Bioinformatics, 36(12), 2020, 3726–3732
doi: 10.1093/bioinformatics/btaa186
Advance Access Publication Date: 14 March 2020
Original Paper

Structural bioinformatics

κ-helix and the helical lock and key model: a pivotal way of looking at polyproline II

Tomer Meirson [1,2,*], David Bomze[3], Gal Markel[2,4] and Abraham O. Samson[1]

[1]Drug Discovery Laboratory, The Azrieli Faculty of Medicine, Bar-Ilan University, Safed 1311502, Israel, [2]Ella Lemelbaum Institute for Immuno-oncology, Sheba Medical Center, Ramat-Gan 526260, Israel, [3]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel and [4]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel-Aviv 6997801, Israel

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Received on July 31, 2019; revised on February 11, 2020; editorial decision on March 11, 2020; accepted on March 12, 2020

Abstract

Motivation: Polyproline II (PPII) is a common conformation, comparable to α-helix and β-sheet. PPII, recently termed with a more generic name—κ-helix, adopts a left-handed structure with 3-fold rotational symmetry. Lately, a new type of binding mechanism—the helical lock and key model was introduced in SH3-domain complexes, where the interaction is characterized by a sliding helical pattern. However, whether this binding mechanism is unique only to SH3 domains is unreported.

Results: Here, we show that the helical binding pattern is a universal feature of the κ-helix conformation, present within all the major target families—SH3, WW, profilin, MHC-II, EVH1 and GYF domains. Based on a geometric analysis of 255 experimentally solved structures, we found that they are characterized by a distinctive rotational angle along the helical axis. Furthermore, we found that the range of helical pitch varies between different protein domains or peptide orientations and that the interaction is also represented by a rotational displacement mimicking helical motion. The discovery of rotational interactions as a mechanism, reveals a new dimension in the realm of protein–protein interactions, which introduces a new layer of information encoded by the helical conformation. Due to the extensive involvement of the conformation in functional interactions, we anticipate our model to expand the current molecular understanding of the relationship between protein structure and function.

Availability and implementation: We have implemented the proposed methods in an R package freely available at https://github.com/Grantlab/bio3d.

Contact: Tomermrsn@gmail.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein secondary structures are repetitive elements with conserved geometric and biophysical properties and are of the utmost importance in many branches of biology. Since Pauling and coworkers proposed the α-helix and the β-sheet in 1951, now known to constitute tens of thousands of proteins, they remain to be the most recognized secondary structures (Eisenberg, 2003). Polyproline II (PPII) helix is yet another fascinating structure which was once considered to be infrequent (Berisio *et al.*, 2006; Creamer and Campbell, 2002; Siermala *et al.*, 2001; Wang *et al.*, 2005), however, the actual prevalence and significance continues to unfold (Adzhubei *et al.*, 2013; Jha *et al.*, 2005; Wang *et al.*, 2005). PPII has an elongated left-handed structure, three residues per turn, and a 3-fold rotational symmetry along the helix axis (Holt and Koffer, 2001).

PPIIs, as indicated by their name, are often enriched with prolines and possess a PxxP motif (Kaneko *et al.*, 2008). However, proline-free PPIIs are almost just as frequent, which evoked a widespread criticism for the misleading name 'polyproline' to the many structures which contain few or none, and calling for a more suitable designation (Adzhubei *et al.*, 2013; Hicks and Hsu, 2004; Hollingsworth *et al.*, 2009; Mansiaux *et al.*, 2011; Martin *et al.*, 2014; Meirson *et al.*, 2020; Milner-White and Russell, 2008). Therefore, 'κ-helix' was proposed as the Greek letter forms three partial triangles. Also, its spelling 'kappa' denotes it is enriched with prolines (aPPa) and requires three residues to complete a turn (AppA) (Meirson *et al.*, 2020).

The κ-helix stands out among secondary structures due to its distinctive structural properties. While κ-helix constructs fibrillar proteins which serve as mechanical support of cells and tissues (Esipova and Tumanyan, 2017; Rhee and Grinnell, 2007), the secondary

structure is markedly more flexible in comparison to the α-helix and β-sheet (Zagrovic *et al.*, 2005). The extended character of κ-helix does not support regular patterns of intrachain hydrogen bonds and is capable of fast conformational changes (Cubellis *et al.*, 2005; Kelly *et al.*, 2001; Mensch *et al.*, 2016; Stapley and Creamer, 1999). The irregular character of κ-helix hydrogen bonds allows it to form intermolecular ligand-target hydrogen bonds and supports the formation of protein–protein interactions (Cubellis *et al.*, 2005; Siligardi and Drake, 1995). The mediation of protein interactions via κ-helix and their recognition motifs are among the best known functional roles of the structure (Adzhubei *et al.*, 2013). Yet, the selective facilitation of protein–protein interactions is discordant with the promiscuous binding of this small and conformationally restricted peptide. This raised speculations on how the conflicting features of the motif can coexist (Agrawal and Kishan, 2002; Ball *et al.*, 2005; Meirson *et al.*, 2020).

