Principles of Inter-Amino-Acid Recognition Revealed by Binding Energies between Homogeneous Oligopeptides

Huiwen Du,†,# Xiaoyu Hu,‡ Hongyang Duan,†,# Lanlan Yu,†,# Fuyang Qu,†,# Qunxing Huang,†,# Wangshu Zheng,‡,# Hanyi Xie,‡,# Jiaxi Peng,‡,# Rui Tuo,§ Dan Yu,† Yuchen Lin,†,# Wenzhe Li,†,# Yongfang Zheng,‡,# Xiaocui Fang,†,# Yimin Zou,†,# Huayi Wang,‡,# Mengting Wang,‡,# Paul S. Weiss,‡,# Yanliang Yang,‡,# and Chen Wang∗,†,#

ABSTRACT: We have determined the interaction strengths of the common naturally occurring amino acids using a complete binding affinity matrix of 20 × 20 pairs of homo-octapeptides consisting of the 20 common amino acids between stationary and mobile states. We used a bead-based fluorescence assay for these measurements. The results provide a basis for analyzing specificity, polymorphisms, and selectivity of inter-amino-acid interactions. Comparative analyses of the binding energies, i.e., the free energies of association (ΔG), reveal contributions assignable to both main-chain-related and side-chain-related interactions originating from the chemical structures of these 20 common amino acids. Side-chain–side-chain and side-chain–main-chain interactions are found to be pronounced in an identified set of amino acid pairs that determine the basis of inter-amino-acid recognition.

Natural proteins contain primarily 20 common amino acids that are distinguished by their chemical side chains. The primary structures of proteins are defined by their sequences of amino acids, which are critically important for protein folding and higher order protein structures and functions.1–4 As with the definition of the primary structures of proteins, peptides are formed from different amino acids, NH2CHRCOOH (where R represents the side chain of the amino acid), conjugated through the amide (–CONH–) link that is the repeating unit of the main chain of the peptide. The main chain contains two moieties, N–H and C==O, which can participate in forming hydrogen bonds. Consequently, the interactions of all of the 20 common amino acids with one another are important in analyzing and predicting protein–protein interactions.5 However, comprehensive quantitative measurements of the inter-amino-acid interactions have not previously been reported.

Inter-amino-acid interactions are reflected in the interpeptide binding affinities (Kb) or binding energies, i.e., free energies of association (ΔG), through the relationship ΔGₐ = RT ln Kb. The interpeptide interactions and those between amino acids are fundamentally correlated since the binding energies are dependent on side-chain–side-chain, side-chain–main-chain, and main-chain–main-chain interactions. In particular, side-chain-related interactions are central to inter-amino-acid interactions and to inter-amino-acid recognitions. Therefore, peptides with homologous sequences and structures should be ideal for determining inter-amino-acid interaction strengths. Given that interaction strengths between proteins vary over a broad range, we expect that the interaction strengths of amino acids should also vary widely. Specifically, inter-amino-acid interactions in the strong interaction regime play fundamental roles in interprotein recognition, as indicated in many biological processes such as antigen–antibody recognition, ligand–receptor interactions, and protein aggregation relating to a broad range of diseases.

The binding energies between biomolecules can be determined by measuring binding affinities. The experimental
quest for quantitative measurements of binding properties of the 20 common amino acids has been challenging due to the significant differences in the chemical structures of side chains (including size, charge, and hydrophobicity) and the sensitivity limitations of the available techniques.5 To our knowledge, no complete set of measurements of the binding characteristics of all the common naturally occurring amino acids under identical conditions has been reported. In related recent efforts, microbead-based and cell-surface-based fluorescence method using flow cytometry (FCM)6–8 has shown promise in obtaining binding affinities of protein–peptide interactions,9 and protein association with chromatin.10 The measured maximal bead fluorescence and median channel fluorescence were shown to be correlated to the binding affinities between proteins immobilized on microbeads and peptides in the mobile phase.9 In another effort, quantitative analyses of the multivalent binding between ligands and nanoparticles with low ligand valency revealed significant heterogeneity in the measured nanoparticle–ligand distributions, which can be characterized by Poisson distributions.11 It was suggested that a threshold valence is required in order to achieve homogeneous binding distributions.12 Therefore, distributions of species formed by ligand-microbead or nanoparticles should be examined statistically in order to perform quantitative assessments of binding affinities.

