Dissecting C—H⋯π and N—H⋯π Interactions in Two Proteins Using a Combined Experimental and Computational Approach

Jia Wang1,2,3 & Lishan Yao1,2*

C—H⋯π and N—H⋯π interactions can have an important contribution for protein stability. However, direct measurements of these interactions in proteins are rarely reported. In this work, we combined the mutant cycle experiments and molecular dynamics (MD) simulations to characterize C—H⋯π and N—H⋯π interactions and their cooperativity in two model proteins. It is shown that the average C—H⋯π interaction per residue pair is ~ — 0.5 kcal/mol while the N—H⋯π interaction is slightly stronger. The triple mutant box measurement indicates that N—H⋯π⋯C—H⋯π and C—H⋯π⋯C—H⋯π can have a positive or negative cooperativity. MD simulations suggest that the cooperativity, depending on the local environment of the interactions, mainly arises from the geometric rearrangement when the nearby interaction is perturbed.

X—H⋯π interactions in biomolecules, where X can be C, N, O, or S, are weak and attractive interactions between the X—H component and aromatic groups. The high incidence in biomolecules makes X—H⋯π interactions an important contributor to the structure and function, and has led to an increasing number of theoretical and experimental studies devoted to characterization of such interactions8–10. Theoretical studies show that N—H⋯π, O—H⋯π and C—H⋯π can have very different optimum geometries, with the interaction strength order O—H⋯π > N—H⋯π > C—H⋯π4,11. The S—H⋯π interaction can be weaker8 or stronger12 than O—H⋯π, but is generally stronger than N—H⋯π and C—H⋯π8,12. The computational interaction energy of the indole-benzene dimer where the N—H⋯π interaction exists can reach −5.2 kcal/mol13. The computational interaction energies between benzene and CH4, NH3, H2S, and H2O are −1.4, −2.5, −2.9 and −3.0 kcal/mol, respectively8. The computational binding energies between indole and CH4, NH3, H2S, and H2O are −2.0, −2.6, −4.9 and −3.6 kcal/mol, respectively14. The importance of the S—H⋯π and C—H⋯π interactions in proteins has also been highlighted by their occurrence in the PDB database search5,11,14. The C—H⋯π interaction has been observed directly in proteins by nuclear magnetic resonance (NMR) spectroscopy methods where the across C—H⋯π J-coupling is detected16. Quantification of C—H⋯π in calix[4]pyrrole receptors yields a magnitude of −1 kcal/mol17. The C—H⋯π interaction in benzene–methane, ethane, propane, and butane, increases monotonically from −1.1 to −2.7 kcal/mol18–20. The measurement of C—H⋯π interactions in a cyclohexylalanine–phenylalanine pair in the core of a synthetic peptide indicates that each C—H⋯π contact can contribute about −0.7 kcal/mol to peptide stability21. In real proteins, C—H⋯π mainly occurs between an aliphatic side chain and an aromatic ring, or between two aromatic rings14. Although C—H⋯π interactions are well documented in proteins1, direct measurements of C—H⋯π and N—H⋯π strength in proteins are scarce.

Another important issue about X—H⋯π interactions is their cooperativity. Cooperativity is a central concept for understanding molecular recognition and supramolecular self-assembly22. By forming networks of weak interactions that compete against the entropy of flexible polypeptides, proteins fold into their biologically functional three-dimensional structures23. As a part of the interaction network, how X—H⋯π interactions coexist and cooperate in proteins is an important question. Only a few studies have addressed the X—H⋯π cooperativity, mainly in small molecules. The cooperativity of C—H⋯π interactions in small molecules is studied using molecular torsional balances24. The average C—H⋯π interaction strength increases as more C—H⋯π pairs are formed.

1Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China. 2Shandong Provincial Key Laboratory of Synthetic Biology, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China. 3University of Chinese Academy of Sciences, Beijing, 100049, China. *email: yaols@qibebt.ac.cn
suggested a positive cooperativity. This is opposite to the findings of an earlier computational study where the negative cooperativity is concluded for the same complexes\textsuperscript{25}. The C−H−π and N−H−π cooperativity in proteins remains largely unexplored.