The κ-helix is found in protein ligands participating in binding to various protein domains, such as the Src homology 3 (SH3), WW (named after a conserved Trp-Trp motif), profilin, major histocompatibility complex class II (MHC-II), Ena/VASP Homology 1 (EVH1) and glycine-tyrosine-phenylalanine (GYF) domains (Adzhubei *et al.*, 2013; Ball *et al.*, 2005). Recently, we have shown that κ-helix interaction with SH3 domains is governed by the rotation angle of the helix backbone around the helical path (Meirson *et al.*, 2020). The helical motif can be described as a 'screw', while its binding interface on the surface of the target offers a semblance of a 'nut thread' to this screw. Therefore, all SH3/κ-helix complexes can be characterized in a single uniform system in rotational space. These findings introduced a novel model of a lock with a rotating and translating key with no known equivalent machinery in molecular biology (Meirson *et al.*, 2020). In contrast to the classical lock and key model where the binding partners have essentially random spatial arrangements, the proposed model offers a coherent apparatus where the binding partners share a common helical interface. In addition, due to the 3-fold rotational symmetry of κ-helix and the rotary translation, the SH3/κ-helix complexes can be separated by 120° rotation, and stratified into three classes α (0–120°), β (120–240°) and γ (240–360°). Also, κ-helices were found to be characterized by a structural reading frame, containing the PxxP motif, and together with the flanking residues dictate the organization of SH3 domains. One of the major questions arising from these findings is whether the helical lock and key model of κ-helix is unique to SH3 domains or is a more general binding mechanism. Therefore, in this study, we focus on the binding interactions of the major κ-helix binding domains and explore a similar helical association to κ-helix.

2 Materials and methods

2.1 Data collection

To obtain all the major domains complexed with PPII (κ-helix) peptides, we queried the PDB database for the domain names in all species—SH3, WW, profilin, MHC-II, EVH1, GYF and UEV domains (Adzhubei *et al.*, 2013; Ball *et al.*, 2005). Each structure was inspected manually and downloaded if it involved an interaction with a peptide. In total, we identified 255 PDB files (Supplementary Table S1). None of the identified UEV complexes passed the assignment threshold for κ-helix and thus were excluded from further analysis. Crystal structures with several chains and NMR structures with multiple conformations were split into separate complexes. For MHC-II structures with multiple chains, only one complex (α/β chains with a peptide) was used. Peptides with less than six residues in the binding groove or synthetic residues in this interface were excluded. Crystal structures with a resolution of more than 3 Å were also excluded. To analyze the structures, the Bio3D package (Grant *et al.*, 2006) was used with the statistical programming language R (https://www.r-project.org/).

2.2 Binding frame assignment

The six-residue reading frame representing the binding interface was assigned based on the position of the peptide relative to a common structurally conserved residue in each domain. The length of the peptide may be longer (e.g., in MHC-II complexes), however only the first six residues are considered. The residues in the target domains representing the boundary of the ligand (Fig. 1) included Trp99 (PDB: 1ABO), Trp280 (PDB: 2HO2), Tyr35 (PDB: 2KJG), Trp61 (PDB: 1JK8), Gln79 (PDB: 5NC7), Tyr6 (PDB: 1L2Z), in SH3, WW, profilin, MHC-II, EVH1 and GYF domains, respectively. Since κ-helix peptides can be situated in various configurations, we used a uniform nomenclature that is based only on the binding frame. The most N-terminal residue in the binding frame is assigned as position '1' and the remaining positions increment positively toward the C-terminus, which allows a comparable nomenclature that is independent of the peptide orientation.

2.3 κ-helix assignment

To assign κ-helix we used the method which we recently introduced (Meirson *et al.*, 2020). We calculated the root mean square dihedral deviations (RMSdD) of the peptide backbone torsional angles ϕ and ψ as a measure of the average deviation from the reference κ-helix. The RMSdD of ϕ and ψ angles is given by

$$\mathrm{RMSdD}_{\phi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\phi_i - \phi_r)^2} \qquad (1)$$

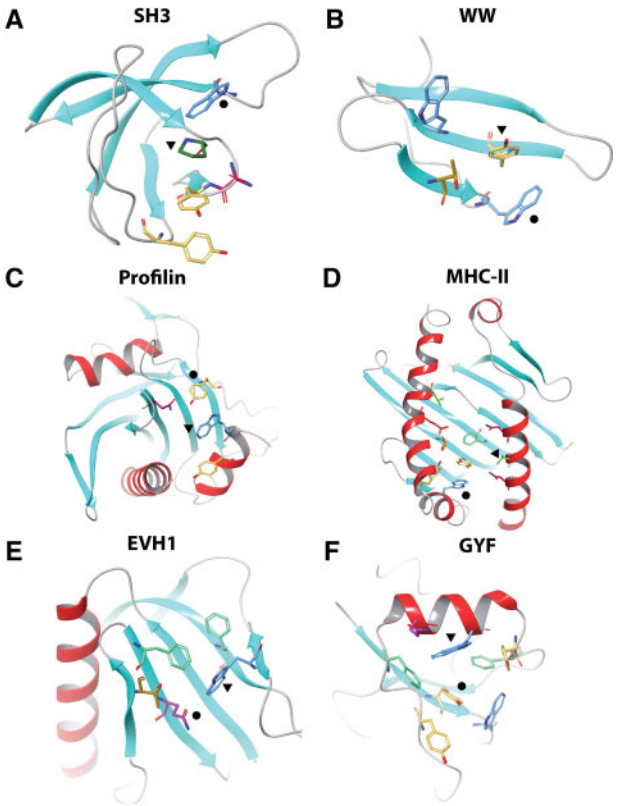

Fig. 1. Three-dimensional structures of representative examples of κ-helix binding domains. (**A**) SH3 domain (PDB: 1ABO). (**B**) WW domain (PDB: 2HO2). (**C**) Profilin (PDB: 2JKG). (**D**) MHC-II (PDB: 1JK8). (**E**) EVH1 domain (PDB: 5ZZ9). (**F**) GYF domain (PDB: 1L2Z). The figure depicts the overall folds of the domains and the location of the peptide-binding site comprising the exposed and aromatic residues. Residues are colored by type. The residues used for binding frame assignment and reference are marked with a circle and triangle, respectively. (Color version of this figure is available at *Bioinformatics* online.)

$$\mathrm{RMSdD}_{\psi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\psi_i - \psi_r)^2} \quad (2)$$

where N is the total number of residues with calculated torsional angles ϕ_i and ψ_i, and ϕ_r and ψ_r are the reference angles of $\phi = -78°$, $\psi = 146°$ (Stapley and Creamer, 1999). The mean RMSdD which incorporates both torsional dihedral angles is then given by

$$\mathrm{RMSdD} = \frac{\mathrm{RMSdD}_{\phi} + \mathrm{RMSdD}_{\psi}}{2} \quad (3)$$

In this study, we used a stricter RMSdD threshold for κ-helix assignment that was set at 70% of the previously calculated cut-off (Meirson *et al.*, 2020).

