinert representation enhances the visualization of tyrosine stacking and glycine loop packing in keratin tail domains, as reported by Steinert et al.^{38,39}. Bold: aromatic residues (mostly tyrosine). Letters in subscript refer to packed glycine loops between two aromatic residues. Each letter represents a type of amino-acid present at least one time in the glycine loop. For simplicity, the flexible loops do not include the numbers of each residue.

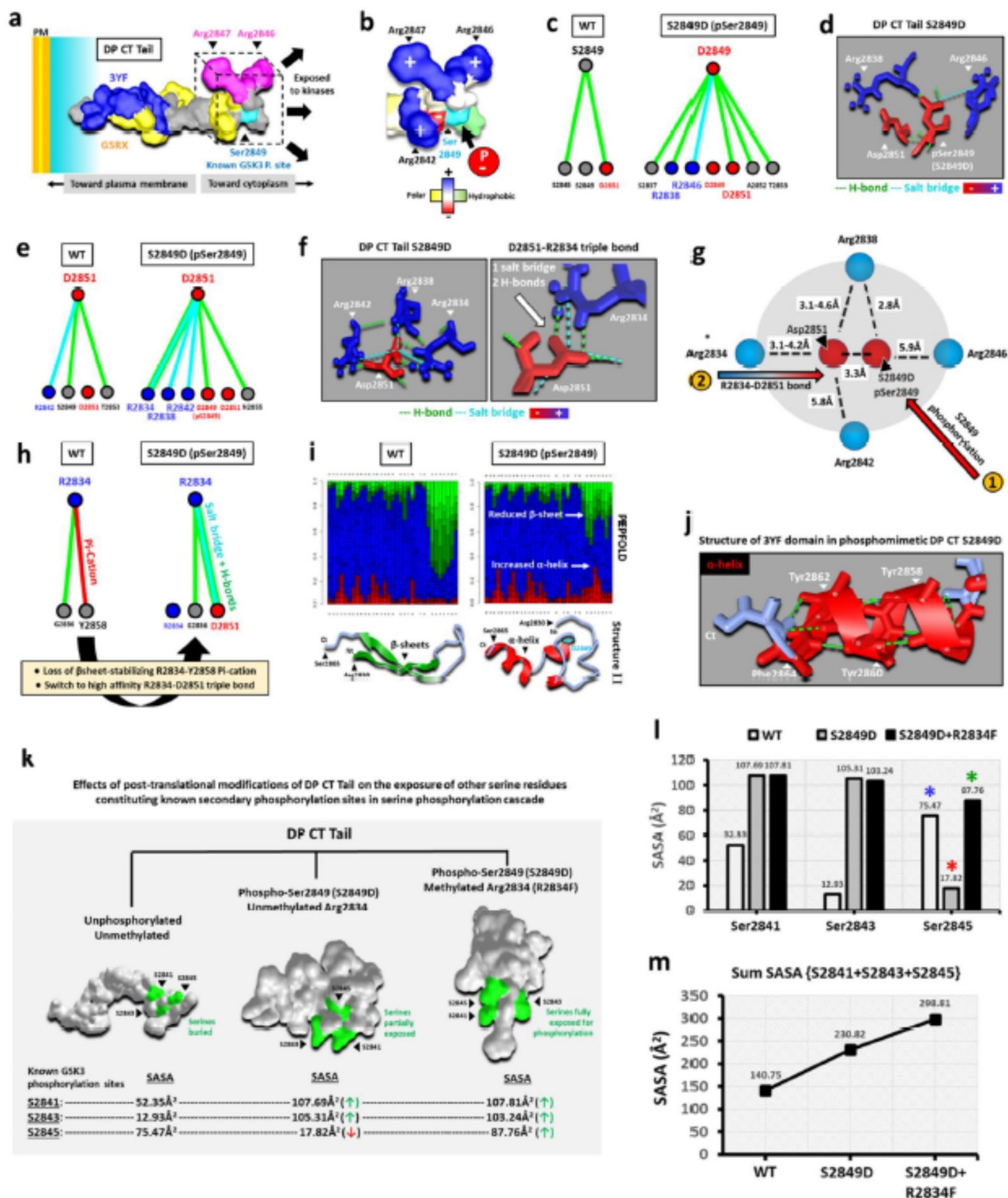
Assessing the effects of Ser2849 phosphorylation and Arg2834 methylation on the structure of DP CT tail and implications of R2834H mutation in PTM cascades in arrhythmogenic cardiomyopathy

Phosphorylation of Ser2849 has been shown to significantly affect the ability of DP to bind intermediate filaments and regulate cell-cell adhesion^{6,9,12}; however, the underlying molecular mechanism and its potential implications in arrhythmogenic cardiomyopathy remain unclear. Since our predictive model of the WT DP CT tail suggests that Ser2849 is likely located within a β -turn of a hairpin structure, this region is expected to be highly sensitive due to its considerable exposure to cytoplasmic kinases such as PKA, PKC and GSK3 which are known to phosphorylate Ser2849^{6,9–12} (Fig. 7a). Moreover, Ser2849 seems to be located in an area that is extensively rich in positively charged arginine residues (Fig. 7b). This suggests that the covalent binding of negative charges associated with Ser2849 phosphorylation may potentially lead to major structural rearrangements in the DP CT tail and possibly alter its IF-binding properties. Indeed, our simulations using the phosphomimetic S2849D DP CT tail suggested that, while Ser2849 would bind only to three other residues in the WT unmodified DP CT tail, phospho-Ser2949 would become prone to interact with arginine residues such as Arg2838 (via H-bond) and Arg2846 (via salt bridge) while retaining its original affinity for Asp2851 (via H-bond) (Fig. 7c, d). Asp2851 itself would then be able to recruit Arg2838 (via H-bond) and Arg2842 (via H-bond) and Arg2834 via 2 H-bonds and 1 salt bridge (Fig. 7e, f). Such electrostatic interactions including the strong Asp2851-Arg2834 triple interactions appear to be made possible through the clustering of positively charged residues (i.e. Arg) and negatively charged residues (i.e. Asp and phospho-Ser) triggered by the phosphorylation of Ser2849 located in an arginine-rich area (Fig. 7g). It is important to note that the Asp2851-Arg2834 interaction would occur at the expense of Arg2834-Tyr2858 π -cation interaction which is predicted to stabilize the β -sheet in the WT unmodified DP CT tail, potentially considerably destabilizing the β -sheet when Ser2849 is phosphorylated (Fig. 7h). Indeed, in silico predictions suggested that the probabilities of forming an α -helix for instance increase when Ser2849 is phosphorylated (Fig. 7i and j). In turn, such a major change in secondary structure could affect the exposure of other residues such as Ser2843 and S2841 which are known to constitute secondary phosphorylation sites in the GSK3-mediated phosphorylation cascade. Measurements of the Solvent Accessible Surface Area (SASA) calculated using the EDTSurf algorithm (iCn3D, NIH) revealed that phosphorylation of Ser2849 would significantly increase the exposure of secondary sites Ser2843 (by ten-fold) and Ser2841 (by two-fold) to potential kinases which, statistically, may putatively increase the probabilities of their phosphorylation (Fig. 7k and l; Table 4). However Ser2845, which is the closest serine to Ser2849 and which is reportedly the next residue to be phosphorylated in the GSK3 cascade (based on in vitro studies), appears to be less exposed upon Ser2849 phosphorylation. Only when Arg2834 is methylated (in combination with Ser2949 phosphorylation), the SASA of Ser2845 seems to increase dramatically, potentially promoting its phosphorylation (Fig. 7k and l; Table 4). As a result, the combined SASA of Ser2841, Ser2843 and Ser2845 (three major serine residues that are phosphorylated after Ser2849) is predicted to increase as well and reach its maximum level when both Arg2834 is methylated and Ser2849 is phosphorylated (Fig. 7m; Table 4). These structural insights closely match in vitro data showing that in the absence of Arg2834 methylation, even in the presence of phospho-Ser2849, Ser2845 cannot be phosphorylated in vitro, effectively stopping the entire phosphorylation cascade⁹. This is particularly relevant considering that in arrhythmogenic cardiomyopathy Arg2834 is mutated and cannot be methylated. Interestingly, it seems that when the mutant DP CT tail is not undergoing massive post-translational modifications, the overall shape of the mutant protein does not appear to be very different from that of the WT unmodified DP CT tail, maintaining its antiparallel β -sheet structure and general hairpin organization (Fig. 8a and b). This is consistent with the fact that the R2834H mutation is ultimately compatible with life and does not induce death at birth for instance. Moreover, the regulatory (REG) region of mutant DP CT tail remains located within a β -turn structure that is exposed to kinases and suitable for phosphorylation. This observation is