One approach to determine the inter-amino-acid interactions is to study the binding of peptides with designated sequences. In particular, peptides with homologous structures should have high fidelity in determining the binding of the 20 common amino acids with one another. Secondary structures should be disfavored in such studies in order to ensure that peptide main chains are ultimately accessible for interpeptide interactions, which is essential for assessing main-chain-related interactions, and clarifying the determinant interactions between amino acids. Previous reports suggest that peptides with fewer than 15 residues are unlikely to develop helices due to the lack of long-range (interhelical) interactions.4,13–15 In addition, β-sheet structures can be avoided by controlling experimental conditions such as solvent polarity and incubation time. In this work, homo-octapeptides were selected for minimizing ambiguity due to sequence effects, reducing the structural fluctuations of terminal moieties, and minimizing conformational effects due to secondary structures, so that quantitative measurements of binding affinity can be made. We introduce a microbead-based FCM method to determine the peptide–peptide binding affinities under constant experimental conditions. The methodology to determine the interpeptide interactions is presented below.16

Let \(I_i\) be the fluorescence threshold for measuring the nonfluorescently tagged polystyrene (PS) microbeads in FCM and \(N_i\) be the number of PS microbeads with measured fluorescence intensity higher than \(I_i\). We write \(N\) as the total number of PS microbeads. Finally, we denote \(X\) as the number of PS microbeads with fluorescence intensity higher than \(I_i\) due to fluorescently tagged peptides bound on the PS microbead surface.

Assuming that each microbead has the same number of binding sites and each binding site binds with peptides independently, \(X\) will have a binomial distribution according to probability model [page 81 in ref 17]: i.e., the probability that \(X\) is equal to a non-negative integer \(k\) is \(\binom{N}{k}p_1^k(1-p_1)^{N-k}\), where \(p_1 = P(I > I_i)\), which represents the probability of PS microbeads with measured fluorescence intensity higher than \(I_i\).

According to the strong Law of Large Numbers (by Borel) [page 202 in ref 18], if the total number of PS microbeads, \(N_i\) is sufficiently large,19 the ratio of the measured number of PS microbeads with fluorescence intensity above \(I_i\) due to adsorbed peptides, \(N_i\), to \(N\) should be approximately equal to \(p_1\), i.e.,

\[
p_1 \approx \frac{N_i}{N}
\]

Note that the fluorescence threshold, \(I_i\), is equivalent to the threshold number of peptides with fluorophores on an individual microbead, \(k_1\) (not measurable in this study), and therefore is equivalent to the threshold peptide coverage of \(\theta_i\) on the microbead surface. As a result, the probability of peptide coverage higher than \(\theta_i\) is equal to \(p_1\).

According to the Central Limit Theorem (by De Moivre) [page 244 in ref 18], the distribution of peptide coverage on the PS microbead surface, \(\theta\), can be approximated by a standard normal distribution.19 Therefore, the distribution of \(\theta\) follows the expression:

\[
P(\theta \geq \theta_i) = 1 - \Phi\left(\frac{\theta_i - \bar{\theta}}{\sigma}\right)
\]

where \(\bar{\theta}\) is the mean of \(\theta\), \(\sigma\) is the standard deviation, and \(\Phi(x)\) is the standard normal distribution function. Combining eqs 1 and 2, one obtains:

\[
\frac{N_i}{N} \approx p_1 = 1 - \Phi\left(\frac{-\bar{\theta}}{\sigma}\right) = \int_{(\theta_i-\bar{\theta})/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-t^2/2} dt
\]