3 Results

3.1 The universal helical nature of κ-helix

To evaluate whether the helical binding mode of κ-helix is restricted to SH3 domains or is a general phenomenon, we analyzed additional domain families. First, we used only complexes with peptides assigned to κ-helix conformation (Supplementary Table S2). The assignment of the κ-helix secondary structure is based on the dihedral angle deviations of the peptide from an ideal PPII. The alignment of the assigned peptides with their cognate targets are shown in Figure 2. The structural alignments of the complexes demonstrate a rotation coupled with translation (i.e. screw motion) along the helix axis which is best observed in the SH3 domain (Fig. 2A). The rest of the domains contain fewer structures but also depict a similar distribution (Fig. 2B–F). However, WW domain (Fig. 2B) and to a lesser extent EVH1 domain (Fig. 2E) present with an additional rotation that is perpendicular to the binding plane in some of the peptides, suggesting a more versatile binding interface or that the interaction is context dependent (Zhou *et al.*, 2019).

To characterize the interaction formed between the peptides and their cognate targets, we modeled the rotation and the rise of the helix relative to a conserved residue in each of the domains (Fig. 1). This analysis allows us to compare all the complexes uniformly, using a single coordinate system. Figure 3 depicts the 2D projection and 3D conformation of the peptides stratified by orientation and protein family membership. The rotary distributions of the residues are continuous in all the groups. SH3 and WW domains (Fig. 3A and B) have the highest number of complexes and display a complete circle of 360° (each residue position shifts by up to 120°), where the positive orientation of MHC-II which has fewer complexes (Fig. 3D) nearly completes a circle. Contrarily, negative MHC-II, profilin, EVH1 and GYF domains (Fig. 3C–F) possess a small number of solved structures and present a more restricted range of distribution, suggesting that the range correlates with the number of structures and the actual range is broader. Also, the transition between the orientations is characterized by a clockwise or anti-clockwise rotation and is also continuous with a certain gap (Fig. 3A–E). The analysis demonstrates that all the peptides present a gradual rotation along the helical path irrespective of the family domain and their distribution can be described as a corkscrew that is distinguished by the relative rise and rotation. Remarkably, this pattern indicates that the helical binding mode is a general property of κ-helices and is not a unique feature of SH3 domain interactions.

3.2 Characterization of the κ-helix interaction

The interaction formed between κ-helices and their target proteins can also be characterized via a unique rotational angle, which represents their relationship (Fig. 4, Supplementary Fig. S1, Table S3). The range of angles is widest in SH3 and WW domains (Fig. 4A and B), followed by the positive orientations of MHC-II, EVH1 and profilin (Fig. 4C–E). The rest cannot be sufficiently assessed due to the small number of solved structures (Fig. 4C–E, Supplementary Fig. S1). The pitch of the peptide represents the height of one complete helix turn, and the pitch distributions are shown in the lower panels in Figure 4A–E and Supplementary Figure S1. The median pitch of

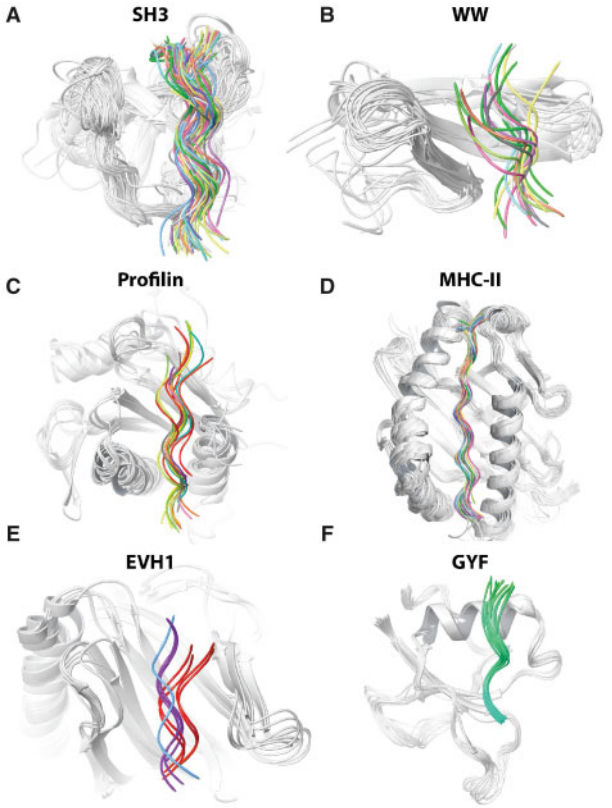

Fig. 2. Structural alignment of κ-helix binding domains in complex with their cognate peptide ligands. (**A**) SH3 domain. (**B**) WW domain. (**C**) Profilin. (**D**) MHC-II. (**E**) EVH1 domain. (**F**) GYF domain; shown is the ensemble of a single GYF domain (PDB: 1L2Z). The figure shows the helical displacement in various degrees relative to the target. Peptide ligands are colored by their PDB ID. (Color version of this figure is available at *Bioinformatics* online.)

the ligands binding to SH3, profilin, EVH1, and GYF domains is around 9 Å. However, the median pitch of the MHC-II helices is 9.57, which is significantly higher than the rest of the domains (Fig. 4F). Furthermore, the pitch distribution of WW domain reveals an interesting pattern in which the opposing orientations are significantly different, but not the rest of the domain families (Fig. 4F). This observation suggests that the classes of WW domains represent two distinct types of pitch selectivity which may impede the formation of dimers, as opposed to the potential dimerization of SH3 domains. The analysis could explain why some domain families dimerize while others do not. However, given that the significance is marginal ($P = 0.04$) more data are required to establish these findings.