in agreement with in vitro data showing that mutant R2834H desmoplakin protein can still be phosphorylated on Ser2849 in vitro⁹ (Fig. 8c, d). Such apparent similarities between WT and mutant R2834H desmoplakin in the absence of PTMs may stem from the fact that, while histidine is only partially positive at cellular pH, cardiac contractions are known to transiently induce a slightly acidic environment in cardiomyocytes⁴⁰. This may increase the positive charge of histidine, potentially minimizing the effects of the R2834H mutation in the cardiac tissue (Fig. 8d). Even without considering this crucial factor, titration curve analyses showed that the mutant R2834H DP CT tail presents a net positive charge of +3.2 (at pH 7.4) slightly down from 4.2 in the case of WT DP CT tail (Fig. 8e). This would still allow the mutant DP CT tail to fold back onto the negatively charged PLEC14 of PRD domain C and effectively recruit desmin rod domains even in patients with arrhythmogenic cardiomyopathy (Fig. 8d, e). However, predictive modeling seems to suggest that the phosphorylation of Ser2849 in the mutant R2834H DP CT tail would not be capable of inducing a shift in the secondary structure with the absence of an α -helix formation for instance (Fig. 8f, g). Consequently, this would prevent the exposure of secondary phosphorylation sites Ser2841, Ser2843 and Ser2845 (Fig. 8h, i), eventually impairing the entire PTM serine phosphorylation cascade as reported by in vitro studies⁹.

Virtual screening of FDA-approved drug libraries identifies several compounds for potential treatment of cardiocutaneous diseases (drug repurposing strategy)

Finding a therapeutic approach to treating R2834H-induced cardiocutaneous diseases is highly challenging. Whilst promotion of the phosphorylation cascade by increasing the activity of certain kinases may appear promising, the outcomes of such a strategy ultimately remain uncertain, since dramatically increasing the pool of cytoplasmic FM/FP desmoplakin (unbound to desmin) for an extended period of time could be detrimental to the formation of desmosomes. On the other hand, preventing the phosphorylation cascade, which is already reduced in R2834H cardiomyocytes, may worsen the disease by increasing the aberrant recruitment of cytoplasmic mutant DP to desmin intermediate filaments instead of desmosomes. What is required is a strategy that allows mutant desmoplakin to both translocate to desmosomes and bind desmin rod domains there, just like the WT desmoplakin protein. According to this concept, the mutant DP CT tail should remain in a “neutral” position where it is incapable of permanently folding back onto the PLEC14 of PRD domain C (which is predicted to be the main reason for the aberrant binding of mutant DP CT tail to cytoplasmic desmin rod domains). This would in theory free up a large amount of desmoplakin proteins and increase their chance of being recruited by desmosomes. In addition, the “neutral” position of the mutant DP CT tail should also prevent it from folding back onto the PLEC17 of PRD domain C (which would completely annihilate any chance of recruiting desmin intermediate filaments to desmosomes). This would maintain a certain degree of interaction with desmin filaments, which would be needed once mutant desmoplakin finally reaches desmosomes. While challenging, these two conditions could perhaps be fulfilled simultaneously by using bulky compounds with a high affinity towards the mutant DP CT tail. Such bulky molecules ectopically “patched” onto mutant DP CT tail could in theory (i) prevent positively charged mutant DP CT tail from folding back onto negative PLEC14 of PRD domain C by inducing a substantial steric hindrance while also masking as much as possible the arginine-bound positive charges on the surface of the DP CT tail (since arginine methylation cannot occur in mutant DP, the next best thing might be to hide those positive charges). They could also (ii) prevent the DP CT tail from folding back onto the PLEC17 of PRD domain C by steric hindrance while also masking as much as possible Ser2849 to prevent its phosphorylation, which would bring negative charges that could be detrimental to the recruitment of desmin filaments to desmosomes (Fig. 9a). To investigate the possibility of such effects whilst retaining a degree of clinical relevance, we performed virtual screening (VS) of FDA-approved drug libraries (PubChem, NIH) aiming to find compounds that are already used in the clinic and which may be useful to treat cardiocutaneous diseases in innovative drug repurposing strategies. This virtual drug screening was conducted as indicated by the pipeline depicted in Fig. 9b with a series of filters to narrow down the number of potential drug candidates from 5317 FDA-approved drugs down to only 3 (i.e. Top 3). Figure 9c shows the different scenarios encountered during the virtual screening with drugs that were either off target (did not cover Ser2849), or indeed capable of masking Ser2849 (partially or fully). Only drugs that at least partially mask Ser2849 were selected. The results of the screening (Top 60) are shown in Fig. 9d. The drug candidates were ranked by their binding affinity towards mutant DP CT tail (kcal/mol, first column on the left with the best compounds highlighted in blue) and can be identified through their CID numbers and brand names. Other features such as molecular weight (g/mol) (the higher the better since large bulky compounds are desired for maximum steric hindrance), Ser2849 masking (%), the higher the better) and remaining SASA of all arginine residues combined after drug interaction ($\text{\AA}^2$, the lower the better) are also shown in Fig. 9d. Using the remaining SASA of all arginine residues as the sole determining factor, it seems that the known drug Natamycin repeatedly appears as one of the major drug candidates that could significantly reduce the exposure of arginine residues, potentially masking detrimental positive charges on the surface of the mutant DP CT tail which could compensate for the lack of arginine methylation (Fig. 9e and f). Importantly, the high number of different docking positions for Natamycin provides a high degree of flexibility and relevance to our method. Although the predictive model of mutant DP CT tail may not be entirely accurate, the versatility of Natamycin docking leaves room for a certain degree of inexactitude: while one docking position may not actually occur in the cellular context, another may. The drugs Kineret and Enasidenib also appear to be quite effective as well, although they display only a limited amount of docking possibilities that could achieve this purpose. Using Ser2849 masking as the sole determining factor, it appears that Natamycin, and to a lower extent Enasidenib, might again be among the best drug candidates (Fig. 9g, h). When performing 2D plots that cross-check two distinct factors to better segregate the data and identify the best compounds based on multiple factors, Natamycin again appears as one of the best drug candidates that satisfies most requirements (Fig. 9i–m), especially when we compare the ability to mask arginine residues and the ability to mask Ser2849 (Fig. 9n). Natamycin has a high molecular weight (which is highly desirable since bulkiness is key here) as well as a high



binding affinity towards mutant DP CT tail (which is directly linked to the high molecular weight). Because of its large size, Natamycin is more likely than other compounds to mask serine and arginine residues on the surface of the mutant DP CT tail. However, the size of the molecule cannot fully explain its efficacy. Acarbose and Lifitegrast, which have similar molecular weights, do not appear to be as effective as Natamycin in masking key residues or binding mutant DP CT tail. Last but not least, the 17 different docking positions of Natamycin (a feature that is unique to this drug) undeniably make it more reliable than other compounds (Fig. 9o). Figure 9p provides key examples of docking positions for Natamycin, Kineret and Enasidenib on mutant DP CT tail as well as related measurements of Ser2849 masking and remaining SASA of arginine residues.

