*Article*

# Structural Achievability of an NH–π Interaction between Gln and Phe in a Crystal Structure of a Collagen-like Peptide

**Ruixue Zhang [1,†], You Xu [2,3,†], Jun Lan [4], Shilong Fan [4,\*], Jing Huang [2,3,\*] and Fei Xu [1,\*]**

[1] Ministry of Education Key Laboratory of Industrial Biotechnology, School of Biotechnology, Jiangnan University, Wuxi, 214122, Jiangsu, China

[2] Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, 310024, Zhejiang, China

[3] Westlake AI Therapeutics Lab, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, 310024, Zhejiang, China

[4] Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, 100084, China

\* Correspondence: fanshilong@mail.tsinghua.edu.cn (S.F.); huangjing@westlake.edu.cn (J.H.); feixu@jiangnan.edu.cn (F.X.)

† These authors contributed equally to this work.

**Abstract:** NH–π interactions between polar and aromatic residues are well distributed in proteins whose stabilizing effects have been investigated in globular and fibrous proteins. In order to gain structural insights into side chain NH–π interactions, we solved a crystal structure of a collagen-like peptide containing Gln-Phe pairs. The Gln-Phe NH–π interactions were further characterized by quantum calculations, molecular simulations, and structural bioinformatics. The analyses indicated that the NH–π interactions are robust under various solvent conditions, can be distributed either on the protein surface or in its hydrophobic core and can form at a wide range of distances between residues. This study suggested that NH–π interactions can play a versatile role in protein design, including engineering hydrophobic cores, solvent accessible surfaces, and protein–protein interfaces.

## 1. Introduction

NH groups are prevalent in proteins given their location in both the backbone and some side chains (e.g., Lys, Arg, Asn, and Gln). A previous study showed that approximately 1 in 10.8 aromatic side chains, not including His, make close contact with an NH group[1]. The nature of NH–π interactions is largely dependent on the polarity of the donor group. When N is from ammonium/guanidinium of Lys/Arg, the interaction energy is dominated by an electrostatic term, so-called cation–π interactions. The stabilizing effects of side chain cation–π interactions have been characterized and subsequently used to enhance protein stability [2–5]. However, when the N from either the backbone or uncharged Asn/Gln side chains is best classified between the conventional hydrogen bonding and a weaker interaction dominated by dispersion. Burley et al. initially noted that aromatic residues could make close contact with both Arg/Lys or Asn/Gln with similar preference [6], suggesting that side chain NH–π interactions could also play an important role in protein stability and/or function. The quantification of NH–π interactions in small molecule systems had been determined experimentally, which was also supported by quantum mechanical theory. In NH–π interactions in the N-methylformamide-benzene complex, the minimum energy distance was 3.2 Å and the binding energy was −4.37 kcal/mol [7]. In contrast, the stabilizing effects of NH–π interactions in Acetyl-Phe-Gly-

Gly-N-Methyl amide peptide model ranged from –1.98 to –9.24 kcal/mol [8]. More detailed studies are needed to understand the NH–π interactions in proteins.

Collagen-like peptides have been utilized as a molecular platform to investigate side chain interactions. The characteristic structure of a collagen-like peptide is a triple helix composed of three chains. Each chain adopts a well-extended PPII conformation with successive Gly-X-Y triplets [9–11]. Although the X and Y positions of this motif can accommodate any amino acid, the highest stability triplet is Gly-Pro-Hyp, where Hyp (O) represents (4R)-hydroxyproline [12]. Chain trimerization is induced by inter-chain backbone hydrogen bonding between Gly and non-Gly residues [9]. In this context, Gly is buried inside the triple helices, while side chains of residues in the X/Y positions are exposed on the collagen surface. With host–guest methods, various amino acids are introduced to replace Pro at X or Hyp at Y-position at the central triplets of the host sequences containing high-imino triplets only, such as $(GPO)_n$, where n = 8, 10 [13–17]. As GPO is the most stabilizing triplet, any substitution could destabilize the peptides with decreased melting temperature ($T_m$). When a peptide containing two substitutions exhibits a higher $T_m$ than expected one by adding the destabilizing effects of two individual substitutions, which could be taken as the energetic contribution of side chain interactions [13]. Stabilizing effects of salt bridges and cation–π interactions, such as Arg-Phe and Lys-Phe pairs, have also been characterized by the host–guest methods [5,14,18–21].

In collagen triple helices, there are two types of inter-chain side chain interaction between two neighboring chains containing the GXYGX'Y' sequence. One type is lateral (Y-X) interaction and the other type is axial (Y-X'). Walker et al. designed guest peptides to calculate the energetic contributions of the axial and lateral interactions [22]. The guest sequence GPOGFQ embedded in $(GPO)_8$ contained two lateral Gln-Phe pairs with $T_m$ 7.5°C lower than the host, while GPQGFO contained one lateral and two axial Gln-Phe pairs with $T_m$ similar to the host [23]. The energetic contribution of the axial and lateral Gln-Phe pair was derived to be 2.13 °C and –0.75 °C, respectively. Axial Gln-Phe pairs were also utilized to induce the formation of a collagen heterotrimer. Molecular dynamics showed that axial Gln-Phe pairs could make close contacts (3.8 Å). However, more detailed structural information is necessary to understand the properties of the NH–π interactions in collagen.

In this study, in order to gain structural insights into an NH–π interaction between Gln and Phe and better understand its chemical properties, we solved a crystal structure of a collagen-like peptides $(GPO)_3GPQGFO(GPO)_3$, abbreviated as GPQGFO, and characterized the side chain interactions using several lines of evidence including circular dichroism, quantum chemical analysis, molecular dynamics simulation, and structural bioinformatics. Two major features of the NH–π interaction were revealed, including its environmental stability and conformational diversity, which provide a framework for exploring the mechanisms of non-covalent XH–π (X = N, S, O, C) interactions and further applying them to protein design.