In this work, we attempt to measure the C−H−π and N−H−π interactions in protein GB3 and staphylococcal nuclease (SNase). GB3 is the third immunoglobulin binding domain of protein G, a model protein that has been extensively studied\textsuperscript{26}. SNase is an enzyme that hydrolyzes nucleotides in DNA or RNA. A stable mutant of SNase, nuclease (SNase). GB3 is the third immunoglobulin binding domain of protein G, a model protein that has been

Figure 1. Putative C−H−π and N−H−π interactions in GB3 (panel A, pdb code: 2OED) Δ + PHS (pdb code: 3BDC). L5 and T18 interact with F30 whereas L5, I7, T16, and N37 interact with Y33 in GB3. L25, V74 and I92 are in contact with F34 in Δ + PHS.

Results
Experimental C−H−π and N−H−π interaction energies. Based on the X-ray crystal structures, a series of X−H−π interactions can be identified in GB3 and Δ + PHS (pdb code: 2OED and 3BDC, respectively). GB3 has five residue pairs that may form C−H−π interactions, L5−F30, T18−F30, L5−Y33, I7−Y33, and T16−Y33, and one residue pair N37−Y33 that can form the N−H−π interaction (Fig. 1A). Δ + PHS has three C−H−π interaction residue pairs, L25−F34, V74−F34, I92−F34 (Fig. 1B). All these C−H−π interactions are between a methyl group and an aromatic ring. A total of nine C−H−π interactions were characterized, including L5−F30, T18−F30, L5−Y33, I7−Y33, T16−Y33, and T16−F33 of GB3, and L25−F34, V74−F34, and I92−F34 of Δ + PHS. Two N−H−π interactions N37−Y33 and N37−F33 in GB3 were also measured. Furthermore, the introduction of triple mutant boxes (TMBs) generates additional 16 C−H−π and 4 N−H−π pairs (Table 1). Therefore, a total of 25 C−H−π and 6 N−H−π interactions were measured.

The folding free energies ΔG of all proteins were derived from the denaturation curves. The values of [D]\textsubscript{590}, m values for the wild type and mutant proteins are given in Supplementary Table S1. The magnitude of noncovalent interactions in the two proteins GB3 and Δ + PHS was obtained using the double mutant cycle (DMC) analysis\textsuperscript{36}. The values of C−H−π interactions are shown in Table 1, ranging from +0.31 (unfavorable) to −0.85 (favorable) kcal/mol, with 22 out of 25 showing favorable interactions. The three small positive interaction energies may come from the secondary interactions, i.e., the interaction changes from the surrounding residues caused by mutations (a caveat of the DMC experiment). The residual secondary interactions can contribute to the measured XH−π energy which may change the sign of the energy (to repulsive) if it is small. The interaction energy of N−H−π ranges from −0.15 to −0.86 kcal/mol.

According to DMC, it is preferable to mutate the two side chains x and y in the X−H−π pair to alanine residues to completely remove the interactions between the two. However, eliminating an aromatic residue in a protein core can be detrimental to protein stability. Instead, we only mutated the aromatic side chain (y) to a leucine (y′) which is still hydrophobic but disrupts the X−H−π interaction (see more details in Materials and Methods). For the X−H component (x), conservative mutations are introduced (x′) to remove the X−H−π interaction and maintain the protein folding at the same time. These mutations may create residual pairwise side chain interactions in x′y′, xy′, and x′y. Furthermore, for a residue like leucine (for example, in L5−F30) which has two CH3 and one CH, it can form multiple C−H−π interactions which complicate the interpretation of the experimental results. These problems can be solved with the assist of MD simulations.

Benchmark of MD simulations. MD simulations were performed for all the experimentally measured mutants with three commonly used force fields, Amber99sb\textsuperscript{39}, Charmm27\textsuperscript{30}, and Gromos53a6\textsuperscript{41}. The...
Experimental Φ−π and Φ−H−π interactions were used as a benchmark to evaluate the accuracy of different force fields. The root mean square deviation (RMSD) between the experimental and predicted Φ−H−π interactions was calculated:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(\Delta \Delta G_{\text{exp}} - \Delta \Delta E_{\text{MD}})^2}{N}}$$

where \(N\) is 31, the total number of measured residue pairs that form Φ−H−π interactions, \(\Delta \Delta G_{\text{exp}}\) is the experimental Φ−H−π interaction energy, and \(\Delta \Delta E_{\text{MD}}\) is the calculated interaction energy. Charmm27 appears to perform better than the other two force fields. Its RMSD value is 0.27 kcal/mol (after removing two apparent outliers), while the RMSDs of Amber99sb and Gromos53a6 are 0.41 and 0.47 kcal/mol, respectively (Fig. 2). Thus, the trajectories produced using Charmm27 were selected for further analyses.