◀**Fig. 7.** Predicted effects of Ser2849 phosphorylation and Arg2834 methylation on the secondary structure of wildtype DP CT tail and the exposure of secondary phosphorylation sites to potential kinases. **(a)** Color-coded molecular surface display of wildtype DP CT tail (PEPFOLD model) showcasing the exposed nature of the regulatory region (delimited by 3D box) and its known phosphorylation site Ser2849 due to the formation of a β -turn in the hairpin structure. Cyan: Ser2849; Magenta: Arg2846, Arg2847; Blue: 3YF region; Yellow: GSRX region. **(b)** 3D molecular surface display combined with a color-coded Wimley-White hydrophobicity scale for the simultaneously visualization of hydrophobic (green), polar (yellow) and charged (blue: positive, red: negative) exposed areas in the β -turn of WT DP CT tail. The phosphorylation of Ser2849 (Cyan) is represented by the transfer of a letter “P” drawn in a red circle within an area extensively rich in positively charged (blue) arginine residues. **(c)** Predicted changes in chemical bonds involving Ser2849 after its phosphorylation (mimicked by S2849D substitution). Green lines: H-bonds. Cyan lines: salt bridge. ICn3D (NIH). **(d)** Detailed representation of the interactions between pSer2949 (S2849D) and Arg2838 (H-bond), Arg2846 (salt bridge) and Asp2851 (H-bond). **(e)** Predicted changes in chemical bonds involving Asp2851 after Ser2849 is phosphorylated (mimicked by S2849D substitution). Green lines: H-bonds. Cyan lines: salt bridge. **(f)** Detailed representation of the interactions between Asp2851 and Arg2838 (salt bridge), Arg2842 (salt bridge) and Arg2834 (2 H-bonds and 1 salt bridge). **(g)** Interaction map summarizing the interactions and distances between negatively charged (red) residues (pSer2849 and Asp2851) and positively charged (blue) residues (Arg2834, Arg2838, Arg2842 and Arg2846). (1) triggering event: Ser2849 phosphorylation; (2) resulting event: rapprochement and Arg2834-Asp2851 triple bonding. **(h)** Predicted changes in chemical bonds involving Arg2834 after Ser2849 is phosphorylated (mimicked by S2849D substitution). Green lines: H-bonds. Cyan lines: salt bridge, Red lines: Pi-cation. **(i)** PEP-FOLD-predicted changes in the secondary structure of DP CT tail after Ser2849 phosphorylation. Green: β -sheet; red: α -helix. **(j)** Detailed illustration of the α -helix formed after Ser2849 phosphorylation. Green dashed lines: H-bonds. **(k–m)** Solvent Accessible Surface Area (SASA) of Ser2841, Ser2843 and Ser2845 after post-translational modifications of DP CT tail (calculated using EDTSurf algorithm, iCn3D, NIH). Blue *: Ser2845’s SASA without PTM; Red *: Ser2845’s SASA after S2849 phosphorylation; Green *: Ser2845’s SASA after S2849 phosphorylation and Arg2834 methylation.

DP	WT		S2849D		S2849D + R2834F		S2849D + R2834H		R2834H	
	SASA (Å ²)	% out	SASA (Å ²)	% out	SASA (Å ²)	% out	SASA (Å ²)	% out	SASA (Å ²)	% out
Main P. sites										
Ser2833	68.65	25	68.67	25	14.83	5	64.58	24	84.41	31
Ser2837	50.62	19	4.59	1	82.97	31	52.33	19	82.21	31
Ser2839	109.26	41	111.58	42	8.09	3	98.74	37	76.27	28
Ser2841	52.35	19	107.69	40 (↑)	107.81	40 (↑)	29.81	11 (↓)	55.30	20
Ser2843	12.93	4	105.31	39 (↑)	103.24	38 (↑)	43.58	16 (↓)	41.04	15
Ser2845	75.47	28	17.82	6	87.76	33 (↑)	79.45	29 (↓)	65.73	24
Ser2841-Ser2845	140.75		230.82		298.81		152.84		162.07	
Other P. sites										
Ser2831	35.69	13	47.07	17	66.10	24	57.54	21	90.69	34
Ser2835	55.95	21	61.88	23	9.46	3	111.10	41	46.11	17

Table 4. Exposure of known GSK3 phosphorylation sites (P. sites) in wildtype DP CT tail (WT), phospho-mimetic wildtype DP CT tail (S2849D), phospho-mimetic/methyl-mimetic wildtype DP CT tail (S2849D + R2834F), phospho-mimetic mutant DP CT tail (S2849D + R2834H) and mutant DP CT tail (R2834H). P. sites refer to serine residues that are known to constitute secondary phosphorylation sites in the GSK3 phosphorylation cascade initiated by S2849 phosphorylation, as described by Albrecht et al.⁶. SASA: Solvent Accessible Surface Area calculated using the EDTSurf algorithm (iCn3D, NIH). Percentage out: percentage of the residue that is “outside” the molecule and exposed to potential kinases.

Discussion

In this study, we investigated the structural impacts of a known pathogenic mutation, R2834H, in the etiology of arrhythmogenic cardiomyopathy, as well as the implications of PTMs such as serine phosphorylation and arginine methylation in the disease process. In order to make any progress in understanding the disease mechanisms, we first needed to understand key molecular mechanisms that regulate the interactions between DP CT tail and desmin intermediate filaments in a physiological context. Structural models that we generated predicted that when WT desmoplakin is bound to desmosomes, its DP CT tail is likely free of PTMs and will display a high positive net charge (notably due to the presence of numerous arginine residues) which will promote its interaction with the negatively charged PLEC14 of neighboring PRD domain C, and consequently enable the recruitment of negatively charged desmin rod domains to desmoplakin-rich desmosomes (Fig. 10). Modelling a series of PTMs (including serine phosphorylation and arginine methylation) which are likely to occur as part of the natural cardiac tissue remodeling cycle, it could be seen that the WT DP CT tail would progressively