## 2. Materials and Methods

### 2.1. Peptide Synthesis and Purification

The peptides were synthesized with FMOC solid-phase chemistry method by GL Biochem Ltd (Shanghai, China), and purified to more than 95% purity. The purity was tested with reversed-phase high-performance liquid chromatography (RP-HPLC) on an Agilent 1260 system using an AdvanceBio peptide map column (C18), 2.1 × 150 mm, pore size 2.7-micron, with a linear gradient of water and acetonitrile gradient each containing 0.5% TFA. Molecular weights were validated with matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) on a Bruker autoflex II using a pre-spotted anchor chip with 10mg/ml 2,5-dihydroxybenzoic acid in water, 50% acetonitrile as the matrix, and a dual-stage reflection electron microscope obtain the spectrum (Figure S5). The sequence of the peptide was as below:

GPQGFO: GPOGPOGPOGP**Q**G**F**OGPOGPOGPO

where O is (4R)-hydroxyproline. The N- and C- termini are acetylated and amidated, respectively.

### 2.2. *Crystallization and Data Collection*

The pure and lyophilized GPQGFO peptide powder was dissolved in a 20 μL buffer at pH 7.0 with 10 mM Tris·HCl making the solutions of 5 mM concentration. Incubate at 4 °C overnight and then set for crystallization by using the hanging drop vapor diffusion method at 4 °C. After ~one month, the best crystals were obtained under 0.2 M potassium nitrate and 20% PEG 3350. X-ray data were collected at 100 K, and the best one was diffracted to 1.385 Å. The space group was $P2_122_1$, with dimensions: a = 24.09 Å, b = 28.01 Å, c = 73.37 Å, $\alpha = \beta = \gamma = 90°$ indexed by the HKL2000 software (HKL Research, Charlottesville, VA, USA). Data collection and processing statistics are summarized in Table S2.

### 2.3. *Structure Determination and Refinement*

The structures of GPQGFO were solved by molecular replacement using a Phaser from the Phenix software suite [24] with a collagen structure (PDB ID: 5YAN [25]) as the search models. Initial phases were improved by rigid body refinement followed by rounds of simulated annealing and anisotropic B-factor refinement using the Phenix suite. Manual model rebuilding was performed using COOT [26]. Further rounds of refinement were performed using the phenix.refine program implemented in the PHENIX package with coordinate refinement, isotropic ADP refinement, and bulk solvent modeling. The final GPQGFO model contains 598 peptide atoms and 115 water molecules. The final GPQGFO structure has an $R_{work}/R_{free}$ value of 16.7/18.0%.

### 2.4. *Circular Dichroism (CD) Spectra*

CD was recorded in 1-mm quartz cuvettes on a Chirascan instrument with a six-cell thermostated cell holder and a temperature controller (Applied Photophysics Ltd, London UK). Peptide solutions of 0.2 mM were characterized in phosphate-buffer (at pH 7.0) or in pH 3, 9 or in 100, 150, 300, 500 mM NaCl. Wavelength scans were conducted from 190 to 260 nm at 4 °C. For temperature-induced denaturation, ellipticity was measured from 4 to 80 °C with a heating rate of 1 °C for 6 minutes. Observed ellipticity was converted to molar ellipticity by dividing raw values by the peptide concentration, number of residues, and cell path length. The melting temperature, $T_m$, was estimated as below:

$$F(T) = \frac{\theta(T) - \theta_U(T)}{\theta_F(T) - \theta_U(T)}$$

where $\theta(T)$ is the observed ellipticity, and $\theta_F(T)$ and $\theta_U(T)$ are predicted ellipticities derived from linear fits to the folded and unfolded baselines. The melting temperature is estimated as the T where F(T) = 0.5.

### 2.5. *Quantum Mechanical (QM) Analysis*

To probe the global energy landscape of Q-F interaction, a three-dimensional potential energy scan was performed between acetamide and toluene. A set of 3-D Cartesian axes was created where the x-y plane was mapped on the toluene ring with atom C1 defined as the origin and vector CM-C1 as the *x*-axis. Because toluene is symmetric about the *x*-axis and also the x-y plane, the whole interaction space between two molecules can be represented by such a quadrant that on only one side of the *x*-axis and x-y plane. The space was gridded at every 0.4 Å from 0 to 4.8 Å along x- and y-axes and every 0.2 Å from 2.0 Å to 5.0 Å along the *z*-axis. In each grid, the coordinate of the N atom in acetamide was constrained while the other atoms were free to move. In total, 2704 pairs of molecule models were optimized and calculated and the energy surface of the amide position related to the toluene ring was scanned. The M06-2X functional with cc-pVTZ basis set was em-

ployed in geometric optimization using Gaussian. Then the single point energy was calculated with the aug-cc-pVTZ basis set and counterpoise correction was employed to calculate the interaction energy. ORCA [27] was used to perform all single-point energy calculations.

Additional QM calculations were carried out to understand the NH–π interaction between acetamide and toluene. Geometry optimization was carried out at the MP2/aug-cc-pVTZ level using psi4 [28] and interaction energy was obtained at the DLNPO-CCSD(T)/aug-cc-pVQZ level with counterpoise correction using ORCA. SAPT calculations were carried out with the aug-cc-pVQZ basis set using psi4.

### 2.6. *Calculation of Geometric Parameters*

To determine the geometric relationship between the pairwise, three bond conformations ($d_{NM}$, θ, ω,) and three side-chain-positions ($d_{C\gamma-\pi}$, τ, χ) were defined. $d_{NM}$ was denoted as the distance between the N of the amide NH groups of Gln or acetamide and the center (M) of the phenyl ring. At this time, the angle between NH and M was defined as ω ($0 < \omega < 180°$), and the acute angle between the MN vector and the normal vector of the phenyl ring was described as θ ($0 < \theta < 90°$). The $d_{C\gamma-\pi}$ and τ were defined as the shortest distance and angle between Cγ of the Gln or acetamide and the plane formed by the phenyl ring. If Cγ is located in the same direction as the normal vector, τ is positive ($0 < \tau < 90°$), otherwise, it is negative ($-90° < \tau < 0°$). Whereas the χ was described as the angle between the vector formed by the projection of the Cγ-N vector onto the plane and the Cβ-Cγ vector formed by the phenyl ring. See SI method for calculation details. All atomic cartesian coordinates were retrieved from the pdb files. Since hydrogen atoms are missing from most pdb files, theoretical positions of hydrogen atoms (X-H = 1.0 Å) were then added with HGEN [29] software, but only those at positions allowing this calculation unambiguously were used further. The following cutoffs are used to screen the Q-F pairs with the NH–π interactions, $d_{N-M} < 4.3$ Å, $\theta < 25°$, $\omega > 120°$ [1].