**Geometric parameters of C−H−π and N−H−π interactions.** The reasonable correlation between the interaction energies from MD simulations and experiments encourages us to investigate the geometric parameters that are important for C−H−π and N−H−π interactions. The pairwise interaction energy \(\Delta \Delta E_{\text{CH3-π}}\) between a CH3 group and a aromatic ring was calculated for all the C−H−π interactions identified above. Two geometric parameters, \(d_{\text{C-H}}\) and \(\omega\), are defined for the C−H−π interaction, where \(d_{\text{C-H}}\) is the distance of the methyl carbon to the center of mass of the aromatic ring (X), and \(\omega\) is the C−H−X angle (Fig. 3A). Since there are three methyl hydrogens, the one with the largest C−H−X angle is defined as \(\omega\). The same geometric parameters can also be defined for N−H−π interactions (Fig. 3B). The 3D plot of \((d_{\text{C-H}}, \omega)\) versus \(\Delta \Delta E_{\text{CH3-π}}\) shows that the geometries with shorter \(d_{\text{C-H}}\) and larger \(\omega\) have more negative interaction energies (Fig. 3C). The distance appears to be

| Protein     | Residue pair | Interaction (kcal/mol) | Protein     | Residue pair | Interaction (kcal/mol) |
|-------------|--------------|------------------------|-------------|--------------|------------------------|
| C−H−π       |              |                        |             |              |                        |
| GB3 (WT)    | L5−F30       | −0.81 ± 0.05           | GB3 (WT)    | L5−Y33       | −0.65 ± 0.06           |
| GB3 (WT)    | L5−Y33       | −0.13 ± 0.06           | GB3 (WT)    | L5−Y33       | −0.85 ± 0.06           |
| GB3 (WT)    | I7−Y33       | −0.13 ± 0.06           | GB3 (Y33F-N37A) | T16−F33   | −0.29 ± 0.06           |
| GB3 (WT)    | T16−Y33      | −0.70 ± 0.05           | Δ + PHS (WT) | L25−F34     | −0.41 ± 0.05           |
| GB3 (WT)    | T16−F33      | −0.68 ± 0.05           | Δ + PHS (WT) | V74−F34     | −0.53 ± 0.07           |
| GB3 (WT)    | L18−F30      | −0.08 ± 0.05           | Δ + PHS (WT) | L19−F34     | −0.46 ± 0.06           |
| GB3 (L5V)   | T16−Y33      | −0.29 ± 0.06           | Δ + PHS (L25V) | V74−F34    | −0.83 ± 0.05           |
| GB3 (L5V)   | T18−F30      | 0.09 ± 0.06            | Δ + PHS (L25V) | L19−F34    | −0.84 ± 0.06           |
| GB3 (T16A)  | T16−Y33      | −0.25 ± 0.06           | Δ + PHS (V74A) | L25−F34    | −0.72 ± 0.04           |
| GB3 (T16A)  | L5−Y33       | −0.26 ± 0.06           | Δ + PHS (V74A) | I92−F34    | −0.51 ± 0.05           |
| GB3 (T16A)  | I7−Y33       | 0.31 ± 0.05            | Δ + PHS (I92V) | L25−F34    | −0.81 ± 0.08           |
| GB3 (T18A)  | I7−F30       | −0.65 ± 0.06           | Δ + PHS (I92V) | V74−F34    | −0.59 ± 0.07           |
| GB3 (N37A)  | L5−Y33       | −0.34 ± 0.06           |                        |             |                        |
| N−H−π       |              |                        |             |              |                        |
| GB3 (WT)    | N37−Y33      | −0.68 ± 0.05           | GB3 (T16A)  | N37−Y33      | −0.86 ± 0.06           |
| GB3 (L5V)   | N37−Y33      | −0.37 ± 0.07           | GB3 (Y33F)  | N37−F33      | −0.83 ± 0.06           |
| GB3 (T17V)  | N37−Y33      | −0.45 ± 0.06           | GB3 (Y33F-T16A) | N37−F33   | −0.15 ± 0.06           |