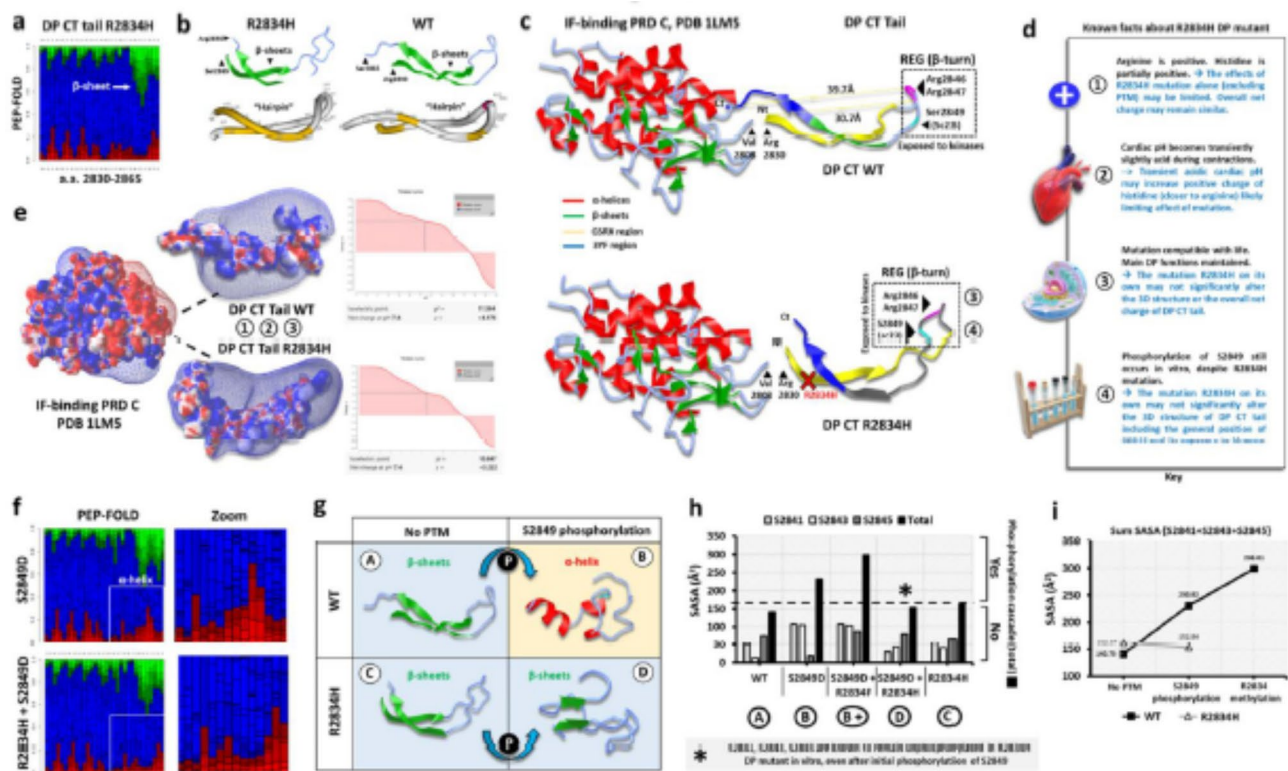


Fig. 8. Predicted effects of R2834H mutation on the secondary structure of DP CT tail and the exposure of secondary phosphorylation sites to potential kinases. **(a)** PEP-FOLD modeling predicting high probabilities of a β -sheet formation in mutant R2834H DP CT tail (a.a. 2830–2865). Green: β -sheet; Red: α -helix. **(b)** Similarities between the predicted antiparallel β -sheets of wildtype (WT) and mutant (R2834H) DP CT tails. The overall “hairpin” conformation is conserved in both proteins. **(c)** Juxtaposition of PRD domain C (PDB 1LM5) and DP CT tail wildtype (WT, PEP-FOLD) or mutant (R2834H, PEP-FOLD). Comparative analysis showed that, in mutant DP CT tail, the GSK3 phosphorylation site Ser2849 (cyan) remains located within a β -turn which also includes Arg2846, Arg2847 (magenta) suggesting that despite the mutation, Ser2848 phosphorylation might still occur in vitro and in vivo which is consistent with known facts ((3), (4)) about mutant DP CT tail summarized in (d). Green: β -sheet; Red: α -helix; Blue: 3YF region; Yellow: GSRX region. REG: regulatory region. **(d)** Known facts about mutant DP CT tail supporting PEP-FOLD predictions. Numbers (1), (2), (3) and (4) refer to specific facts that are closely related to matching observations in PEP-FOLD models of WT and mutant DP CT tails displayed in (c, e). **(e)** DelPhi equipotential maps (iCn3D, NIH) and titration curves (Prot Pi) of WT and mutant DP CT tails showing similar positive charges. Numbers (1), (2) and (3) refer to known facts detailed in (d) that are closely related to matching observations in PEP-FOLD models of WT and mutant DP CT tails. DelPhi settings: potential contour set at 2kT/e (25.6mV at 298 K) with salt concentration at 0.15 M. Titration curves and net charges at pH7.4 from Prot pi (<https://www.protpi.ch/Calculator/ProteinTool#Results>). **(f, g)** PEP-FOLD modeling predicting high probabilities of an absence of α -helix in phosphoSer2849 mutant DP CT tail (mimicked by R2834H + S2849D) which is supposedly present in phosphoSer2849 wildtype DP CT tail (mimicked by S2849D). Green: β -sheet; Red: α -helix. No PTM: no post-translational modification. The letters A, B, C and D of (g) refer to specific PTM states of WT and mutant DP CT tails which are showcased in (h). **(h, i)** Solvent Accessible Surface Area (SASA) of Ser2841, Ser2843 and Ser2845 in WT and mutant DP CT tails in various states of post-translational modifications. The letters A, B, C and D of (h) refer to specific PTM states of WT and mutant DP CT tails which are showcased in (g). SASA calculated using EDTSurf algorithm, (iCn3D, NIH).

become negatively charged and therefore become unable to bind negatively charged desmin rod domains while being released into the cytoplasm. This soluble pool of cytoplasmic desmoplakin may then be used to form new desmosomes, where all PTMs on the WT DP CT tail would then be removed again in order to connect the newly formed desmoplakin-rich desmosomes with the desmin cytoskeleton. However, our numerous analyses revealed that the mutation R2834H would break this cycle and allow the spontaneous formation of a cytoplasmic mutant desmoplakin with a DP CT tail that is constitutively positively charged due to the reported impairment of the serine phosphorylation cascade associated with the lack of Arg2834 methylation⁹ (Fig. 10). As a result, cytoplasmic mutant desmoplakin with its constitutively positively charged DP CT tail would be largely sequestered by the negatively charged rod domains of desmin intermediate filaments located in the cytoplasm. This would deprive the membrane-anchored desmosomes from critically-needed desmoplakin proteins, which

is consistent with in vitro observations⁹. The resulting poor strength in cell-cell contacts may compromise the natural transmission of mechanical forces and/or chemo-electrical signals between cardiomyocytes, potentially leading to cardiac arrhythmia. In an attempt to discover possible therapeutic approaches for this deadly disease, we conducted virtual screening of FDA-approved drug libraries in the hope of identifying compounds that could be used through drug repurposing strategies. To date we have identified several compounds such as Natamycin, Kineret and Enasidenib as potential drug candidates for treatment of R2834H-induced arrhythmogenic cardiomyopathy. However these promising findings will need to be validated in vitro and in preclinical studies before clinical trials can be considered.

Materials and methods

PEP-FOLD algorithm

PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino acid sequences^{28–33}. It is available at <https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/>. We here summarize below the supporting information regarding PEP-FOLD concepts, development and validation that is available at the weblink provided above. Essentially, this method, based on structural alphabet SA letters to describe the conformations of four consecutive residues, couples the predicted series of SA letters to a greedy algorithm and a coarse-grained force field. The structural alphabet is an ensemble of elementary prototype conformations able to describe the whole diversity of protein structures⁴¹. The enhanced greedy algorithm that assembles HMM-SA letters is described in³³. The default short simulations correspond to 100 model generation runs (sufficient to identify the correct fold). Once generated, models are clustered using Apollo⁴² to identify groups of similar models. In the clustering, 10 features are evaluated including avg, gdt (predicted GDT_TS for Global Distance Test Total Score, predictioncenter.org), max, q (predicted Qmean score), tm (predicted TM score) which are the scores predicted by Apollo⁴². The clusters are then sorted using either the sOPEP energy value (coarse grained energy of PEP-FOLD), or Apollo predicted TMscore (tm). The cluster representatives correspond to the models of the clusters having the best scores, i.e. with the lowest sOPEP energy (resp. highest tm value). For peptides up to 36 residues, sOPEP will result in proposing native or near native conformations. For the coarse grained force field, the OPEP potential (Optimized Potential for Efficient structure Prediction) helps limit the roughness of the peptides energetic landscape, by simplifying side chains representation by a single bead. OPEP v3 parameters are optimized by a genetic algorithm procedure using a large ensemble of protein decoys⁴³. OPEP is the objective function that drives the greedy algorithm during the rebuilding process.