Using the analogy of the helical peptide 'screw' that is driven into the protein 'nut', the target can also be represented by two structural features: the side-chain ridges that constitute the 'nut thread', and the scaffold that constitutes the 'nut body'. To evaluate the relationship between these three actors, we calculated the vertical displacement along the helical axis in SH3/κ-helix complexes (Supplementary Fig. S2A). The displacements between the n-SRC loop ('nut thread'), represented by the conserved SH3-Trp and the center ('nut body'), represented by the conserved SH3-Pro, significantly correlate with the displacement of the peptide ('screw'). This correlation is more pronounced in the positive than the negative orientation (Supplementary Fig. S2B).

We further analyzed the physico-chemical interactions, focusing on hydrogen bonds (H-bonds), to generalize the observations about the interaction patterns of κ-helix complexes (Supplementary Materials). The results show different H-bonding profiles between the domains and orientations (Fig. 5), with some consistency in the

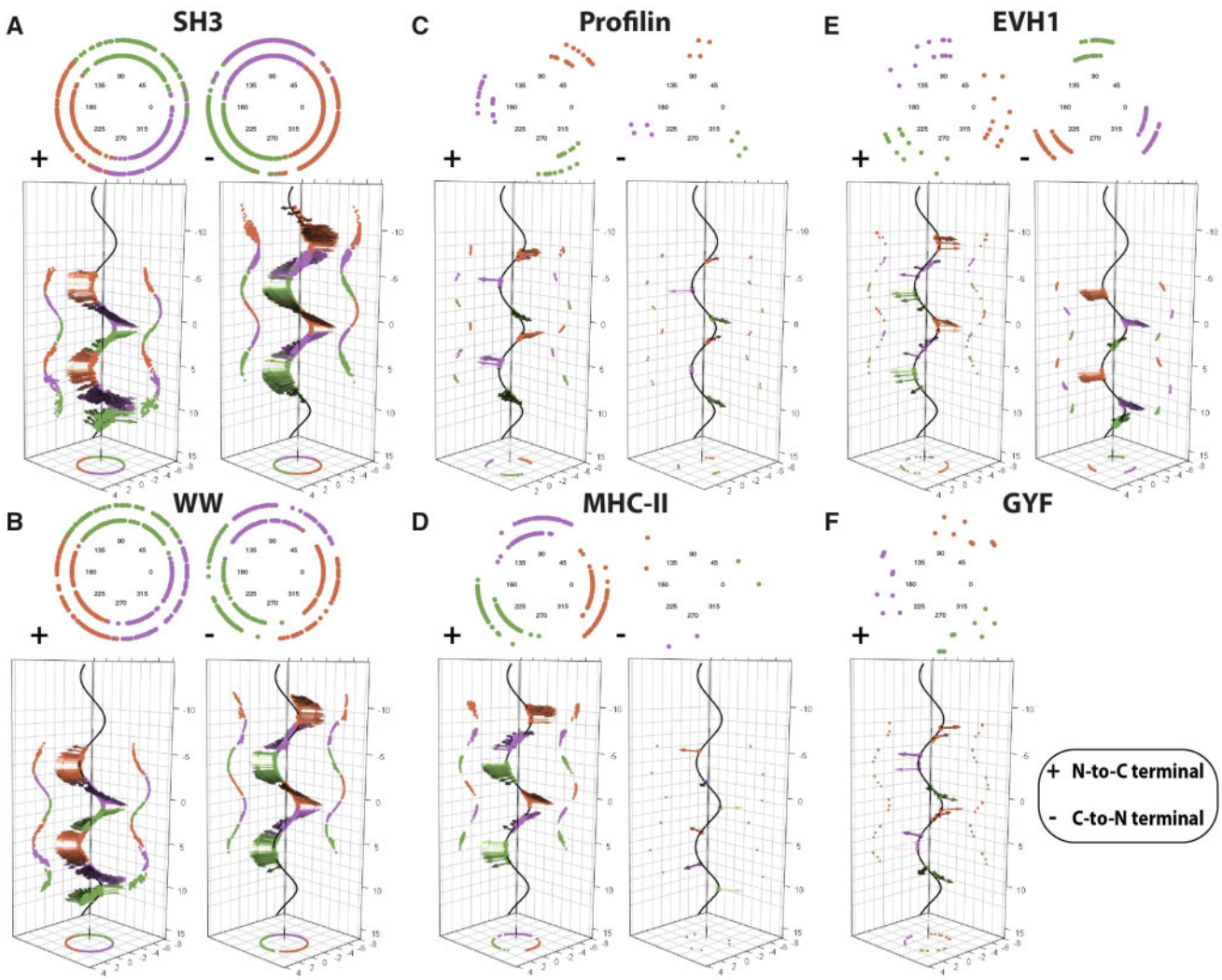

Fig. 3. K-helix sliding helical pattern. The residue angles of κ-helices are shown as a 2D projection (top) and in 3D (bottom) stratified by orientation. (**A**) SH3 domain ($n = 405$ conformations). (**B**) WW domain ($n = 151$ conformations). (**C**) Profilin ($n = 13$ confirmations). (**D**) MHC-II ($n = 65$ conformations). (**E**) EVH1 domain ($n = 27$ conformations). (**F**) GYF domain ($n = 7$ conformations). Red, purple and green dots represent the 1st, 2nd and 3rd residues of each κ-helix turn, respectively. In the 2D projection (top), the 1st and 2nd turns are designated by the inner and outer circles, respectively. Perpendicular arrows and *X*, *Y* and *Z* projections are shown to aid in the spatial interpretation. X- and Y-axes are in Angstroms. *Z*-axes are in radians. Each panel is stratified into positive (N-to-C) and negative (C-to-N) orientations, designated by + and − signs, respectively. NMR ensembles are included in the dataset