Table 1. Experimental interaction energies of C−H−π and N−H−π interactions from double mutant cycle analysis.
the most important parameter, with the energy dropping quickly as the distance decreases. Meanwhile, the angle \( \omega \) can also be important. The average \( \Delta E_{\text{HS-anti}} \) for all the C–H⋯π interactions is \(-0.36\) kcal/mol. The number of N–H⋯π interactions is less than that of C–H⋯π, and they appear to be stronger than C–H⋯π interactions with the same geometric parameters.

\[ \Delta \Delta G_{\text{coop}} \] from TMB measurements. On the basis of double mutant cycles we had, we established several TMBs to elucidate the cooperativity in C–H⋯π–C–H⋯π and C–H⋯π–N–H⋯π interactions. In protein GB3, the cooperativity is positive in L5–F30–T18, L5–Y33–T16, I7–Y33–T16, L5–Y33–N37, I7–Y33–N37, and T16–F33–N37, with \( \Delta \Delta G_{\text{coop}} \) varied from \(-0.16\) to \(-0.55\) kcal/mol (Supplementary Table S2, Fig. 4), suggesting that they are cooperative with each other. In contrast, the C–H⋯π–N–H⋯π in T16–Y33–N37 of GB3, and the C–H⋯π–C–H⋯π in L25–F34–V74, L25–F34–I92, and V74–F34–I92 of \( \Delta + \text{PHS} \) are anticooperative, with \( \Delta \Delta G_{\text{coop}} \) varied from 0.04 to 0.37 kcal/mol (Supplementary Table S2). The cooperativity difference in different C–H⋯π–C–H⋯π and C–H⋯π–N–H⋯π suggests that it depends on the local interaction network.

Cooperativity mechanism from MD simulations. The \( \Delta \Delta G_{\text{coop}} \) are in a good correlation with the computational \( \Delta \Delta E \) (cooperativity energy, see more details in Materials and Methods), although the absolute value of \( \Delta \Delta G_{\text{coop}} \) is generally larger than that of \( \Delta \Delta G_{\text{coop}} \) (Fig. 4). One likely cause is that the entropic contribution, which is not calculated in MD simulations, may offset the large change of \( \Delta \Delta E \). The entropy calculation is far more difficult (less reliable) and thus not pursued. As discussed above, the residual interactions caused by the experimental non-alanine mutations complicate the interpretation of \( \Delta \Delta G_{\text{coop}} \). To solve this problem, we rebuilt TMBs by mutating the three side chains, for example L25, F34, and V74 in L25–F34–V74, to alanines systematically in MD simulations. The cooperativity energy \( \Delta \Delta E \) was calculated for the residue groups listed above with the same procedure (Fig. 5A). The cooperativity from \( \Delta \Delta E \) generally agrees with that from \( \Delta \Delta E \), except that L5–Y33–T16 and I7–Y33–T16 show a weak negative instead of positive cooperativity.

The cooperativity energy \( \Delta \Delta E \) varies from \(-0.39\) to 0.16 kcal/mol (Fig. 5A). Although they appear to be small, the percentagewise \( \Delta \Delta E \) (\( \Delta \Delta E \) divided by the average of the two C–H⋯π interactions in
C−H⋯π−C−H⋯π or the average of the C−H⋯π and the N−H⋯π interaction energy in C−H⋯π−N−H⋯π can vary from −40% (cooperative) to +60% (anticooperative) (Fig. 5B). So it is obvious that cooperativity can be very important for C−H⋯π and N−H⋯π interactions in an interaction network. To further understand the origin of cooperativity, the geometric changes in the TMB are investigated. It is known that $d_{\text{CX}}$ or $d_{\text{CX,MUT}}$ (Fig. 3) is an important parameter for C−H⋯π or N−H⋯π. Using L5−Y33−T16 as an example, $\Delta d$, the change of $d_{\text{CX}}$, was calculated by

$$
\Delta d = d_{\text{CX,Wt}} - d_{\text{CX,MUT}}
$$

(2)