OPEP version 3 is expressed as a sum of local, nonbonded and hydrogen-bond (H-bond) terms:

$$E = E_{\text{local}} + E_{\text{nonbonded}} + E_{\text{H-bond}}$$

The local potentials are expressed by:

$$E_{\text{local}} = w_b \sum_{\text{bonds}} K_b (r - r_{eq})^2 + w_a \sum_{\text{angles}} K_a (\alpha - \alpha_{eq})^2 + w_{\Omega} \sum_{\text{impr-torsions}} k_{\Omega} (\Omega - \Omega_{eq})^2 + w_{\phi, \psi} \left(\sum_{\phi} E_{\phi} + \sum_{\psi} E_{\psi} \right)$$

The term Elocal contains force constants associated with changes in bond lengths and bond angles of all particles as well as force constants related to changes in improper torsions of the side-chains and the peptide bonds.

The nonbonded potentials are expressed by:

$$E_{\text{nonbonded}} = w_{1,4} \sum_{1,4} E_{VdW} + w_{C\alpha,C\alpha} \sum_{C\alpha,C\alpha} E_{VdW} + w_{1>4} \sum_{M',M''} E_{VdW} + w_{1>4} \sum_{M',C\alpha} E_{VdW} + w_{1>4} \sum_{M',Sc} E_{VdW} + w_{Sc,Sc} \sum_{Sc,Sc} E_{VdW}$$

with 1, 4 the 1–4 interactions along each torsional degree of freedom, M' the N, C, O and H main chain atoms, and Sc the side-chain. Short-range interactions are separated from long-range ($j > i + 4$) interactions, and the C alpha atom from the other main chain atoms. More information on EVdW potential can be found in previous reports^{42,44}.

The hydrogen-bonding potential (EH – bond) consists of two-body (EHB1) and four-body (EHB2) terms. Two-body H-bonds are defined by:

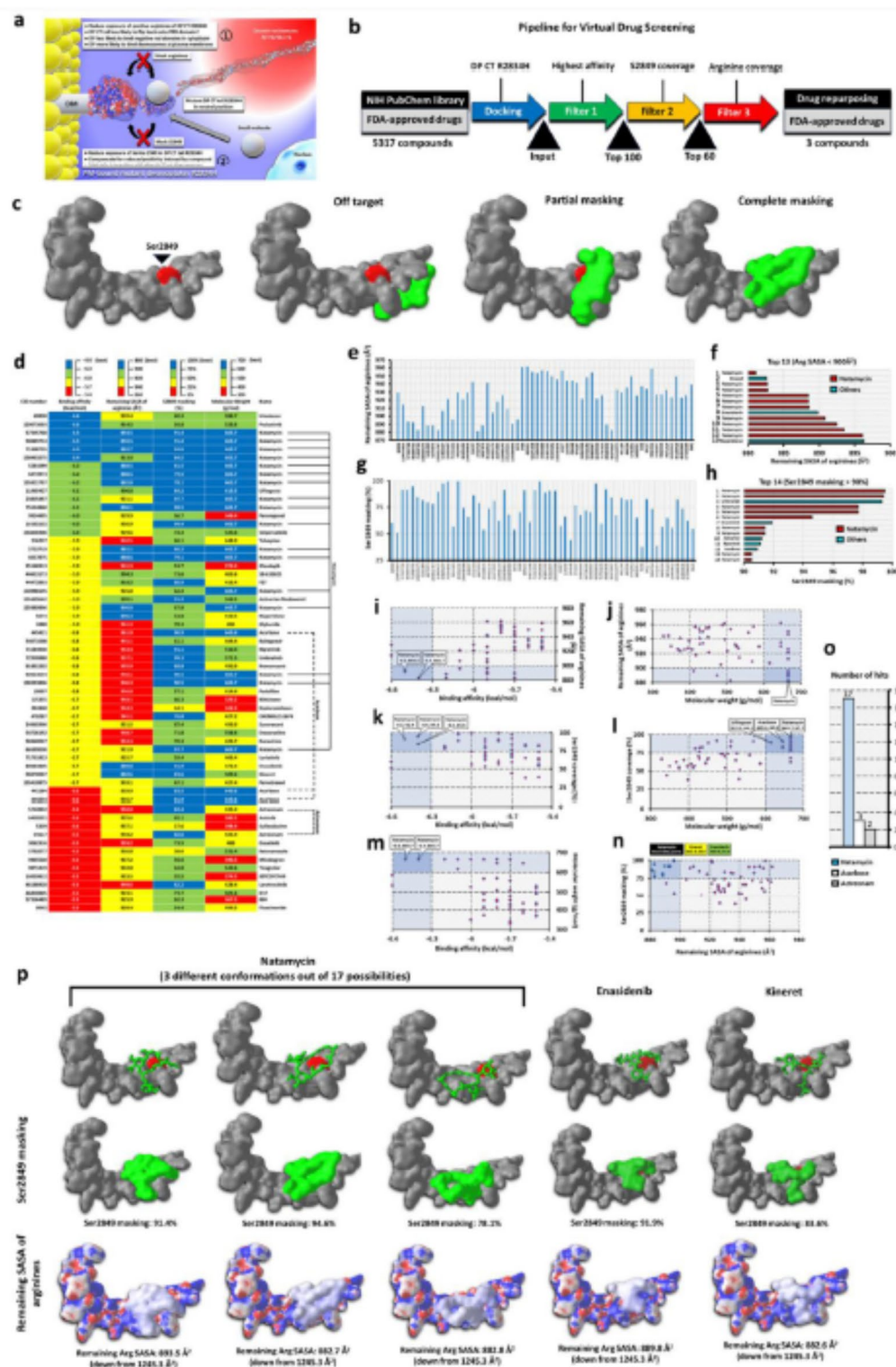
$$E_{\text{HBI}} = w_{\text{HBI-4}} \sum_{i,j>i+4} \epsilon_{\text{HBI-4}} \mu(r_{ij}) \nu(\alpha_{ij}) + w_{\text{HBI}>4} \sum_{i,j>i+4} \epsilon_{\text{HBI}>4} \mu(r_{ij}) \nu(\alpha_{ij})$$

Four-body effects, which represent cooperative energies between hydrogen bonds ij and kl , are defined by:

$$E_{\text{HBI2}} = \sum_{\alpha} \epsilon_{\alpha}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2) \Delta(ijkl) + \sum_{\beta} \epsilon_{\beta}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2) \Delta'(ijkl)$$

All details about the OPEP force field used in PEP-FOLD are available in previous reports^{33,43}.

During the validation process, validation tests are performed using two different peptide sets namely, (i) the “pepLook cyclic peptide set” which includes 34 peptides containing disulfide bonds characterized by NMR spectroscopy⁴⁴ with an average RMSd for the best model of 2.75 Å, and (ii) the “pepStr set”, which includes 42 linear bioactive peptides free of any disulfide bridge characterized by NMR spectroscopy in both aqueous and non-aqueous solutions⁴⁵. This is followed by an additional test using a “PEP-FOLD set” constituted of two subsets of PDB structures solved in aqueous solution, selected for their sizes and topology diversity as described below:



Short peptides: PDB codes (10 targets from 10 to 23 aa) : 1dep, 1k43, 1le1, 1le3, 1pei, 1uao, 1wbr, 1wz4, 2evq and the beta hairpin fragment of 2gb1.

Long peptides: PDB codes (14 targets from 27 to 49 aa) : 1abz, 1aie, 1bb1, 1bdd, 1e0l, 1e0n, 1f4i, 1fsd, 1i6c, 1kjk, 1psv, 1vii, 1vpv and 2p81.

Web-based 3D structure viewer iCn3D

Web-based 3D structure viewer iCn3D (NCBI) can be accessed at the following website: <https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>.