We find that \(\Phi(x)\) is approximately linearly related to \(x\) when \(|x| < 1\), with coefficient \(\alpha\), which is approximately \(\frac{1}{\sqrt{2\pi}}\) [page 178 in ref 18]:

\[
\frac{N_i}{N} \approx p_1 = \frac{1}{2} - \frac{\alpha}{\sigma} \theta_i + \frac{\alpha}{\sigma} \bar{\theta} = \frac{1}{2} - \frac{\alpha}{\sigma} \theta_i + \frac{\alpha}{\sigma} \bar{\theta}
\]

According to the Langmuir isotherm:

\[
\frac{\bar{\theta}}{\bar{\theta}} = \frac{C_i}{K_D + C_i}
\]

where \(K_D\) is the equilibrium dissociation constant and \(C_i\) is the peptide concentration in solution. The equilibrium between the octapeptide in solution (mobile state) and those adsorbed is described by the Langmuir adsorption isotherm. Therefore, the probability of fluorescence intensity can be expressed as

\[
p_1 \approx \frac{1}{2} - \frac{\alpha}{\sigma} \theta_i + \frac{\alpha}{\sigma} \frac{C_i}{K_D + C_i} = p_0 + \frac{nC_i}{K_D + C_i}
\]

\[
p_0 = \frac{1}{2} - \frac{\alpha}{\sigma} \theta_i
\]

\[
n = \frac{\alpha}{\sigma}
\]

In the present study, \(p_0 \approx 2\%\), and is related to the preselected threshold in the FCM measurements. \(n\) is a constant relating to the standard deviation of peptide coverage on the PS microbead surface.

Note that in the actual experiments, the PS microbeads inevitably have size distributions that could lead to fluctuations in the numbers of peptide binding sites on their surfaces; the
assumption that each microbead has identical numbers of binding sites with peptides is an idealized situation. However, we expect that, within experimental error, the relationship provides an accurate estimate of the interpeptide binding affinities.

The measured binding affinities enable us to determine the free energies of association for octapeptide pairs

\[
\Delta G_\text{AA8}(\text{AA8} \leftrightarrow \text{AA8}) = RT \ln K_D(\text{AA8} \leftrightarrow \text{AA8}) \tag{6}
\]

in which AA8 represents the bound octapeptides on the microbead surfaces (in the stationary states) and AA8 represents the fluorescent-tagged octapeptides (in the mobile phase), respectively. Note that since the homo-octapeptides lead to reduced ambiguity on interpeptide interactions due to heterogeneity in chemical structures of side chains, the interaction strengths between pairs of octapeptides can be correlated to the corresponding pairs of amino acids. Consequently, analyses of main-chain-related and side-chain-side-chain interactions may be performed by comparative

Figure 1. Schematic of the experimental configuration for obtaining interpeptide binding affinities using a microbead-based fluorescence method.

Figure 2. Binding affinity measurement of G8 ↔ G8 using microbead-based fluorescence. (A) Fluorescence distribution of microbeads modified with G8 from flow cytometry measurements. The red distribution corresponds to control microbeads (i.e., G8 concentration is zero), and the blue distribution corresponds to a G8 concentration of 5 \times 10^{-6} M. The fluorescence threshold is marked by the arrow. (B) The dependence of the probability of fluorescence intensity on G8 concentration. The solid curve is the fit based on eq 5 based on the Bayesian analysis described in the Experimental Section.
analyses of the free energies of association of octapeptides pairwise.

RESULTS AND DISCUSSION

The binding of octapeptides is measured using microbead-based flow cytometry (as shown schematically in Figure 1). The binding of the fluorescein isothiocyanate (FITC)-labeled octapeptides in the mobile state (referred to as (8) and the octapeptides in the stationary state (hereafter referred to as AA8) leads to fluorescence of the microbeads proportional to the equilibrium coverage. The binding affinities of octapeptide pairs were obtained from the microbead population above the fluorescence threshold as a function of the (8 concentration.