where $d_{\text{CX,Wt}}$ is $d_{\text{CX}}$ between the methyl of L5 and the aromatic side chain of Y33 in the wild type, and $d_{\text{CX,MUT}}$ is $d_{\text{CX}}$ in the single mutant T16A. A similar $\Delta d$ can be defined for C−H⋯π−N−H⋯π interactions. $\Delta d$ was calculated for 10 residue groups shown in Fig. 5. The positive $\Delta d$ corresponds to the increase of the first C−H⋯π (or N−H⋯π) distance when the aliphatic side chain of the second C−H⋯π (or N−H⋯π) is mutated to alanine. In other words, removing the second C−H⋯π (or N−H⋯π) interaction weakens the first C−H⋯π (or N−H⋯π) interaction, suggesting a positive cooperativity. For 9 out of 10 groups, the distance change $\Delta d$ predicts the cooperativity consistent with the interaction energy result (Fig. 5B,C), indicating that the cooperativity in C−H⋯π−C−H⋯π or C−H⋯π−N−H⋯π mainly arises from the geometric rearrangement.

Discussion

DMC experiments are commonly used to measure residue−residue interactions, such as salt bridges and hydrogen bonds. However, measuring C−H⋯π interactions in the protein interior using DMC can be challenging because removing an aromatic side chain can destabilize and even unfold the protein. In this work, we only mutate the aromatic residue to leucine which maintains the protein folding and removes the C−H⋯π interaction. Two very stable proteins GB3 and Δ+ PHS were selected for the purpose. One caveat of the F or Y to L mutation is that side chain interactions with leucine complicate the data interpretation. Molecular dynamics simulations were used to decompose the various contributions and help us focus on the C−H⋯π interactions. The good agreement between experimental and computational interaction energies validates the procedure which provides important insights about the C−H⋯π and N−H⋯π interactions.

The energy of C−H⋯π interactions obtained from the DMC experiments of two proteins in this work is smaller than $\sim 0.9$ kcal/mol, with an average of $\sim 0.3$ kcal/mol. This C−H⋯π interaction strength is generally weaker than those reported for small molecules. It is likely that different interactions compete with each other in proteins so that the C−H⋯π interaction of a specific residue pair is not in an optimum geometry. This is evident from the interaction energy landscape of methyl−aromatic ring pair (Fig. 3). The lower corner, with $d_{\text{CX}}$ of $\sim 0.4$ nm and $\omega$ of $\sim 165^\circ$, has the lowest interaction energy in the plot. But many C−H⋯π pairs are clustered around $d_{\text{CX}}$ of $0.4−0.6$ nm and $\omega$ of $120^\circ−150^\circ$. The optimal $d_{\text{CX}}$ of 0.4 nm is close to the distance obtained from the quantum mechanical calculations. For C−H⋯π pairs with larger $d_{\text{CX}}$, the C−H group moves away from the top of the aromatic ring to form a side-by-side configuration which has an optimal $d_{\text{CX}}$ of $\sim 0.5$ nm, as suggested from the QM calculations. The non-optimum geometry also implies that different C−H⋯π interactions with the same aromatic ring are interdependent. A small perturbation of one C−H⋯π pair may affect the geometry of another C−H⋯π pair nearby which creates the cooperativity effect.

The cooperativity analysis from TMB clearly suggests that the C−H⋯π−C−H⋯π and C−H⋯π−N−H⋯π can be either cooperative or anticooperative (Fig. 4). Although in the experimental TMB analysis, the cooperativity information is contaminated by the residual interactions in the mutants, the computational TMB analysis where the residual interactions are removed suggests that the side chain C−H⋯π and N−H⋯π interactions have a major contribution to the experimentally determined $\Delta \Delta G$ (Figs. 4 and 5). Moreover, the $d_{\text{CX}}$ or $d_{\text{CX,MUT}}$ distance change $\Delta d$ is an important indicator for the cooperativity. But when comparing the computational cooperativity energy $\Delta \Delta E$ and $\Delta d$, the linear correlation between the two is only moderate, suggesting that the distance change is not the only contributor to the cooperativity change. The change of angles such as $\omega$ may also play a role.