### 2.7. *Molecular Dynamics (MD) Simulation*

The molecular triplex model with 24 residues in each strand was built from the crystal structure using CHARMM [30]. After energy minimization in a vacuum, the system was solvated with TIP3P water [31] in a cubic box that was 89 Å in length. No counterions or salt concentration was included. Periodic boundary conditions were used and the particle mesh Ewald was applied on both electrostatic [32] and Lennard–Jones (LJ-PME) [33,34] interactions, with 12 Å as the real space cutoff. Applying CHARMM36m [35,36] force field, the simulation was performed in an isothermal-isobaric ensemble (NPT) at 298 K and 1 atm by employing the Andersen thermostat and Monte Carlo barostat, respectively. With the constraints on hydrogen-involved bonds, a time step of 2 fs was used in the velocity Verlet integrator. After 0.64 ns equilibrium simulation, the system was further run for 200 ns using OpenMM [37] with snapshots saved every 100 ps.

### 2.8. *Database Analysis of Q-F Pair Interaction*

A non-redundant set was curated containing 9658 protein structures all determined by X-ray diffraction to a resolution of 2.0 Å or higher, refined to R ≤ 0.25 and less than 30% sequence homology by PIESCES suits [38,39].

The following criteria were employed to assemble the set: (1) no theoretical model structures and no incomplete entries, i.e., Cα coordinates only, were accepted; (2) only crystal structures with a resolution of 2.0 Å or better and a crystallographic R-factor of 0.25 or lower were accepted; (3) the minimum chain length was 40 amino acid residues; and (4) the maximum acceptable amino acid identity between any two protein chains of the set was 30%.

## 3. Results

### 3.1. Triple Helices Remained Stable under Various Ionic Strengths and pH

In a previous study of GPQGFO peptides, it was indicated that a favorable Q–F interaction could be generated to stabilize the triple helix in a neutral environment [23]. To experimentally assess the tolerance of the Q-F pairs in different environments, the thermal stability of GPQGFO under various NaCl (100–500 mM) concentrations and pH (3, 7, 9) was characterized by circular dichroism spectra. The molar residual ellipticity (MRE) at 220 nm and $T_m$ (~40°C) were apparently unaffected (Figure 1b, c). These results indicated that the stability contribution of the Q-F pairs was maintained under various environmental conditions, which distinguished the GPQGFO motif from that of less tolerant motifs such as GPKGEO where the stabilizing electrostatic interactions are more affected [18]. Notably, the $T_m$ values of peptides recorded here were higher than those reported by Persikov et al. [12] and Walker et al. [23], potentially due to differences in melting schedule, host sequence, or blocking of peptide termini.
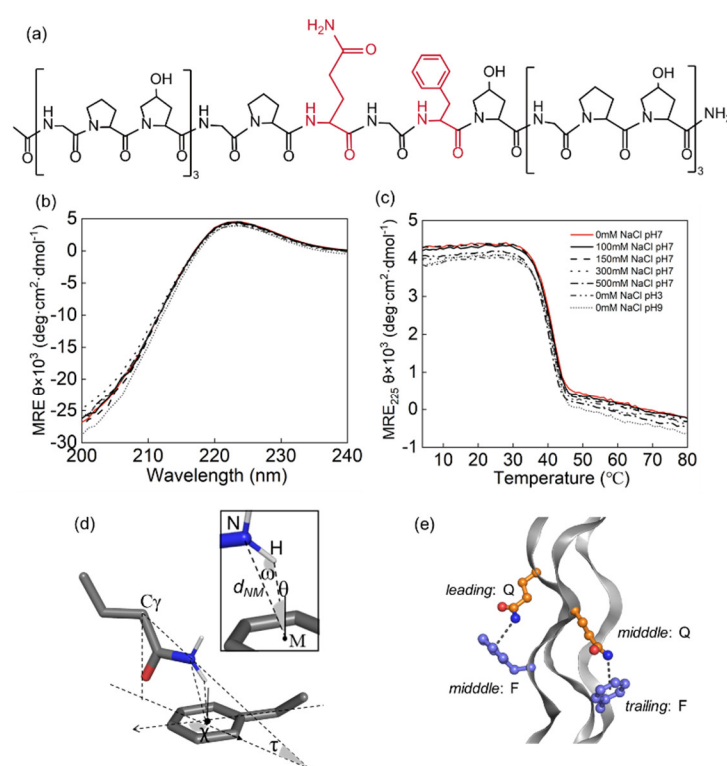


**Figure 1.** Experimental characterization and geometric description of Gln–Phe interactions: (**a**) primary sequence of the collagen-like peptide, [(GPO)₃GPQGFO(GPO)₃]₃ is shown, where O represents (4R)-hydroxyproline. Gln and Phe are colored in red; (**b–c**) triple helix formation and stability of peptide GPQGFO were characterized by CD in 10mM phosphate buffer under various NaCl concentrations and pH values as labeled in the plot. Wavelength spectra at 4 °C are shown in (**d**), and temperature melting experiments monitored at 225nm from 4 to 80 °C are shown in (**e**); (**d**) geometric parameters of an NH–π interaction in either a Q-F pair or an acetamide-toluene complex are defined. Three side-chain-position parameters, $d_{C\gamma-\pi}$, τ, and χ, describe relative position and orientation between Gln Cγ-Nε (acetamide C-N) and the phenyl plane. Inset on top-right corner shows three bond conformation parameters, $d_{NM}$, θ, and ω, describing how Gln Nε-H (acetamide N-H) is positioned relative to the phenyl ring center M; (**e**) the crystal structure of the collagen triple helices was solved (PDB ID:7VEG). Two closely contacted Q-F pairs are shown in ball-and-stick representation. The chain identities are as labeled in figure (**d**)–(**e**), figures were generated with Pymol.