An example of the experimental results for G8 ↔ 8i s presented in Figure 2A. The gate was set at the arrow, where the population of the background signal above the gate value is 1.8%. The microbead population above the fluorescence threshold, p1, is 48.2% for the G8 ↔ 8 binding pair at a (8 concentration of 5 × 10⁻⁸ M. The equilibrium dissociation constant K_D of G8 ↔ 8 was obtained by fitting the fluorescence intensity distribution on 8 as a function of concentration using Bayesian analysis (Figure 2C). The values for the constants p₀ and n were obtained from the Bayesian analysis, as shown in Figure 2C. The constant p₀ corresponds to the microbead population above the fluorescence threshold in the control measurements in which no fluorescently labeled peptides were introduced into the solution. The constant n is related to the standard deviation of the peptide coverage on the PS microbead surface. The complete results of binding affinity measurements for 20 × 20 pairs of octapeptide consisting of the common amino acids (Figures S1−S400) and the tabulated binding affinities K_D for octapeptide pairs (Table S1) are presented in the Supporting Information.

This complete set of measurements of equilibrium dissociation constants (K_D) of the 20 × 20 pairs of octapeptides in the stationary (AA8) and mobile (AA8) states (Table S1) reveals the interaction strengths between all of the 20 amino acids. From the relationship in eq 6, the free energies of association for interpeptide interactions are obtained and are presented in Table 1, in which the values of the free energies of association ΔG of AA8 ↔ AA8 are categorized and color coded. The binding energies are categorized into four regimes, very strong (below −12 kcal/mol), relatively strong (−10 to −12 kcal/mol), relatively weak (−8 to −10 kcal/mol), and very weak (above −8 kcal/mol). The diagonal values in Table 1 represent the binding energies between pairs of identical octapeptides. The off-diagonal values correspond to the binding energies between dissimilar octapeptide pairs distinguished by the physicochemical properties such as size, shape, charge, and hydrophobicity of constituent amino acids. The magnitude of ΔG is between −7.7 ± 0.3 kcal/mol (for D8 ↔ T8) and −13.0 ± 0.1 kcal/mol (for L8 ↔ W8, I8 ↔ W8). We conclude from Table 1 that the interpeptide interactions are exclusively attractive.

With the obtained complete set of the binding energies of octapeptides, consisting of the common amino acids under identical conditions, we can elucidate the contributions of inter-amino-acid recognition and understand the sequence dependence of interprotein interactions at the level of individual amino acids. In order to understand the inter-amino-acid interactions from the measured free energies of association of homopeptides, it is important to differentiate the contributions from side-chain–side-chain, side-chain–main-chain, and main-
chain–main-chain interactions. We illustrate here that comparative analyses of the free energies of interpeptide association enable elucidation of the key contributions to the overall interactions. Note that the net charges of the basic and acidic amino acids are significantly dependent on pH, and associated electrostatic interactions are screened by the electrolyte in buffer solution in the present study. Therefore, the contributions to the inter-amino-acid interactions in this work should not be dominated by ionic interactions.

In order to determine the main-chain- and side-chain-related effects in binding interactions, we discuss below the contributions from side chains and main chains via comparative analyses of amino acids with structural similarities. An example of such a comparison can be seen in the effect of hydrophobicity in the alkyl side-chain series, such as the binding energies of aspartic acid (D) vs glutamic acid (E). These two amino acids have the same terminal carboxyl moiety, −COOH, with a difference of only one methylene, −CH=−, in the side chain. As a result, the binding energy \( \Delta G_8(E8 \leftrightarrow E8) \) is lower than \( \Delta G_8(D8 \leftrightarrow D8) \) by 0.6 kcal/mol (−9.7 kcal/mol vs −9.1 kcal/mol). This and related differences indicate that longer alkyl side chains and hydrophobicity can contribute appreciably to enhance binding affinities. The same effect can be observed in the relative binding energies of glutamine (Q) and asparagine (N) (with terminal amide groups). The binding energy of \( Q8 \leftrightarrow Q8 \) is again lower than that of \( N8 \leftrightarrow N8 \) by 0.5 kcal/mol (−11.0 kcal/mol vs −10.5 kcal/mol). The magnitudes of the binding energy differences due to the methylene moiety are in general agreement with those derived from mutational analyses such as the stability of the enzyme barnase (ca. 0.8 kcal/mol).20