type of H-bonds (donor/acceptor) and contributing atoms (side-chain/main-chain) within the domain families, especially in SH3 and MHC-II complexes. The most common H-bond patterns are summarized in Supplementary Figure S3. Also, we investigated whether there is a relationship between the rotation angle and the strength or angle of the donor-acceptor H-bonds. There was a significant correlation ($P = 0.03$) between the rotations and the donor-acceptor distances at positions 0 and 3 in the SH3 positive and negative orientations, respectively (Supplementary Fig. S4). Furthermore, significant correlations between the rotations and H-bond angles were observed in three positions (−2, 1 and 6) in MHC-II (Supplementary Fig. S5). Overall, the structural analysis demonstrates the shared attributes and complexities of κ-helix interactions on different levels and their relationship to the rotational angle.

3.3 Analysis of κ-helix NMR ensembles

NMR ensemble structures with multiple conformers may provide valuable information regarding the dynamic motion of protein complexes. To assess whether the helical displacement occurs not just between different complexes in a domain family but also within the same structure, we analyzed the conformations of each NMR structure. The complexes of SH3, WW, EVH1 and GYF domains demonstrate a helical displacement within the bundles, confirming that this notable feature is a general property of κ-helix interaction (Supplementary Fig. S6). Since there are no NMR structures of profilin and MHC-II domains in the dataset, they were not evaluated for this purpose. However, the helical displacement is consistent in any of the NMR structures in the dataset (Supplementary Table S4). One of the major limitations of the analysis involves the result of an NMR structure determination procedure, which reflects the uncertainty of the structure-solving method (Billeter, 2015). This imprecision represents the sparse nature of the data that is to some extent, due to intrinsic structural dynamics (Billeter, 2015). Validation of local dynamics using better techniques such as NMR relaxation studies (Lindorff-Larsen *et al.*, 2005) should be performed.

3.4 The pivotal role of κ-helix in protein interaction and signaling

The significance of a domain family can be discerned using various attributes, including the size of the family and the number of interactions with counterpart proteins. Therefore, we used the SMART database to summarize the number of proteins that are involved in

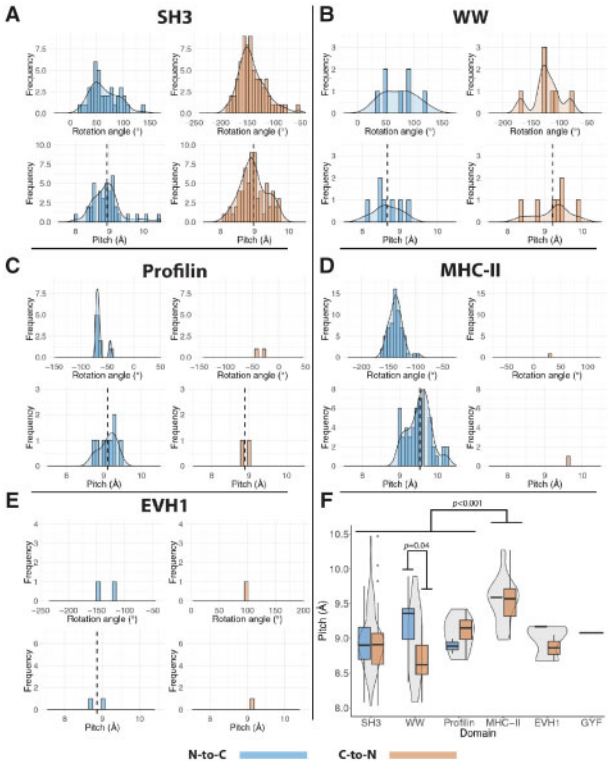

Fig. 4. Histograms of κ-helix rotational angles and pitch. The distribution of rotational angles of the 3rd residue of κ-helix relative to the target is shown at the top panel. The distribution of κ-helices pitch is shown at the bottom panel. (**A**) SH3 domain. (**B**) WW domain. (**C**) Profilin. (**D**) MHC-II. (**E**) EVH1 domain. (**F**) Boxplots comparing the pitch distributions between different domain families and their orientations. *P*-values were calculated using two-tailed Student's *t*-test. Density plots (A–E) and violin plots (F) are shown where applicable. Median pitch is depicted with a dashed line. For NMR structures with multiple conformations, the average rotation angle and pitch was calculated