### 3.2. Side Chain Conformations in Crystal Structure

In order to gain a better understanding of the Q-F pairs in the context of a triple helix, and especially to help understand the difference between the axial and lateral sequential geometry [23]. The GPQGFO peptide crystallizes with the triple helical structure characteristic of collagen molecules with 1.385 Å resolution. More data acquisition and refinement statistics are available in Table S2. Two axial pairs and one lateral were observed in the structure depending on the relative positions of the interacting amino acids (Figure 1e). Two axial pairs were, respectively, composed of Gln in the leading and Phe in the middle chains of the triple helices, the other was between the middle and trailing chains. Three bond-conformation parameters, $d_{NM}$, θ, and ω, describe whether the Gln Nε-H perpendicularly points to the phenyl ring center, forming a hydrogen bond-like conformation (Figure 1d). The conformations were located at the region with $d_{NM}$ ~3.3 Å, θ ~10° and ω ~150°, well within the empirical cutoffs for NH–π interactions ($d_{NM} < 4.3$ Å, $-25° < θ < 25°$, and $ω > 120°$) extracted from the previous data-mining of protein structures [1].

Except for the bond-conformation parameters, side-chain-position parameters, $d_{Cγ-π}$, τ and χ, describe relative positions between the two molecules or side chains (Figure 1d). The separation distances were located at the regions with $d_{Cγ-π}$ ~5 Å, τ ~150° and χ ~93°. There was a third pair between Gln in the trailing and Phe in the leading chains. The side chains were pointed in opposite directions ($d_{NM} = 7.4$ Å), which did not form any interactions.

### 3.3. Energetic Favorability of Amide–π Interactions Characterized with Quantum Methanics

To explore whether the Q-F pairs captured in the crystal structures are energetically favored, ab initio QM calculations were performed on an acetamide-toluene complex as a model system of Gln–Phe interactions. A global minimum of interaction energy was obtained at the M06-2X/aug-cc-pVTZ level by scanning the relative positions of the acetamide atom N, representing Gln Nε, with respect to the coordinate set on the phenyl ring (Figure S1). The global minimum of interaction energy (ΔE) was −5.81 kcal/mol similar to hydrogen bonding (approximately −5 kcal/mol) [40,41] and smaller than cation–π interactions (about −19 kcal/mol) [42].

In order to better understand the relationship between the conformation and interaction energy (ΔE), geometric parameters were plotted against ΔE (Figure 2a). The optimal (ΔE = −5.81 kcal/mol) and sub-optimal (−5.75 < ΔE < −5.48 kcal/mol) conformations were located at the region with $d_{NM}$ ~3.2 Å, θ ~10° and ω ~150°, were highly similar to those of axial Q-F pairs in the crystal structure, which suggested the latter should also be energetically favored (Table 1). At $0 < θ < 20°$, ΔE was plotted with respect to $d_{NM}$ and the curve was similar to the Leonard–Jones potential. The minimal ΔE occurs at $d_{NM}$ ~3.0–3.5 Å. Beyond $d_{NM}$ ~5 Å, there was little interaction, while within $d_{NM}$ ~3 Å, ΔE increased dramatically (Figure 2c). This result indicated that the NH–π interaction was favorable in specific geometry of distance and orientation.

**Table 1.** Geometric parameters of Q-F pairs extracted from QM calculation of the acetamide-toluene complex, crystal structure and MD trajectories of peptide GPQGFO, and a non-redundant dataset of protein crystal structures.

| Methods | Pairs | $d_{NM}$ (Å) | θ (°) | ω (°) | $d_{Cγ-π}$ (Å) | τ (°) | χ (°) |
|---|---|---|---|---|---|---|---|
| QM | Acetamide-Methylbenzene | 3.2 | 8.2 | 149.1 | 4.9 | 46.7 | 65.9 |
| Crystal | $L^a$:Q-$M^a$:F | 3.3 | 8.4 | 159.2 | 5.3 | 61.1 | 93.0 |
| | $M^a$:Q-$T^a$:F | 3.5 | 20.9 | 141.9 | 5.5 | 62.3 | 107.8 |
| MD[b] | $L^a$:Q-$M^a$:F | 3.3 | 15.9 | 145.4 | 5.6 | 72.3 | 43.8 |
| | | | | | 5.2 | −50.5 | 115.1 |
| | $M^a$:Q-$T^a$:F | 3.2 | 10.5 | 139.9 | 5.5 | 73.0 | 113.4 |
| | | | | | 5.5 | −78.4 | 122.4 |

| Data mining[c] | Q-F | 3.5 ± 0.23 | 14.1 ± 6.1 | 146.3 ± 14.1 | 3.9 ± 0.8 | 72.1 ± 19.8 | 103.9 ± 37.1 |
| | | | | | 3.9 ± 0.6 | −16.0 ± 14.9 | 122.2 ± 43.5 |

[a] Leading, middle and trailing chain of the triple helices were represented as L, M, and T, respectively. [b] The parameter values with the highest occurring frequencies were listed here. [c] The mean and standard deviation values were listed here.

The side-chain-position parameters of optimal and sub-optimal states were located at the regions with $d_{C\gamma\text{-}\pi}$ ~5 Å, $\tau$ ~50° and $\chi$ ~66°, indicating that acetamide (Gln C$\gamma$-N$\varepsilon$) was tilted against the phenyl ring (Figure 2b). Notably, the sub-optimal states were clustered near the toluene C atom (Phe C$\beta$) with $\chi$ ~65°, which might enhance interaction between acetamide and toluene methyl group. With 40° < $\tau$ < 50°, the data of $d_{C\gamma\text{-}\pi}$ vs. ΔE were projected on a plane (Figure 2d). ΔE decreases as $d_{C\gamma\text{-}\pi}$ decreases. When 4.5 Å < $d_{C\gamma\text{-}\pi}$ < 5 Å, ΔE decreases as $\tau$ increases (Figure S2). This suggests that, although the side chain proximity and orientation could affect the interaction energy, an NH–$\pi$ interaction could be formed under various side chain conformations in protein.