In addition, significant stereoselective effects can be elucidated involving amino acids with branched side chains, i.e., isoleucine (I), leucine (L), and valine (V). From the molecular structures, I and L are allotropes with branched methyl groups at the \( \beta \) and \( \gamma \) carbon sites, respectively. The effects of steric repulsion can be seen in the binding energies of \( \Delta G_8(L8 \leftrightarrow L8) \) (−10.1 kcal/mol), which is lower than \( \Delta G_8(I8 \leftrightarrow I8) \) (−9.0 kcal/mol) by 1.1 kcal/mol. The stereoselectivity due to the branched side chains contributes to the selectivity inter-amino-acid interactions.

One can further compare side chains with the same alkyl lengths to determine the contributions from the terminal moieties to the binding energies. These effects are reflected in the variations of binding energies for peptides with one methyl unit in the side chains, phenylalanine (F) (F8 \( \leftrightarrow \) F8: −10.4 kcal/mol, tryptophan (W) (W8 \( \leftrightarrow \) W8: −12.0 kcal/mol), tyrosine (Y) (Y8 \( \leftrightarrow \) Y8: −11.1 kcal/mol), histidine (H) (H8 \( \leftrightarrow \) H8: −9.6 kcal/mol), aspartic acid (D) (D8 \( \leftrightarrow \) D8: −9.1 kcal/mol), and leucine (L) (L8 \( \leftrightarrow \) L8: −10.1 kcal/mol). With one methyl unit, the binding energies of F8, W8, Y8, and H8 are significantly lower than those of D8 and L8, suggesting that aromatic interactions between side chains are more effective in enhancing binding affinities. With two methylene units in the side chain moiety, the binding energy of methionine (M) (M8 \( \leftrightarrow \) M8: −11.6 ± 0.1 kcal/mol) is noticeably lower than that of glutamic acid (E) (E8 \( \leftrightarrow \) E8: −9.7 ± 0.3 kcal/mol) and glutamine (Q) (Q8 \( \leftrightarrow \) Q8: −11.0 ± 0.3 kcal/mol).

The above discussion qualitatively illustrates the contributions of side-chain-related interactions to the binding affinities of amino acids. We further demonstrate the feasibility of analyzing main-chain-related interactions, which are also critical in determining inter-amino-acid binding energies.

We consider the contributions from main chains by comparing the binding energies between glycine (G) with other amino acids. Since glycine is the simplest amino acid, with only one hydrogen atom in the side chain, only main-chain–main-chain interactions are involved in \( G8 \leftrightarrow G8 \). Therefore, the interactions of glycine with other amino acids can be used to illustrate the main-chain-side-chain interactions.

Figure 3 presents the distribution of binding energies between \( G8 \) and octapeptides of the 20 common amino acids. The magnitudes of the binding energies vary between −8.0 kcal/mol (\( G8 \leftrightarrow I8 \)) and −12.7 kcal/mol (\( G8 \leftrightarrow W8 \)) in the \( G8 \leftrightarrow A8/A8 \) and \( AA8 \leftrightarrow G8 \) series, which indicate that the main chains of \( G8 \) and \( G8 \) interact with the side chains of other octapeptides. The contributing interactions can be attributed to hydrogen bonding between the C=O and N–H groups of the main chains with the side chains of other octapeptides.21,22 Other interactions, including electrostatics and steric hindrance, can also contribute. The analogous comparisons of free energies of association for other AAs are provided in the Supporting Information (Figures S401–S419).