After uploading the PDB file of our PEP-FOLD model of DP CT Tail, the identification of intramolecular interactions was achieved by using the “Interactions” tool in the “Analysis” tab of iCn3D. The thresholds used to determine the interactions were the ones set as default in iCn3D, in particular, 3.8Å for H-bonds, 6Å for π -cation

◀ **Fig. 9.** Virtual screening of FDA-approved drug libraires identifies several promising drug candidates for treatment of R2834H-induced cardiocutaneous diseases. **(a)** Rationale of drug-repurposing strategy based on the use of bulky FDA-approved compounds with high affinity towards mutant DP CT tail to maintain it in a neutral position (binding neither PLEC14 nor PLEC17) through the masking of positively charged arginine residues and the known GSK3 phosphorylation site Ser2949. **(b)** Pipeline detailing the workflow of the virtual screening (VS) and the different filters used to narrow down the number of drug candidates from 5317 to the 3 best compounds. **(c)** Several examples of molecular docking illustrating the different types of Ser2849 masking obtained during VS. Only drugs masking at least partially Ser2849 were selected. Grey: R2834H DP CT tail (PEP-FOLD). Red: Ser2849. Green: FDA-approved compound (PubChem, NIH). **(d)** Heat map displaying the top 60 candidates sorted by binding affinity (kcal/mol, left column) towards mutant DP CT tail. Other parameters include: remaining SASA of arginines ($\text{\AA}^2$), Ser2849 masking (%) and molecular weight (g/mol). SASA calculated using EDTSurf algorithm, (iCn3D, NIH). For each parameter, the best values are highlighted in blue, the worst values in red. Each drug can be identified through his CID number (PubChem, NIH) (left) and brand name (right). Note that a same compound can appear several times due to different docking positions. **(e)** Arginine-masking properties of top 60 candidates. **(f)** Top 13 candidates that are the most capable of masking arginine residues in mutant DP CT tail (remaining arginine SASA $< 900 \text{ \AA}^2$). **(g)** Ser2849-masking properties of top 60 candidates. **(h)** Top 14 candidates that are the most capable of masking Ser2849 in mutant DP CT tail (masking $> 90\%$). **(i-n)** 2D scatter plots comparing two distinct parameters at a time namely: binding affinity (kcal/mol), remaining SASA of arginines ($\text{\AA}^2$), Ser2849 masking (%) and molecular weight (g/mol). The best candidates are located in the dark blue zone with respective labels. **(o)** Number of hits referring to the number of times a same compound appears in the output of VS (number of docking positions). **(p)** Key examples of top 3 FDA-approved drug candidates in docking positions on mutant R2834H DP CT tail. Respective abilities in Ser2849 masking and arginine masking are indicated. Grey: R2834H DP CT tail (PEPFOLD). Red: Ser2849. Green: FDA-approved compound (PubChem, NIH). DelPhi surface with potential in the last row obtained from iCn3D (NIH) with following settings: potential contour set at $2kT/e$ (25.6mV at 298 K) with salt concentration at 0.15 M.

interactions and $6\text{\AA}$ for salt bridge interactions. The 3D molecular surface display combined with a color-coded Wimley-White hydrophobicity scale to simultaneously visualize all hydrophobic, polar and charged exposed areas was obtained by selecting “Wimley-White” in the “Hydrophobicity” tab of the “Color” tab of iCn3D, and then by selecting “Molecular surface” in the “Surface type” tab of the “Style” tab of iCn3D. Solvent Accessible Surface Area (SASA) values were obtained by selecting “Surface area” in the “Analysis” tab of iCn3D. DelPhi equipotential maps and surfaces with potential were also obtained from iCn3D (NIH) by using the “Analysis” tab and selecting “DelPhi potential”, with the subsequent settings: potential contour set at $2kT/e$ (25.6mV at 298 K) with salt concentration at 0.15 M. Grid size: 65.

Hydrophobicity plots and polarity plots

Hydrophobicity plots (Hopp & Woods scale) were obtained from <http://www.pepcalc.com/>. Polarity plots (Zimmerman scale) were obtained from <https://web.expasy.org/protscale/>.

Titration curves

Titration curves and net charges at pH7.4 were obtained by inputting relevant sequences into Prot pi (<https://www.protpi.ch/Calculator/ProteinTool#Results>).

Virtual drug screening

Virtual screening (VS) was achieved by first uploading the PEPFOLD model of mutant R2834H DP CT tail (in .pdb format) into the virtual screening tool PyRx and labeling it as “AutoDock Macromolecule” (AutoDock → Make macromolecule). The library of FDA-approved drugs (in .sdf format) here referred to as “Ligands” was downloaded from PubChem (NCBI, NIH) and imported into PyRx (Controls → Open Babel → Insert New Item → PubChem_Compound_list.sdf). Energy minimization was achieved by selecting “Minimize All” with the following default parameters: Force Field: uff; Optimization Algorithm: Conjugate Gradients; Total number of steps: 200; Number of steps for update: 1; Stop if energy difference is less than: 0.1. Subsequently, all ligands with minimized energy where converted into AutoDock Ligands in pdbqt format by clicking “Convert All to AutoDock Ligand (pdbqt)”. After adjusting the Vina Search Space (ROI) to be inclusive of Ser2849, virtual drug screening was achieved by clicking “Run Vina” in the Vina Wizard tab and the results were saved under .csv format. As depicted on Fig. 9b, the docking results were then ranked based on their binding affinity (Kcal/mol) (Filter 1) and the Top 100 with the lowest energy levels were selected. Subsequently, only docking results from the Top 100 that were capable of at least partially masking Ser2849 were considered (Filter 2). This reduced the number of potential candidates down to 60 (Top 60, Fig. 9b). Each of these docking results from the Top 60 was then saved as a .sdf file and imported into iCn3D (and then saved as .pdb) for further analyses including measurement of the Solvent Accessible Surface Area (SASA) of Ser2849 and several arginine residues (Arg2830, Arg2838, Arg2842, Arg2846, Arg2847 combined) in mutant DP CT tail, calculated using the EDTSurf algorithm from iCn3D (NIH). Through rigorous selections illustrated in Fig. 9d-p, these analyses eventually led to the discovery of 3 major drug candidates (Top 3) for putative treatment of R2834H-induced cardiocutaneous diseases including arrhythmogenic cardiomyopathy.