### 3.4. Balanced Contributions from Various Types of Chemical Forces

To further understand the chemical property of NH–$\pi$ interactions, the single point energy of the global minimum of acetamide-toluene was calculated at a higher level of DLNPO-CCSD(T)/aug-cc-pVQZ. The interaction energy was −5.95 kcal/mol. Symmetry-adapted perturbation theory (SAPT) was then used to separately examine the electrostatic ($E_{ele}$), exchange ($E_{ee}$), induction ($E_{ind}$), and dispersion ($E_{disp}$) terms that contribute to the Q–F interaction energy (Table S1). The SAPT2+3/aug-cc-pVQZ calculation revealed a total interaction energy of −7.03 kcal/mol, with 1 kcal/mol overestimated from the CCSD(T) value. Both $E_{ele}$ (−6.6 kcal/mol) and $E_{disp}$ (−8.5 kcal/mol) made comparably large contributions (Table S1). The $E_{disp}/E_{ele}$ ratio (1.29) indicated a relatively balanced interaction between dispersion and electrostatic energies that indeed affected characteristics of both the protein interior and exterior [43].
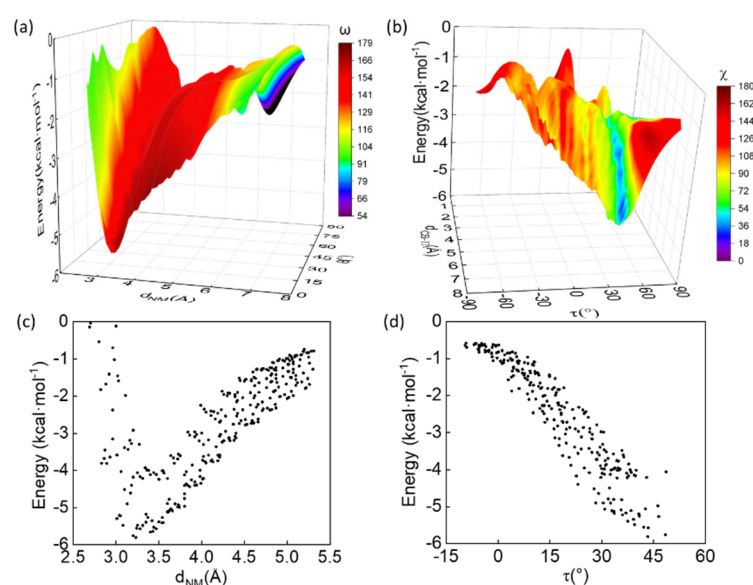


**Figure 2.** (**a**) Three-dimensional (3D) interaction energy (ΔE) surface is plotted against $d_{NM}$ and θ. The energy surface is color-coded with ω. The color scales are located on the right. (**b**) 3D ΔE surface is plotted against $d_{C\gamma\text{-}\pi}$ and $\tau$. The energy surface is color-coded with $\chi$. The color scales are located on the right. (**c**) The data points with 0 < θ < 20° in the 3D-plot in (**a**) are projected on the ΔE-$d_{NM}$ plane. (**d**) The data points with 40° < $\chi$ < 50° in the 3D-plot in (**b**) are projected on the ΔE-$\tau$ plane.

### 3.5. Environmental Dielectric Constants

To explore how the NH–$\pi$ interaction is influenced under a wide range of environments, such as protein interior and aqueous solution, the DLNPO-CCSD(T) calculations with implicit solvent model CPCM were performed. The interaction energy moderately decreased to −4.54 kcal/mol in protein interior (dielectric constant $\varepsilon$ = 4) and −4.03 kcal/mol in water ($\varepsilon$ = 78.4), respectively, compared to −5.95 kcal/mol in a vacuum. In contrast, cation–$\pi$ interactions, such as $NH_4^+$-benzene, were much more favorable in vacuum (approximately −19 kcal/mol) but significantly weakened to −3.6 kcal/mol in protein interior and −1.3 kcal/mol in water, which displays a high dependency on the polarity of the environment [42].

### 3.6. Conformational Distribution Sampled by Molecular Dynamics Simulations

In addition to the energetic favorability revealed by QM, atomistic molecular dynamics (MD) simulations were performed with the crystal structure of the collagen triple helix to investigate the dynamics of NH–$\pi$ interactions in solution. To this end, the geometric parameters were extracted from the 200 ns MD trajectory and plotted as color-coded density maps (Figure 3). For the two closely contacted Q-F pairs, ~22% of the bond conformations, represented by $d_{NM}$, $\theta$, and $\omega$, fell within the empirical cutoffs. Due to the flipping of the phenyl ring, the $\tau$ values could be either positive or negative and there were two symmetrical peaks around the line $\tau$=0° on the $d_{C\gamma\text{-}\pi\text{-}\tau}$ plane (Figure 3b). The most frequently occurring conformations in the trajectory had the similar geometric values to those in the crystal structure and QM models (Table 1). The MD data suggested that the Q-F pairs in the crystal structure should be maintained in solution to stabilize the protein, consistent with the observations in the CD experiments. For the third pair of Gln in the trailing and Phe in the leading chain, no interaction was detected in the MD trajectories (Figure S3).
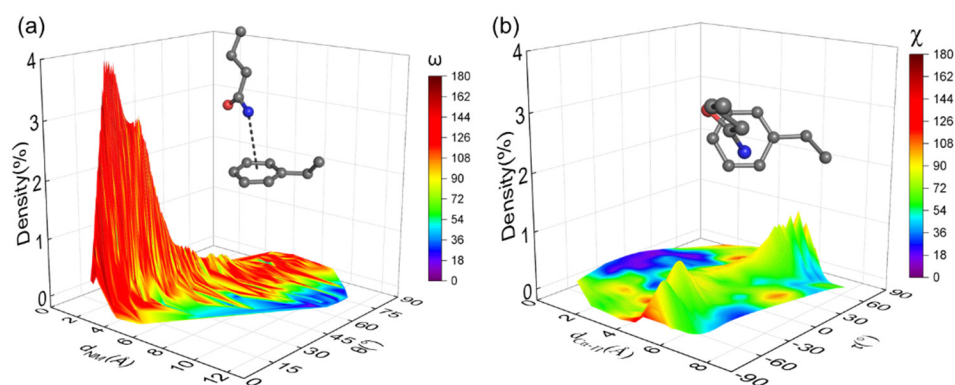


**Figure 3.** Molecular dynamics simulation of the collagen-like peptide GPQGFO. The geometric parameters of the Q-F pair between the leading and middle chain were extracted from the trajectory and plotted into three-dimensional density maps. (**a**) Conformation occurring percentage is plotted against $d_{NM}$ and $\theta$. The surface is color-coded with $\omega$. (**b**) Conformation occurring percentage is plotted against $d_{C\gamma\text{-}\pi}$ and $\tau$. The surface is color-coded with $\chi$. The color scales are located on the right. Insets of (**a, b**) show the most occurring conformations.