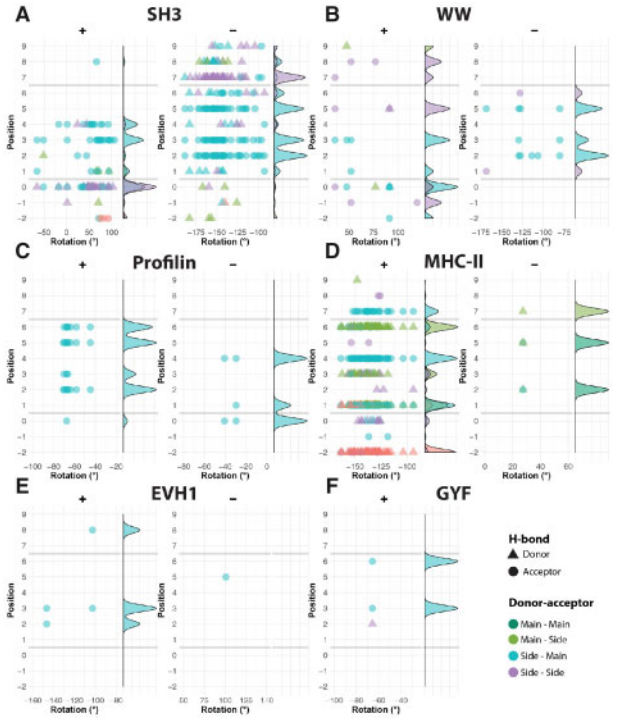

Fig. 5. Distribution of κ-helix intermolecular hydrogen bonds. The relationships between the positions of hydrogen bonds and the rotational angles (of the 3rd residue) of κ-helices are shown. (**A**) SH3 domain. (**B**) WW domain. (**C**) Profilin. (**D**) MHC-II. (**E**) EVH1 domain. (**F**) GYF domain. The densities of the contributing side-chain/main-chain atoms are shown on the right side. Each panel is stratified into positive (N-to-C) and negative (C-to-N) orientations, designated by + and − signs, respectively. Flanking N-terminal (positions −2 to −0) and C-terminal (positions 7–9) residues are indicated by two grey lines

protein-protein interactions (Supplementary Material). Supplementary Table S6 lists the top 20 domain families ranked according to the abundance of proteins in the domain. The table shows that SH3 and WW domains are among the most abundant families. SH3 domain family is ranked 3rd and contains 211 proteins which correspond to 1.03% of the proteome, whereas WW domain is ranked 14th with 52 proteins, which corresponds to 0.25%. Furthermore, we calculated the extensiveness of the protein–protein network using the 1st- and 2nd-degree neighbors associated with the domain members. The results in Supplementary Table S6 show that SH3 and WW domains interact in total with 87.8% and 83.1% of the proteome, respectively. Also, Supplementary Figure S7 illustrates the overall functional interaction map of the six κ-helices binding domains, which comprise 1.8% of the proteome. The 1st and 2nd neighbors together comprise 89.2% of the proteome, indicating how substantial is the scope of the interaction network.

4 Discussion

The discovery of the helical lock and key model that portrays the binding interaction between SH3 and PPII as helical rotations, unveiled new insights into this exceptional binding mechanism and raised questions regarding the generality of the binding model (Meirson *et al.*, 2020). Here, we show that the helical nature of the interaction between SH3 domain and its ligand equally occurs in WW, profilin, MHC-II, EVH1 and GYF domains. The consistency throughout all the investigated protein domain families signifies the universal nature of this apparatus.

The designation 'PPII' was initially coined in 1958 to describe the configuration of polymers enriched with prolines (Harrington and Sela, 1958). However, other secondary structures are not termed based on the propensities of their residues. For example, α-helix is not named 'polyalanine helix' despite the high propensity of alanine (Blaber *et al.*, 1993), yet six decades after the introduction of PPII helix, the structure is still named after the high propensity of proline. Moreover, almost half of the PPII helices contain not a single proline (Cubellis *et al.*, 2005). Therefore, despite the familiar acronym, we refer throughout the paper to PPII as κ-helix as outlined previously (Meirson *et al.*, 2020), to conform with the growing criticism of the historical term.

For many years, α-helix was considered to be the most abundant secondary structure (Buxbaum, 2015; Fisher *et al.*, 2018; Mohammed *et al.*, 2009; Rosu, 1996; Zhang *et al.*, 2000), however, this conclusion might be due to the limited assessment which is restricted to globular regions in proteins (Kohn and Hodges, 1998; Mondal and Gazit, 2016; Paoli *et al.*, 2010). If the coil regions in folded proteins are considered, the κ-helical content is comparable with the α-helical conformation (Adzhubei *et al.*, 2013). Also, in fibrillar proteins, such as collagen and elastin, κ-helix is the most common structure, with almost 100% of the proteins (Esipova and Tumanyan, 2017). Collagen alone, which is the major component of the extracellular matrix accounts for 30% of the total protein mass (Vanakker *et al.*, 2015). Therefore, we make the case that κ-helix is the most abundant secondary structure in the human body. Yet, this conformation was overlooked for many years, partly because it is not defined by hydrogen bonds and is not assigned by the most widely used secondary-structure assignment program, DSSP, which is employed in the PDB (Mansiaux *et al.*, 2011). Unlike α-helices and β-sheets, which represent structural building blocks in proteins, κ-helices represent 'functional blocks' as they are often linked to a specific function (Adzhubei *et al.*, 2013). Also, as a distinct secondary structure class, these helices play an important structural role in proteins (Esipova and Tumanyan, 2017). Their extended