Two simpler cooperativity models were built using two methane and one benzene molecules, with methanes on the same side (MMB) or opposite side (MBM) of the benzene. The cooperativity energies of MMB and MBM models were calculated at the MP2/aug-cc-pvtz level. According to the quantum mechanical (QM) calculations, the cooperativity energy of MMB is 0.74 kcal/mol, indicating that C−H⋯π−C−H⋯π is anticooperative in this
model, while the cooperativity energy of MBM is 0.03 kcal/mol, suggesting that there is no cooperativity in this model. Similar to the result in the MD simulations, the geometric reorganization occurs in the MMB model where the two methanes compete for the binding site. No such competition exists in the MBM model where the cooperativity energy is close to zero. The QM calculations highlight the importance of geometric reorganization to cooperativity.

Conclusion
In this study, we measured the strength of C−H⋯π and N−H⋯π interactions in GB3 and SNase. The C−H⋯π interaction is about 0.3 to −0.9 kcal/mol whereas the N−H⋯π interaction is about −0.2 to −0.9 kcal/mol. The energy decomposition from MD simulations helps determine the C−H⋯π and N−H⋯π interactions for individual methyl−aromatic and amino−aromatic pairs and identify important geometric parameters δCGUX and ω. The experimental TMB analysis suggests that the cooperativity of X−H⋯π interactions can be either positive or negative, depending on the local environment. The cooperativity trend is successfully captured by MD simulations where the cooperativity energy can reach ~ −40% to 60% of C−H⋯π or N−H⋯π interactions, highlighting its importance in proteins. The geometric rearrangement is the main cause for the cooperative interactions. It is worth noting that the C−H⋯π and N−H⋯π interactions and the cooperativity were only measured for two proteins GB3 and Δ + PHS. More measurements will be needed to see whether the conclusions also hold for other proteins. But we expect that the mechanism behind the interactions is universal for all protein molecules.

Materials and Methods
Protein expression and purification. The wild type and mutants of GB3 and Δ + PHS were prepared with the PCR-based site-directed mutagenesis on vector pET-11b. These plasmids were transformed into the E. coli strain BL21 (DE3) cells for protein expression. The purification procedure for GB3 and its variants has been described previously35. Δ + PHS and its variants were purified using the same procedure as described by Shortle and Meeker36.

Thermodynamic stability measurements. All the denaturation measurements were performed using a HITACHI f-4600 fluorescence Spectrophotometer. Mixtures consisted of up to 6.0 M GdnHCl and 50 μM proteins (final concentration) were incubated for 30 min at 30°C. The signal intensity at 340 nm for GB3 and 348 nm for SNase was extracted and fitted using the following equation,

$$S = \frac{(\alpha_N + \beta_N[D]) + [(\alpha_U + \beta[D])\exp[\{m(\Delta - [D]_{50\%})/RT\}]}{1 + \exp[m(\Delta - [D]_{50\%})/RT]}$$

where S is the measured Fluor340nm or Fluor348nm, m and α are the intercepts and β and β are the slopes of the Fluor340nm or Fluor348nm baselines at low (N) and high (U) denaturant concentrations, R is the Boltzmann constant, T is the temperature, [D] is the denaturant concentration, [D]_{50\%} is the denaturant concentration at which the protein is 50% denatured.

Double mutant cycle analysis. Double mutant cycle (DMC), proposed by Fersht and co-workers, can eliminate the contribution of the secondary interactions and obtain accurate binding energy for the interaction between two residues37,38. Double mutant cycles were performed to quantify C−H⋯π interactions and N−H⋯π interactions in this work. To build the DMC, dozens of single and double mutants were prepared. Single mutants included L5V, I7V, T16A, T18A, N37A, F30L, Y33L, Y33F in GB3 and L25V, V74A, I92V, F34L in Δ + PHS. Double mutants contained two substitutions, L5V-I7V, T16A-W18A, N37A-D30L, Y33A-W33F in GB3, and L25V-V74A, I92V-F34L in Δ + PHS. The geometric rearrangement occurs in the MMB model where the two methanes compete for the binding site. No such competition exists in the MBM model where the cooperativity energy is close to zero. The QM calculations highlight the importance of geometric reorganization to cooperativity.