### 3.7. Data Mining of Protein Structures

In order to survey the situation of NH–$\pi$ interactions between Gln and Phe in proteins, a non-redundant protein structure dataset was built consisting of 9658 well-resolved (resolution ≤ 2 Å) crystal structures. Among the 1747 Q-F pairs ($d_{NM}$ ≤ 4.5 Å) in the dataset, ~5% of the pairs (94 out of 1747) fell within the cutoffs that defined NH–$\pi$ interactions ($d_{NM}$ < 4.3 Å, −25°< $\theta$ < 25°, and $\omega$ > 120°). The phenyl rings of the 94 Q-F pairs were superimposed and formed two umbrella-shaped distributions of Gln above and beneath the ring, where all of the N$\varepsilon$-H groups were positioned approximately perpendicular to the

ring center with similar $d_{NM}$, $\theta$ and $\omega$ values (Figure 4). The values of $d_{C\gamma\text{-}\pi}$ and $\tau$ range from 2.9 Å to 5.9 Å and $-74.7°$ to 56.2°, respectively. The Gln side chains in the axial Q-F pairs were located in the center of the umbrella and had extremely high $d_{C\gamma\text{-}\pi}$ (~5.3 Å) and $\tau$ (~61.1°). Notably, rather than uniformly distributed around the phenyl ring, the Gln side chains are less frequently projected toward the Phe $C\alpha$ with $\chi$ ranging from 4.6° to 175.7°, potentially preventing steric clashes. This finding suggests that NH–$\pi$ interactions could occur in proteins through diverse side chain conformations within a wide range of distances between Gln and Phe.

It is noted that Phe is hydrophobic while Gln is hydrophilic, which is relevant to their functions in the protein. To explore locations of the NH–$\pi$ interactions between them in proteins, i.e., the hydrophobic cores or the protein surfaces, the solvent accessible surface (SAS) area of the Q-F pairs were calculated with DSSP [44,45]. The majority (56 out of 94) of Q-F pairs were buried relatively deep inside the protein core (SAS < 30 Å$^2$) (Figure 4d). Eighteen Q-F pairs were partially exposed to solvent (60 Å$^2$ < SAS < 120 Å$^2$). Gln had higher SAS than Phe (Figure 4e). Except for the Q-F pairs in peptide GPQGFO, there were three fully exposed pairs (Figure S4). In one case, Gln and Phe were located in the i and i+4 positions of $\alpha$-helices, which may stabilize the $\alpha$-helices. SAS analysis indicated that the Gln-Phe NH–$\pi$ interactions could occur in both the core and the surface of proteins. Consistently, when the interaction was transferred from a vacuum to either low or high dielectric constant, simulating the protein core or surface, energy losses of only ~33% and ~25% were detected, respectively.
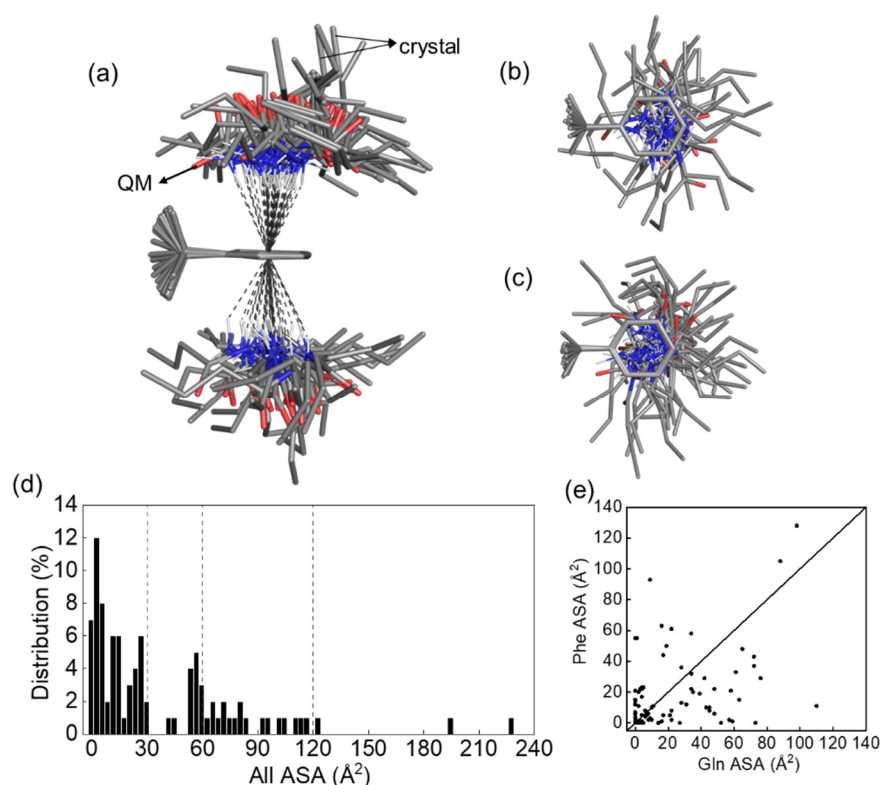


**Figure 4.** Q-F pairs with a typical NH–$\pi$ interaction extracted from the non-redundant dataset of protein crystal structures. (**a–c**) The Q-F pairs are positioned by superimposing Phe with each other. The side, bottom-up, and top-down views are shown in (**a**), (**b**) and (**c**), respectively. (**d–e**). Accessible surface (ASA) of the Q-F pairs were calculated. Distribution of the total ASA is shown in (**d**). The scatter plot of ASA of Gln vs. Phe is shown in (**e**).