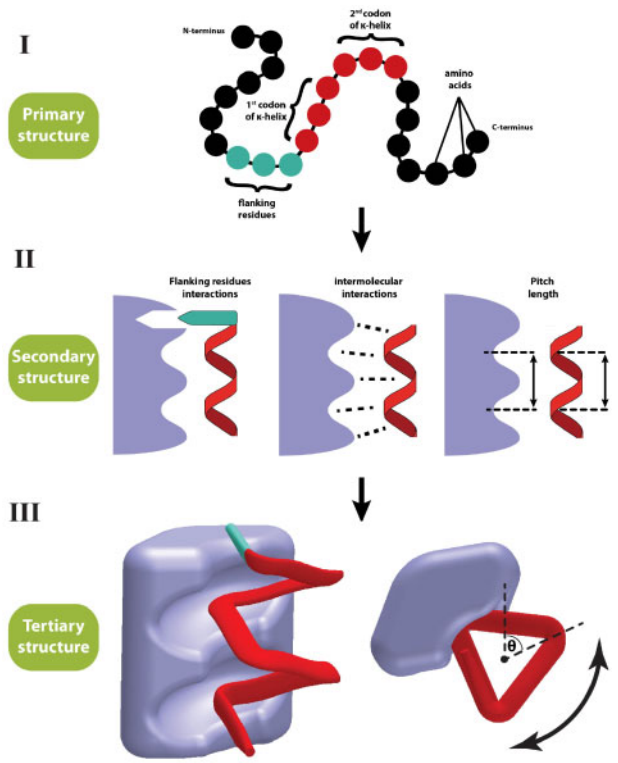

Fig. 6. Contributing factors for κ-helix interaction. (**I**) The primary structure comprising the reading frame—six residues of κ-helix, where every three residues form a structural 'codon' that completes one turn in the helix. Optionally, flanking residues are positioned upstream, downstream or in both directions. Note that the length of the peptide might be different from six residues. (**II**) The overall contribution to the binding interaction is affected by various factors including flanking residues (left), intermolecular interactions between the target and κ-helix (middle), and matching pitch between the binding interface and the ligand (right). (**III**) The resulted interaction is shown in 3D representation (left), which is characterized by a distinct rotational angle of the peptide ('screw') relative to the target (right–bottom view). The target side-chain ridges constitute the 'nut thread', and the scaffold constitutes the 'nut body'

conformation, non-regular hydrogen bonds and preferred location on the surface of proteins, makes them ideal structure for a wide range of molecular interactions (Adzhubei and Sternberg, 1993, 1994; Cubellis *et al.*, 2005). Furthermore, we show that a subset of κ-helix binding domains constitutes substantial interconnectivity covering most of the proteome (Supplementary Table S6 and Fig. S7), corroborating the idea that κ-helix represents the most widely spread binding motif in proteins (Adzhubei *et al.*, 2013). Together, κ-helices are directly involved or mediate most or possibly all major physiologically and pathological processes.

Therefore, the helical lock and key model could potentially represent the most widespread type of binding mechanism involved in functional protein–protein interactions.

The extended linear motifs, or more specifically helical motifs, formed by κ-helices, represent a structural code that encrypts the specific rotational angle to the target. The produced angle is affected by the overall contribution of various factors (Fig. 6) including: helical pitch, κ-helix intermolecular interactions (Ball *et al.*, 2005), flanking residues (Meirson *et al.*, 2020) and the structural context of the peptide (Zhou *et al.*, 2019). Discerning these relationships may provide a substantial advancement in structural biology; however, the lack of graphical representation of κ-helix dooms the observer to be blind to the visual interpretation of these structural elements (Meirson *et al.*, 2020).

The global properties of helical binding and helical displacement of κ-helices provide further support to the potential involvement of this conformation in dynamic propagation (Meirson *et al.*, 2020). Since κ-helix is characterized by irregular hydrogen bonds, low-affinity binding, and can complete a full rotational displacement, it serves as a suitable candidate for facilitating motion. This could be true not just for protein–protein interactions such as sarcomeres and motor proteins but also in protein–nucleic acid interactions where κ-helix forms non-specific interaction with the minor and major grooves of DNA and interaction with RNA molecules (Hicks and Hsu, 2004). Future studies should test the hypothesis that κ-helix can facilitate motion.

Predicting which peptides will be presented by MHC-II molecules are essential for understanding the activation of T-helper cells which orchestrate the outcome of many host immune responses (Castellino *et al.*, 1997; Rudolph *et al.*, 2006). However, despite many efforts, most methods still have low performance compared to MHC class I (MHC-I) (Jensen *et al.*, 2018). Different explanations were proposed to explain the difficulty in predicting MHC-II binding affinities, but we provide a fundamentally different explanation of why these two similar complexes produce dissimilar behaviors. Since MHC-II binds κ-helices, whereas MHC-I mostly not (Adzhubei *et al.*, 2013), the peptide–MHC-II interaction is also affected by the rotational space, facilitated by the exposed backbone of κ-helices. Therefore, similar sequences may bind in a different angle or binding register, whereas unalike sequences may bind similarly, contributing to the complexity of generating predictive models in MHC-II.

Since the introduction of the κ-helix structure over six decades ago, scientists have pondered about the unusual nature of this conformation. Here, we find that the rotating helical motif is omnipresent in κ-helix interactions. The ability of the motif to function in a helical pattern acquires a new physical layer of complexity, enabling a simple structure, restricted by a limited set of interactions to form virtually unlimited arrangements in the rotational space. Thus, the helical lock and key model may provide a solution to the enigma of how small, promiscuous, and conformationally limited structures perform surprisingly rich and directed functions. While once considered infrequent and insignificant with disordered conformational composition, the κ-helix structure may reveal to be the master conductor of biological interactions.

Author contributions

T.M. conceived the original idea, designed the study and the computational framework and analyzed the data. T.M. wrote the manuscript. A.O.S. supervised and contributed to the final version of the manuscript. D.B. and G.M. aided in interpreting the results. A.O.S., G.M. and D.B. discussed the results and commented on the manuscript.

Funding

This work was supported by the Leir foundation and the Ginzburg foundation. T.M. was supported by the Foulkes Foundation fellowship for MD/PhD students.

Conflict of Interest: none declared.

References

Adzhubei,A.A. and Sternberg,M.J. (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J. Mol. Biol.*, **229**, 472–493.

Adzhubei,A.A. and Sternberg,M.J. (1994) Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. *Protein Sci.*, **3**, 2395–2410.

Adzhubei,A.A. *et al.* (2013) Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.*, **425**, 2100–2132.

Agrawal,V. and Kishan,K. (2002) Promiscuous binding nature of SH3 domains to their target proteins. *Protein Pept. Lett.*, **9**, 185–193.

Ball,L.J. *et al.* (2005) Recognition of proline-rich motifs by protein–protein-interaction domains. *Angew. Chem. Int. Ed. Engl.*, **44**, 2852–2869.

Berisio,R. *et al.* (2006) Polyproline helices in protein structures: a statistical survey. *Protein Pept. Lett.*, **13**, 847–854.