Triple mutant box analysis. Two double mutant cycles can be combined to produce a TMB, which can be used for quantification of cooperative effects. Extensive studies have been performed by Hunter and co-workers using triple mutant box experiments to evaluate cooperativity in non-covalent interactions39,40. Double mutants of GB3 (L5V-I7V, L5V-T16A, L5V-T18A, I7V-T16A, L5V-N37A, I7V-N37A, and T16A-N37A) and Δ + PHS (L25V-V74A, L25V-I92V and L74A-I92V) were used to set TMBs. All of these double mutant proteins could be expressed except L5V-I7V of GB3. Triple mutants were prepared, including L5V-T16A-Y33L, L5V-T18A-F30L, I7V-T16A-Y33L, L5V-N37A-Y33L, I7V-N37A-Y33L, T16A-N37A-Y33L, and T16A-N37A-F30L for GB3, and L25V-V74A-F34L, L25V-F34L-I92V, and V74A-I92V-F34L for Δ + PHS. These mutants were used to quantify the cooperativity in C−H⋯π−C−H⋯π interactions and C−H⋯π−N−H⋯π interactions. The folding free energy for each mutant was measured using the same method mentioned above. The cooperativity energy was then calculated using:

$$\Delta \Delta G = \Delta G_{xy} - \Delta G_{x'y'} - \Delta G_{y'y} + \Delta G_{x'y}$$

where $\Delta G_{xy}$, $\Delta G_{x'y}$, $\Delta G_{y'y}$, and $\Delta G_{xy'}$ are the folding free energy for the wild type protein xy, single mutants x'y and y'x, and the double mutant x'y', respectively. The symbols x and y denote the aliphatic and aromatic side chains in the C−H⋯π or N−H⋯π interaction energy with the aromatic ring was then calculated using:

$$\Delta \Delta G_{xy} = \Delta G_{xy} - \Delta G_{x'y'} - \Delta G_{y'y} + \Delta G_{x'y}$$

where $\Delta G_{xy}$, $\Delta G_{x'y}$, $\Delta G_{y'y}$, and $\Delta G_{xy'}$ are the folding free energy for the wild type protein xy, single mutants x'y and y'x, and the double mutant x'y', respectively. The symbols x and y denote the aliphatic and aromatic side chains in the C−H⋯π or N−H⋯π interaction energy with the aromatic ring was then calculated using:

$$\Delta \Delta G_{xy} = \Delta G_{xy} - \Delta G_{x'y'} - \Delta G_{y'y} + \Delta G_{x'y}$$

where $\Delta G_{xy}$, $\Delta G_{x'y}$, $\Delta G_{y'y}$, and $\Delta G_{xy'}$ are the folding free energy for the wild type protein xy, single mutants x'y and y'x, and the double mutant x'y', respectively. The symbols x and y denote the aliphatic and aromatic side chains in the C−H⋯π or N−H⋯π interaction energy with the aromatic ring was then calculated using:
\[
\Delta \Delta G_{\text{coop}} = \Delta G_{xy} - \Delta G_{xyz} + \Delta G_{x'z'}
\]

where \( y \) represents the aromatic residue, \( x \) and \( z \) represent nonaromatic residues, \( \Delta G_{xy} \), \( \Delta G_{xyz} \), \( \Delta G_{x'z'} \), and \( \Delta G_{x'y'z'} \) are the folding free energy of the wild type protein \( x'yz \), single mutants \( x'y'z \), \( x'y'z' \), double mutants \( x'y'z' \), and triple mutants \( x'y'z' \), respectively.