## 4. Conclusions

From thermal unfolding experiments, we have demonstrated that NH–$\pi$ interactions allow the side chain of Gln to interact with the $\pi$-system of Phe in an axial geometry that

is less influenced by the environment. The interaction is energetically highly favorable with its magnitude similar to hydrogen bonding [40,41]. Computational and experimental data from structural biology, QM, and structural bioinformatics demonstrate that the polar N$\epsilon$-H group is typically positioned perpendicular to the phenyl ring and aligned with the phenyl ring center, which is well consistent with a previous study [1].

Two major features of the NH–$\pi$ interactions have been revealed. First, the NH–$\pi$ interactions are stable in various environments, i.e., the interaction energy is insensitive to dielectric constants. The stabilizing effect also remains almost constant under a wide range of pH and ionic strengths. When transferred from a vacuum to either low or high dielectric constant, the NH–$\pi$ interaction loses only ~33% and ~25% energy, respectively, while cation–$\pi$ interaction decays by ~70% and ~90% [42]. The Q-F pairs, both buried in the hydrophobic core or exposed on the protein surface, are identified in the data mining of the protein structures.

Secondly, the NH–$\pi$ interactions can be formed by diverse side chain conformations under a range of distances between Gln and Phe. In the GPQGFO structure, Gln and Phe, with C$\alpha$ atoms separated by 7.7 Å, made a strong NH–$\pi$ interaction. Multiple sub-optimal interacting conformations were observed in the QM models.

Similar interactions between side chain amino and aromatic groups, such as between Asn and Phe, can be further explored, which will enable the use of NH–$\pi$ interactions in the design of hydrophobic protein cores, solvent accessible protein surfaces, and protein–protein binding interfaces.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom12101433/s1, Figure S1: Energy surface of an acetamide-toluene complex with quantum chemical calculation; Figure S2: Scatter plot of the data points with $40° < \tau < 50°$ are projected on the $\Delta$E- $d_{C\gamma-\pi}$ lane and $d_{NM} \leq 4$ Å are projected on the $\Delta$E-$\theta$ plane; Table S1: Terms of interaction energy from SAPT2+3/aug-cc-pVQZ and MP2 methods; Table S2. Crystal statistics of crystal data collection and structure refinement one crystal was used for each structure; Figure S2: Molecular dynamics simulation of the collagen-like peptide GPQGFO; Figure S3: Molecular dynamics simulation of the collage-like peptide GPQGFO; Figure S4: Side chain interactions of Q-F pairs exposed on protein surface in the crystal structure of PDB ID:4OV4, 7ODC, and 3CK6; Figure S5: HPLC Chromatograms and Mass Spectrometry analysis of peptides; Figure S6: The electron density map of the Q-F pairwise.

**Author Contributions:** F.X., J.H. and R.Z. designed research; R.Z., Y.X. and J.L. collected and analyzed data; F.X., R.Z., Y.X., J.H. and S.F. composed the manuscript; Y.X and J.H provided quantum chemical calculation resources; J.L. and S.F. dissecting the crystal structure. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.
**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Steiner, T.; Koellner, G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305*, 535–557, doi:10.1006/jmbi.2000.4301.
2. Andrew, C.D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J.D.; Jones, G.R.; Doig, A.J. Stabilizing Interactions between Aromatic and Basic Side Chains in $\alpha$-Helical Peptides and Proteins. Tyrosine Effects on Helix Circular Dichroism. *J. Am. Chem. Soc.* **2002**, *124*, 12706–12714, doi:10.1021/ja027629h.
3. Mecozzi, S.; West, A.P., Jr.; Dougherty, D.A. Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. United States Am.* **1996**, *93*, 10566–10571, doi:10.1073/pnas.93.20.10566.

4.  Hughes, R.M.; Waters, M.L. Arginine methylation in a beta-hairpin peptide: implications for Arg-pi interactions, DeltaCp(o), and the cold denatured state. *J. Am. Chem. Soc.* **2006**, *128*, 12735–12742, doi:10.1021/ja061656g.

5.  Chen, C.C.; Hsu, W.; Hwang, K.C.; Hwu, J.R.; Lin, C.C.; Horng, J.C. Contributions of cation-pi interactions to the collagen triple helix stability. *Arch. Biochem. Biophys.* **2011**, *508*, 46–53, doi:10.1016/j.abb.2011.01.009.

6.  Burley, S.K.; Petsko, G.A. Amino-aromatic interactions in proteins. *FEBS Letters* **1986**, *203*, 139–143, doi:10.1016/0014-5793(86)80730-x.

7.  Cheng, J.; Kang, C.; Zhu, W.; Luo, X.; Puah, C.M.; Chen, K.; Shen, J.; Jiang, H. N-Methylformamide−Benzene Complex as a Prototypical Peptide N−H···π Hydrogen-Bonded System: Density Functional Theory and MP2 Studies. *J. Org. Chem.* **2003**, *68*, 7490–7495, doi:10.1021/jo026910b.

8.  Tóth, G.; Murphy, R.F.; Lovas, S. Investigation of Aromatic-Backbone Amide Interactions in the Model Peptide Acetyl-Phe-Gly-Gly-N-Methyl Amide Using Molecular Dynamics Simulations and Protein Database Search. *J. Am. Chem. Soc.* **2001**, *123*, 11782–11790, doi:10.1021/ja011245u.

9.  Rich, A.; Crick, F.H.C. The molecular structure of collagen. *J. Mol. Biol.* **1961**, *3*, 483–484, doi:10.1016/s0022-2836(61)80016-8.

10. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **1994**, *266*, 75–81, doi:10.1126/science.7695699.

11. Fraser, R.D.B.; MacRae, T.P.; Suzuki, E. Chain conformation in the collagen molecule. *J. Mol. Biol.* **1979**, *129*, 463–481, doi:10.1016/0022-2836(79)90507-2.

12. Persikov, A.V.; Ramshaw, J.A.M.; Kirkpatrick, A.; Brodsky, B. Amino Acid Propensies for the Collagen Triple-Helix. *Biochem.* **2000**, *39*, 14960–14967, doi:10.1021/bi001560d.

13. Shah, N.K.; Ramshaw, J.A.; Kirkpatrick, A.; Shah, C.; Brodsky, B. A host-guest set of triple-helical peptides: stability of Gly-X-Y triplets containing common nonpolar residues. *Biochemistry* **1996**, *35*, 10262–10268, doi:10.1021/bi960046y.