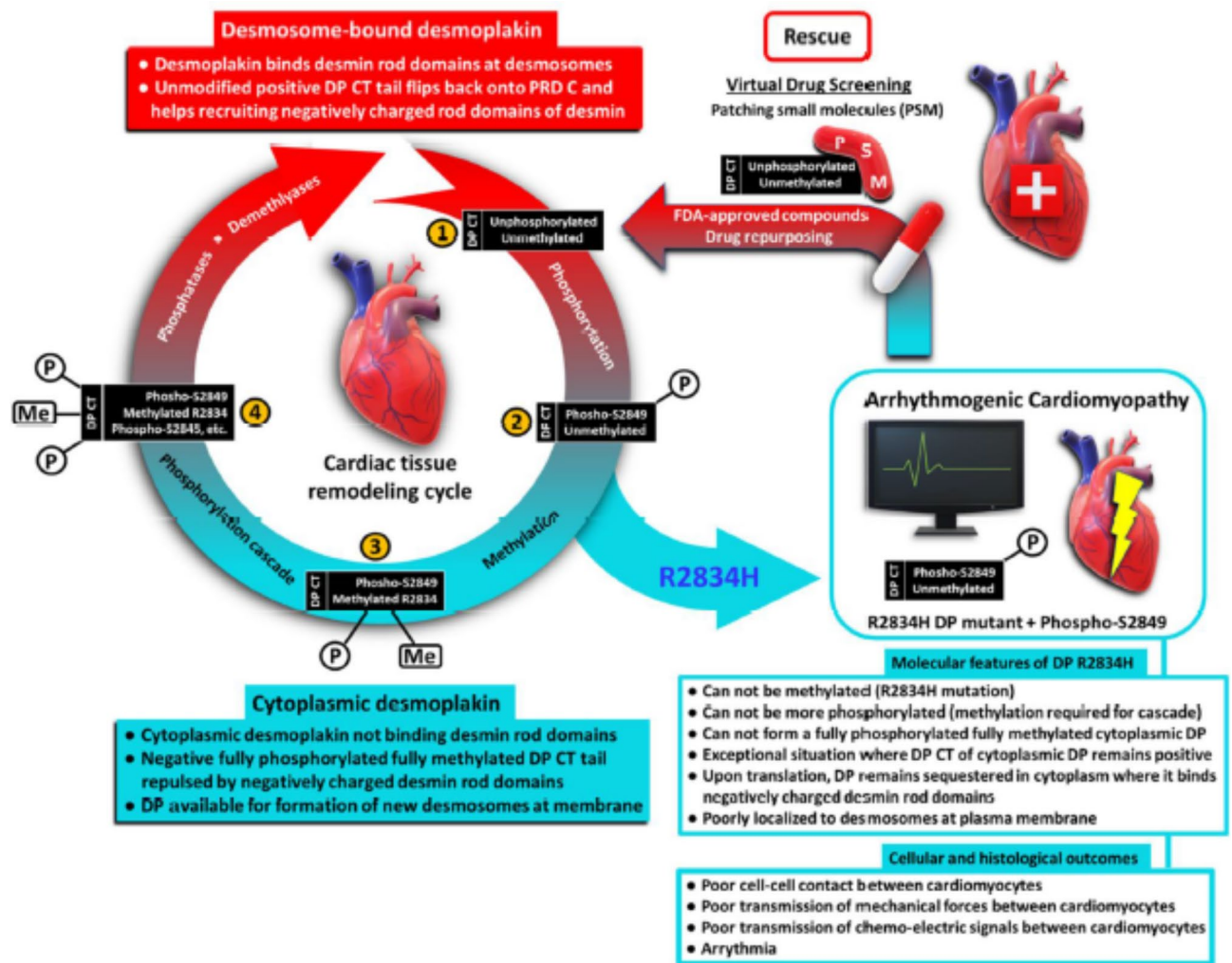


Fig. 10. Graphical summary detailing the key findings of this study placed in the physiological context of normal cardiac tissue remodeling and pathological setting of arrhythmogenic cardiomyopathy. This model is based on original structural data from PEP-FOLD de novo predictive modeling of wildtype and mutant R2834H DP CT tail (a.a. 2830–2865) as well as known in vitro data from previous studies⁹. PEP-FOLD-designed mutant DP CT tail was used for molecular docking in the virtual screening of FDA-approved drug libraries (PubChem, NIH) aiming to identify potential drug candidates for treatment of R2834H-induced cardiocutaneous diseases such as arrhythmogenic cardiomyopathy. Detailed explanations can be found in the “Discussion” section. PRD plakoin repeat domain, P phosphorylation, Me methylation.

Data availability

The data generated and/or analyzed in the current study are available from the corresponding author on reasonable request. The websites used to generate or represent the data are provided below: <https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html> <https://www.protpi.ch/Calculator/ProteinTool#Results> <http://www.pepcalc.com/> <https://web.expasy.org/protscale/>. All data generated or analyzed during this study are included in this published article.

Received: 21 May 2024; Accepted: 19 September 2024

Published online: 05 October 2024

References

- Kowalczyk, A. P. et al. The amino-terminal domain of desmoplakin binds to plakoglobin and clusters desmosomal cadherin-plakoglobin complexes. *J. Cell. Biol.* **139**, 773–784. <https://doi.org/10.1083/jcb.139.3.773> (1997).
- Kouklis, P. D., Hutton, E. & Fuchs, E. Making a connection: direct binding between keratin intermediate filaments and desmosomal proteins. *J. Cell. Biol.* **127**, 1049–1060. <https://doi.org/10.1083/jcb.127.4.1049> (1994).
- Bornslaeger, E. A., Corcoran, C. M., Stappenbeck, T. S. & Green, K. J. Breaking the connection: displacement of the desmosomal plaque protein desmoplakin from cell-cell interfaces disrupts anchorage of intermediate filament bundles and alters intercellular junction assembly. *J. Cell. Biol.* **134**, 985–1001. <https://doi.org/10.1083/jcb.134.4.985> (1996).