Molecular dynamics simulations. MD simulations were performed using the GROMACS 4.5 package\(^{40}\) with Amber99sb\(^{29}\), Charmm27\(^{30}\), or Gromos53a6\(^{31}\) force fields. The structures of all variants of GB3 and \( \Delta + \text{PHS} \) were produced by FoldX\(^{41}\) with the protein backbone fixed. Each protein was solvated by adding 10.0 Å TIP3P water\(^{42}\) (or SPC water when the Gromos53a6 force field was used) in a rectangular box, and counter ions were used to neutralize the system. 500,000 steps of energy minimization followed by 1 ns MD simulation at constant pressure (1 atm) and temperature (303 K) were performed to equilibrate the system before the production run. Three 10 ns MD production runs with different random starting velocities were performed with snapshots saved every 50 ps which were then used in the data analysis and error estimation. All backbone heavy atoms are restrained in the equilibrium and production runs. Temperature was regulated by a modified Berendsen thermostat\(^\text{1}^{43}\) and pressure was controlled by the extended ensemble Parrinello-Rahman approach\(^\text{1}^{44,45}\). The long-range electrostatic interactions were evaluated by the Particle mesh Ewald method\(^\text{1}^{46,47}\). The nonbonded pair list cutoff was 10 Å and the list was updated every 10 fs. The LINCS algorithm\(^\text{48}\) was used to constrain all bonds linked to hydrogen in the protein, whereas the SETTLE algorithm\(^\text{49}\) was used to constrain bonds and angles of water molecules, allowing a time step of 2 fs. In the energy decomposition analysis, only the interaction energy between the paired residues of \( \text{C}^-\text{H}--\pi \) or \( \text{N}--\pi \) was calculated. The computational interaction energy \( \Delta \Delta E \) was calculated by

\[
\Delta E_{xy} = E_{xy} = \frac{E_{xy} - E_{xy - J}}{J}
\]

where \( \Delta E_{xy} \), \( \Delta E_{x'y'} \), \( \Delta E_{xy'z'} \), and \( \Delta E_{x'y'z'} \) are the interaction energy of \( x'y'z \), single mutant \( x'y'z \), and \( x'y'z' \) in the double mutant \( x'y'z' \), and \( x'y'z' \) and \( x'y'z' \) in the wild type protein, \( x'y'z \) and \( x'y'z' \) respectively.

QM calculations. Two methane and one benzene molecules were built to model the cooperativity of \( \text{C}^-\text{H}--\pi \) and \( \text{H}--\pi \). The geometries of the two models, MMB and MBM, were optimized at the MP2/6-31+G(d,p)\(^{20}\) level. The energy calculations were performed at the MP2/aug-cc-pvtz\(^{14}\) level. All the calculations were done using the Gaussian 09 software\(^{15}\).

Received: 4 September 2019; Accepted: 12 December 2019; Published online: 27 December 2019

References

1. Nishio, M., Umezawa, Y., Fantini, J., Weiss, M. S. & Chakrabarti, P. CH-pi hydrogen bonds in biological macromolecules. Phys. Chem. Chem. Phys. 16, 12648–12683 (2014).
2. Neel, A. J., Hilton, M. J., Sigman, M. S. & Toste, F. D. Exploiting non-covalent pi interactions for catalyst design. Nature. 543, 637–646 (2017).
3. Parsons, Z. D., Bland, J. M., Mullins, E. A. & Eichman, B. F. A Catalytic Role for C-H/pi Interactions in Base Excision Repair by Bacillus cereus DNA Glycosylase A lkD. J. Am. Chem. Soc. 138, 11485–11488 (2016).
4. Mohan, N., Vijayalakshmi, K. P., Koga, N. & Suresh, C. H. Comparison of aromatic NH--\pi, OH--\pi, and CH--\pi interactions of alamine using MP2, CCSD, and DFT methods. J. Comput. Chem. 31, 2874–2882 (2010).
5. Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. Angew. Chem. Int. Ed. 42, 1210–1250 (2003).
6. Kamps, J. J. A. G. et al. Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. Nat. Commun. 6, 8911–8911 (2015).
49. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 13, 952–962 (1992).
50. Head-Gordon, M., Pople, J. A. & Frisch, M. J. MP2 energy evaluation by direct methods. Chem. Phys. Lett. 153, 503–506 (1988).
51. Frisch, M. J. T. et al. Gaussian 09, revision B.01; Gaussian, Inc.: (Wallingford, CT, 2010).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (Grant nos. 21773280 and 3166143036), Natural Science Foundation of Shandong Province (Grant no. ZR2018ZB0207), and the Taishan Scholars Program of Shandong Province.

Author contributions
J.W. and L.Y. conceived the project. J.W. and L.Y. planned the experiments and MD simulations. J.W. performed the experiments and MD simulations. J.W. and L.Y. analyzed the data. J.W. and L.Y. wrote the paper. All authors approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-56607-4.

Correspondence and requests for materials should be addressed to L.Y.

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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019