14. Persikov, A.V.; Ramshaw, J.A.; Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **2005**, *280*, 19343–19349, doi:10.1074/jbc.M501657200.

15. Ramshaw, J.A.; Shah, N.K.; Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J. Struct. Biol.* **1998**, *122*, 86–91, doi:10.1006/jsbi.1998.3977.

16. Beck, K.; Chan, V.C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J.A.; Brodsky, B. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. United States Am.* **2000**, *97*, 4273–4278, doi:10.1073/pnas.070050097.

17. Persikov, A.V.; Ramshaw, J.A.; Kirkpatrick, A.; Brodsky, B. Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* **2002**, *316*, 385–394, doi:10.1006/jmbi.2001.5342.

18. Persikov, A.V.; Ramshaw, J.A.M.; Kirkpatrick, A.; Brodsky, B. Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability. *Biochem.* **2005**, *44*, 1414–1422, doi:10.1021/bi048216r.

19. Chiang, C.H.; Horng, J.C. Cation-pi Interaction Induced Folding of AAB-Type Collagen Heterotrimers. *J. Phys. Chemistry. B* **2016**, *120*, 1205–1211, doi:10.1021/acs.jpcb.5b11189.

20. Chen, C.C.; Hsu, W.; Kao, T.C.; Horng, J.C. Self-assembly of short collagen-related peptides into fibrils via cation-pi interactions. *Biochem.* **2011**, *50*, 2381–2383, doi:10.1021/bi1018573.

21. Chiang, C.H.; Fu, Y.H.; Horng, J.C. Formation of AAB-Type Collagen Heterotrimers from Designed Cationic and Aromatic Collagen-Mimetic Peptides: Evaluation of the C-Terminal Cation-pi Interactions. *Biomacromolecules* **2017**, *18*, 985–993, doi:10.1021/acs.biomac.6b01838.

22. Walker, D.R.; Hulgan, S.A.H.; Peterson, C.M.; Li, I.C.; Gonzalez, K.J.; Hartgerink, J.D. Predicting the stability of homotrimeric and heterotrimeric collagen helices. *Nat. Chem.* **2021**, *13*, 260–269, doi:10.1038/s41557-020-00626-6.

23. Walker, D.R.; Alizadehmojarad, A.A.; Kolomeisky, A.B.; Hartgerink, J.D. Charge-Free, Stabilizing Amide-pi Interactions Can Be Used to Control Collagen Triple-Helix Self-Assembly. *Biomacromolecules* **2021**, *22*, 2137–2147, doi:10.1021/acs.biomac.1c00234.

24. Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkoczi, G.; Chen, V.B.; Croll, T.I.; Hintze, B.; Hung, L.W.; Jain, S.; McCoy, A.J.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst. D Struct Biol* **2019**, *75*, 861–877, doi:10.1107/S2059798319011471.

25. Zheng, H.; Lu, C.; Lan, J.; Fan, S.; Nanda, V.; Xu, F. How electrostatic networks modulate specificity and stability of collagen. *Proc. Natl. Acad. Sci. United States Am.* **2018**, *115*, 6207–6212, doi:10.1073/pnas.1802171115.

26. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst. D Biol Cryst.* **2004**, *60*, 2126–2132, doi:10.1107/S0907444904019158.

27. Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.* **2017**, *8*, doi:10.1002/wcms.1327.

28. Smith, D.G.A.; Burns, L.A.; Simmonett, A.C.; Parrish, R.M.; Schieber, M.C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; et al. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 184108, doi:10.1063/5.0006002.

29. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D Biol Cryst.* **1994**, *50*, 760–763, doi:10.1107/S0907444994003112.

30. Brooks, B.R.; Brooks, C.L., 3rd; Mackerell, A.D., Jr.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem* **2009**, *30*, 1545–1614, doi:10.1002/jcc.21287.

31.  Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935, doi:Doi 10.1063/1.445869.
32.  Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092, doi:Doi 10.1063/1.464397.
33.  Wennberg, C.L.; Murtola, T.; Pall, S.; Abraham, M.J.; Hess, B.; Lindahl, E. Direct-Space Corrections Enable Fast and Accurate Lorentz-Berthelot Combination Rule Lennard-Jones Lattice Summation. *J. Chem Theory Comput.* **2015**, *11*, 5737–5746, doi:10.1021/acs.jctc.5b00726.
34.  Xu, Y.; Huang, J. Validating the CHARMM36m protein force field with LJ-PME reveals altered hydrogen bonding dynamics under elevated pressures. *Commun. Chem.* **2021**, *4*, 99, doi:10.1038/s42004-021-00537-8.
35.  Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273, doi:Doi 10.1021/Ct3004000x.
36.  Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.L.; Grubmuller, H.; MacKerell, A.D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73, doi:10.1038/nmeth.4067.
37.  Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol* **2017**, *13*, e1005659, doi:10.1371/journal.pcbi.1005659.
38.  Wang, G.; Dunbrack, R.L., Jr. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res.* **2005**, *33*, W94-98, doi:10.1093/nar/gki402.
39.  Wang, G.; Dunbrack, R.L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589–1591, doi:10.1093/bioinformatics/btg224.
40.  Huang, J.; Meuwly, M. Explicit Hydrogen-Bond Potentials and Their Application to NMR Scalar Couplings in Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 467–476, doi:10.1021/ct9005695.
41.  Sheu, S.-Y.; Yang, D.-Y.; Selzle, H.L.; Schlag, E.W. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci.* **2003**, *100*, 12683–12687, doi:10.1073/pnas.2133366100.
42.  Kumar, K.; Woo, S.M.; Siu, T.; Cortopassi, W.A.; Duarte, F.; Paton, R.S. Cation–$\pi$ interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **2018**, *9*, 2655–2665, doi:10.1039/c7sc04905f.
43.  Rezac, J.; Riley, K.E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438, doi:10.1021/ct2002946.
44.  Touw, W.G.; Baakman, C.; Black, J.; te Beek, T.A.; Krieger, E.; Joosten, R.P.; Vriend, G. A series of PDB-related databanks for everyday needs. *Nucleic Acids. Res.* **2015**, *43*, D364-368, doi:10.1093/nar/gku1028.
45.  Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637, doi:10.1002/bip.360221211.