4. Choi, H. J., Park-Snyder, S., Pascoe, L. T., Green, K. J. & Weis, W. I. Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure. *Nat. Struct. Biol.* **9**, 612–620. <https://doi.org/10.1038/nsb818> (2002).
5. Favre, B. et al. Desmoplakin interacts with the coil 1 of different types of intermediate filament proteins and displays high affinity for assembled intermediate filaments. *PLoS ONE* **13**, e0205038. <https://doi.org/10.1371/journal.pone.0205038> (2018).
6. Albrecht, L. V. et al. GSK3- and PRMT1-dependent modifications of desmoplakin control desmoplakin-cytoskeleton dynamics. *J. Cell. Biol.* **208**, 597–612. <https://doi.org/10.1083/jcb.201406020> (2015).
7. Fontao, L. et al. Interaction of the bullous pemphigoid antigen 1 (BP230) and desmoplakin with intermediate filaments is mediated by distinct sequences within their COOH terminus. *Mol. Biol. Cell.* **14**, 1978–1992. <https://doi.org/10.1091/mbc.e02-08-0548> (2003).
8. Mohammed, F., Trieber, C., Overduin, M. & Chidgey, M. Molecular mechanism of intermediate filament recognition by plakins proteins. *Biochim. Biophys. Acta Mol. Cell. Res.* **1867**, 118801. <https://doi.org/10.1016/j.bbamcr.2020.118801> (2020).
9. Stappenbeck, T. S., Lamb, J. A., Corcoran, C. M. & Green, K. J. Phosphorylation of the desmoplakin COOH terminus negatively regulates its interaction with keratin intermediate filament networks. *J. Biol. Chem.* **269**, 29351–29354 (1994).
10. Godsel, L. M. et al. Desmoplakin assembly dynamics in four dimensions: multiple phases differentially regulated by intermediate filaments and actin. *J. Cell. Biol.* **171**, 1045–1059. <https://doi.org/10.1083/jcb.200510038> (2005).
11. Bass-Zubek, A. E. et al. Plakophilin 2: a critical scaffold for PKC alpha that regulates intercellular junction assembly. *J. Cell. Biol.* **181**, 605–613. <https://doi.org/10.1083/jcb.200712133> (2008).
12. Meng, J. J., Bornslaeger, E. A., Green, K. J., Steinert, P. M. & Ip, W. Two-hybrid analysis reveals fundamental differences in direct interactions between desmoplakin and cell type-specific intermediate filaments. *J. Biol. Chem.* **272**, 21495–21503. <https://doi.org/10.1074/jbc.272.34.21495> (1997).
13. Dehner, C., Rotzer, V., Waschke, J. & Spindler, V. A desmoplakin point mutation with enhanced keratin association ameliorates *Pemphigus vulgaris* autoantibody-mediated loss of cell cohesion. *Am. J. Pathol.* **184**, 2528–2536. <https://doi.org/10.1016/j.ajpath.2014.05.016> (2014).
14. Iyer, V. R. & Chin, A. J. Arrhythmogenic right ventricular cardiomyopathy/dysplasia (ARVC/D). *Am. J. Med. Genet. C Semin. Med. Genet.* **163**, 185–197. <https://doi.org/10.1002/ajmg.c.31368> (2013).
15. Mazurek, S. & Kim, G. H. Genetic and epigenetic regulation of arrhythmogenic cardiomyopathy. *Biochim. Biophys. Acta Mol. Basis Dis.* **1863**, 2064–2069. <https://doi.org/10.1016/j.bbdis.2017.04.020> (2017).
16. Liang, Y. et al. Desmosomal COP9 regulates proteome degradation in arrhythmogenic right ventricular dysplasia/cardiomyopathy. *J. Clin. Invest.* **131**, 689. <https://doi.org/10.1172/JCI137689> (2021).
17. Zhang, J., Liang, Y., Bradford, W. H. & Sheikh, F. Desmosomes: emerging pathways and non-canonical functions in cardiac arrhythmias and disease. *Biophys. Rev.* **13**, 697–706. <https://doi.org/10.1007/s12551-021-00829-2> (2021).
18. Qiu, X. et al. Mutations of plakophilin-2 in Chinese with arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Am. J. Cardiol.* **103**, 1439–1444. <https://doi.org/10.1016/j.amjcard.2009.01.356> (2009).
19. Castelletti, S. et al. Desmoplakin missense and non-missense mutations in arrhythmogenic right ventricular cardiomyopathy: genotype-phenotype correlation. *Int. J. Cardiol.* **249**, 268–273. <https://doi.org/10.1016/j.ijcard.2017.05.018> (2017).
20. Dalal, D. et al. Clinical features of arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in plakophilin-2. *Circulation* **113**, 1641–1649. <https://doi.org/10.1161/CIRCULATIONAHA.105.568642> (2006).
21. Biernacka, E. K. et al. Pathogenic variants in plakophilin-2 gene (PKP2) are associated with better survival in arrhythmogenic right ventricular cardiomyopathy. *J. Appl. Genet.* **62**, 613–620. <https://doi.org/10.1007/s13353-021-00647-y> (2021).
22. Yang, Z. et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ. Res.* **99**, 646–655. <https://doi.org/10.1161/01.RES.0000241482.19382.c6> (2006).
23. Najor, N. A. et al. Epidermal growth factor receptor neddylation is regulated by a desmosomal-COP9 (constitutive photomorphogenesis 9) signalosome complex. *Elife* **6**, 599. <https://doi.org/10.7554/eLife.22599> (2017).
24. Norgett, E. E. et al. Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. *Hum. Mol. Genet.* **9**, 2761–2766. <https://doi.org/10.1093/hmg/9.18.2761> (2000).
25. Pantou, M. P. et al. A truncating variant altering the extreme C-terminal region of desmoplakin (DSP) suggests the crucial functional role of the region: a case report study. *BMC Med. Genom.* **16**, 95. <https://doi.org/10.1186/s12920-023-01527-6> (2023).
26. Choi, H. J. & Weis, W. I. Crystal structure of a rigid four-spectrin-repeat fragment of the human desmoplakin plakins domain. *J. Mol. Biol.* **409**, 800–812. <https://doi.org/10.1016/j.jmb.2011.04.046> (2011).
27. Choi, H. J. & Weis, W. I. Purification and structural analysis of Desmoplakin. *Methods Enzymol.* **569**, 197–213. <https://doi.org/10.1016/bs.mie.2015.05.006> (2016).
28. Maupetit, J., Derreumaux, P. & Tuffery, P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **37**, W498–W503. <https://doi.org/10.1093/nar/gkp323> (2009).
29. Thevenet, P. et al. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **40**, W288–W293. <https://doi.org/10.1093/nar/gks419> (2012).
30. Shen, Y., Maupetit, J., Derreumaux, P. & Tuffery, P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.* **10**, 4745–4758. <https://doi.org/10.1021/ct500592m> (2014).
31. Alland, C. et al. RPBS: a web resource for structural bioinformatics. *Nucleic Acids Res.* **33**, W44–W49. <https://doi.org/10.1093/nar/gki477> (2005).
32. Neron, B. et al. Moby: a new full web bioinformatics framework. *Bioinformatics* **25**, 3005–3011. <https://doi.org/10.1093/bioinformatics/btp493> (2009).
33. Maupetit, J., Derreumaux, P. & Tuffery, P. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.* **31**, 726–738. <https://doi.org/10.1002/jcc.21365> (2010).
34. Bouameur, J. E. et al. Phosphorylation of serine 4,642 in the C-terminus of plectin by MNK2 and PKA modulates its interaction with intermediate filaments. *J. Cell. Sci.* **126**, 4195–4207. <https://doi.org/10.1242/jcs.127779> (2013).
35. Bennett, E. J., Rush, J., Gygi, S. P. & Harper, J. W. Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. *Cell* **143**, 951–965 (2010).
36. Mostaqul Huq, M. D. et al. Suppression of receptor interacting protein 140 repressive activity by protein arginine methylation. *EMBO J.* **25**, 5094–5104. <https://doi.org/10.1038/sj.emboj.7601389> (2006).
37. Badowski, C. et al. Modeling the structure of keratin 1 and 10 terminal domains and their misassembly in keratoderma. *J. Invest. Dermatol.* **137**, 1914–1923. <https://doi.org/10.1016/j.jid.2017.03.038> (2017).
38. Steinert, P. M. et al. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**, 130–139. [https://doi.org/10.1016/0141-8130\(91\)90037-u](https://doi.org/10.1016/0141-8130(91)90037-u) (1991).
39. Steinert, P. M. Structure, function, and dynamics of keratin intermediate filaments. *J. Invest. Dermatol.* **100**, 729–734. <https://doi.org/10.1111/1523-1747.ep12475665> (1993).
40. Lyu, Y. et al. Beat-to-beat dynamic regulation of intracellular pH in cardiomyocytes. *iScience* **25**, 103624. <https://doi.org/10.1016/j.isci.2021.103624> (2022).
41. Camproux, A. C., Gautier, R. & Tuffery, P. A hidden Markov model derived structural alphabet for proteins. *J. Mol. Biol.* **339**, 591–605. <https://doi.org/10.1016/j.jmb.2004.04.005> (2004).
42. Beaufays, J., Lins, L., Thomas, A. & Brasseur, R. In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. *J. Pept. Sci.* **18**, 17–24. <https://doi.org/10.1002/psc.1410> (2012).

43. Maupetit, J., Tuffery, P. & Derreumaux, P. A coarse-grained protein force field for folding and structure prediction. *Proteins* **69**, 394–408. <https://doi.org/10.1002/prot.21505> (2007).
44. Wang, Z., Eickholt, J. & Cheng, J. APOLLO: a quality assessment service for single and multiple protein models. *Bioinformatics* **27**, 1715–1716. <https://doi.org/10.1093/bioinformatics/btr268> (2011).
45. Kaur, H., Garg, A. & Raghava, G. P. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept. Lett.* **14**, 626–631. <https://doi.org/10.2174/092986607781483859> (2007).

Acknowledgements

We would like to thank Microsoft (<https://www.microsoft.com/en-us/>) for their permission to use their 3D models from Paint 3D.

Author contributions

C.B. conceived the project, led the study and generated the 3D models of WT and mutant DP CT tail using PEPFOLD. C.B. also performed the 3D rendering analyses with web-based 3D structure viewer iCn3D (NIH), the net charge analyses with titration curves (with ProtPi) as well as the hydrophobicity and polarity plots (with Pepcalc and Expasy). C.B. also conducted the virtual screening of FDA-approved drug libraires (PubChem, NIH) using PyRx. P.B. carried out the NCBI protein-protein BLAST search which led to the identification of a sequence homologous to desmoplakin in plectin (Fig. 1). P.B. also performed the multiple alignment of desmoplakin sequences from different species which led to the characterization of the 3 main conserved regions of DP CT tail namely: GSRS region, REG region and 3YF region (Fig. 3). P.B. also curated the available PDB files for human desmoplakin (Fig. 1) and plectin (Table 1). C.S.V. provided critical assessment of scientific data on protein structure modelling, especially regarding the effects of post-translational modifications on secondary structures as well as net charge analyses. E.B.L. conceived the project with C.B and provided critical input and ideas that were essential to the successful development of the project. E.B.L. also provided financial support for the study. All authors read and revised the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-73705-0>.

Correspondence and requests for materials should be addressed to C.B. or E.B.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024