Billeter,M. (2015) A consensus on protein structure accuracy in NMR? *Structure*, **23**, 255–256.

Blaber,M. *et al*. (1993) Structural basis of amino acid alpha helix propensity. *Science*, **260**, 1637–1640.
Buxbaum,E. (2015) Protein structure. In: *Fundamentals of Protein Structure and Function*. Springer International Publishing, Switzerland, pp. 15–64.
Castellino,F. *et al*. (1997) Antigen presentation by MHC class II molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. *Hum. Immunol*., **54**, 159–169.
Creamer,T.P. and Campbell,M.N. (2002) Determinants of the polyproline II helix from modeling studies. *Adv. Protein Chem*., **62**, 263–282.
Cubellis,M. *et al*. (2005) Properties of polyproline II, a secondary structure element implicated in protein–protein interactions. *Proteins*, **58**, 880–892.
Eisenberg,D. (2003) The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. USA*, **100**, 11207–11210.
Esipova,N.G. and Tumanyan,V.G. (2017) Omnipresence of the polyproline II helix in fibrous and globular proteins. *Curr. Opin. Struct. Biol*., **42**, 41–49.
Fisher,B.F. *et al*. (2018) Thermodynamic scale of β-amino acid residue propensities for an α-helix-like conformation. *J. Am. Chem. Soc*., **140**, 9396–9399.
Grant,B.J. *et al*. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–2696.
Harrington,W.F. and Sela,M. (1958) Studies on the structure of poly-L-proline in solution. *Biochim. Biophys. Acta*, **27**, 24–41.
Hicks,J.M. and Hsu,V.L. (2004) The extended left-handed helix: a simple nucleic acid-binding motif. *Proteins*, **55**, 330–338.
Hollingsworth,S.A. *et al*. (2009) On the occurrence of linear groups in proteins. *Protein Sci*., **18**, 1321–1325.
Holt,M.R. and Koffer,A. (2001) Cell motility: proline-rich proteins promote protrusions. *Trends Cell Biol*., **11**, 38–46.
Jensen,K.K. *et al*. (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, **154**, 394–406.
Jha,A.K. *et al*. (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry*, **44**, 9691–9702.
Kaneko,T. *et al*. (2008) The SH3 domain—a family of versatile peptide-and protein-recognition module. *Front. Biosci*., **13**, 4938–4952.
Kelly,M.A. *et al*. (2001) Host– guest study of left-handed polyproline II helix formation. *Biochemistry*, **40**, 14376–14383.
Kohn,W.D. and Hodges,R.S. (1998) *De novo* design of α-helical coiled coils and bundles: models for the development of protein-design principles. *Trends Biotechnol*., **16**, 379–389.
Lindorff-Larsen,K. *et al*. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, **433**, 128–132.
Mansiaux,Y. *et al*. (2011) Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PLoS One*, **6**, e18401.
Martin,C. *et al*. (2014) Silaproline helical mimetics selectively form an all-trans PPII helix. *Chem. A Eur. J*., **20**, 14240–14244.
Meirson,T. *et al*. (2020) A helical lock and key model of polyproline II conformation with SH3. *Bioinformatics* **36**, 154–159.
Mensch,C. *et al*. (2016) Ramachandran mapping of peptide conformation using a large database of computed Raman and Raman optical activity spectra. *Phys. Chem. Chem. Phys*., **18**, 31757–31768.
Milner-White,E.J. and Russell,M.J. (2008) Predicting the conformations of peptides and proteins in early evolution. *Biol. Direct*, **3**, 3.
Mohammed,O.F. *et al*. (2009) Primary peptide folding dynamics observed with ultrafast temperature jump. *Angew. Chem. Int. Ed. Engl*., **48**, 5628–5632.
Mondal,S. and Gazit,E. (2016) The self-assembly of helical peptide building blocks. *ChemNanoMat*, **2**, 323–332.
Paoli,B. *et al*. (2010) Slow folding of cross-linked α-helical peptides due to steric hindrance. *J. Phys. Chem. B*, **114**, 2023–2027.
Rhee,S. and Grinnell,F. (2007) Fibroblast mechanics in 3D collagen matrices. *Adv. Drug Deliv. Rev*., **59**, 1299–1305.
Rosu,H. (1996) On the kinks and dynamical phase transitions of α-helix protein chains. *Il Nuovo Cimento D*, **18**, 477–481.
Rudolph,M.G. *et al*. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol*., **24**, 419–466.
Siermala,M. *et al*. (2001) On preprocessing of protein sequences for neural network prediction of polyproline type II secondary structures. *Comput. Biol. Med*., **31**, 385–398.
Siligardi,G. and Drake,A.F. (1995) The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. *Biopolymers*, **37**, 281–292.
Stapley,B.J. and Creamer,T.P. (1999) A survey of left-handed polyproline II helices. *Protein Sci*., **8**, 587–595. doi: 10.1110/ps.8.3.587.
Vanakker,O. *et al*. (2015) The genetics of soft connective tissue disorders. *Annu. Rev. Genomics Hum. Genet*., **16**, 229–255.
Wang,M.L. *et al*. (2005) Prediction by support vector machines and analysis by Z-score of poly-L-proline type II conformation based on local sequence. *Comput. Biol. Chem*., **29**, 95–100.
Zagrovic,B. *et al*. (2005) Unusual compactness of a polyproline type II structure. *Proc. Natl. Acad. Sci. USA*, **102**, 11698–11703.
Zhang,M. *et al*. (2000) Conformational characterization of a helix-nucleated bicyclic GCN4 decapeptide by proton NMR. *J. Pept. Res*., **55**, 398–408.
Zhou,P. *et al*. (2019) Is protein context responsible for peptide-mediated interactions? *Mol. Omics*, **15**, 280–295.