The exceptionally strong binding affinity of \( W8 \) with \( G8 \) can be attributed to the indole moiety of the tryptophan side chain that forms NH···π hydrogen bonds, while the indole NH-group can participate in hydrogen bonding as an acceptor.23 The quadrupole of the aromatic moiety of tryptophan can also contribute weak electrostatic interactions.23,24 In particular, \( W8 \) in the mobile phase interacts almost invariably with peptides of all 20 amino acids, as seen in Table 1. This observation is consistent with the pronounced stabilization effect of tryptophan in protein structures, which has been extensively reported.25

As seen in Figure 3, the binding energy of \( G8 \) with \( I8 \) (−8.0 kcal/mol) is significantly reduced from the baseline interactions between the main chains (\( G8 \leftrightarrow G8 \) : −9.4 kcal/mol). We attribute such repulsive effects to side-chain desolvation associated with shielding of the branched structure of isoleucine (I) blocking the access of water to C=O in the main chain, which could be energetically unfavorable.27 In contrast, the enhanced interaction can be seen between the side...
chains of I8 and the main chain of G8 (I8 ↔ G8: −10.7 kcal/mol), showing that branched side chains can enhance interaction strength, which may be conformation dependent.

Note that in Figure 3, most of the binding energies for \( \Delta G_\text{AA}(G8 ↔ A\,A8) \) and \( \Delta G_\text{AA}(A8 ↔ G8) \) are close (with differences of less than 1.1 kcal/mol) to that of \( \Delta G_\text{AA}(G8 ↔ G8) \) (−9.4 kcal/mol), except for \( \Delta G_\text{AA}(G8 ↔ W8) \), \( \Delta G_\text{AA}(G8 ↔ I8) \), \( \Delta G_\text{AA}(I8 ↔ G8) \), and \( \Delta G_\text{AA}(P8 ↔ G8) \), suggesting that the dominant inter-amino-acid interactions are due to the main chains. This observation is consistent with the statistical analyses of protein structures that reveal hydrogen-bonding interactions are common between the C=O and N–H groups of main chains.21,22

The extensive efforts focusing on the interactions between main chains of proteins and peptides have revealed a rich range of energetics that are sensitive to solvent quality and conformations. The hydrogen bonds between main chains have been at the center of this pursuit and determining the free energy of an individual amide hydrogen bond N–H–O=C has been attempted in a number of studies.28 The dependence of the hydrogen-bond strength on conformation and local environment leads to significant variations.4 It is unlikely that all amide moieties in G8 (or G8) participate in hydrogen bond formation in G8 ↔ G8. Rigorous quantification of the inter-amino-acid binding energies will require further experiments to clarify the contributions of van der Waals interactions and conformational effects of the whole peptide, which are likely dependent on peptide length as well as sequence.

The peptide conformational entropy may be different in the mobile and stationary states, leading to observable asymmetries in binding affinities. An interesting comparison is found in the asymmetry in the binding energies of AA8 ↔ G8 in the mobile state with other homopeptides in the stationary state. This effect is more pronounced in W8 ↔ G8, D8 ↔ G8, and C8 ↔ G8. The dramatic differences between G8 ↔ W8 (−12.7 kcal/mol) and W8 ↔ G8 (−9.6 kcal/mol), G8 ↔ D8 (−10.3 kcal/mol), and D8 ↔ G8 (−8.8 kcal/mol), as well as G8 ↔ C8 (−10.3 kcal/mol) and C8 ↔ G8 (−8.8 kcal/mol) are likely manifestations of conformational entropy effects for the stationary and mobile peptides.4 Side chain conformation is dependent on the state of immobilization, resulting in variations in association rates and binding affinities5 and the observed asymmetry in binding energies between the amino acid pairs.

The binding energies between G8 and peptides of other common amino acids in Figure 3 also reflect inter-amino-acid recognition, i.e., specificity, selectivity, and polymorphism. The specificity of glycine can be defined as its profile of interactions with the common amino acids in both stationary and mobile states. The selectivity of glycine, as well as other amino acids, can be defined as the most pronounced interactions in the specificity profiles, which provide the basis for studying the sequence dependence and the mutagenesis effects on peptide–peptide and protein–protein interactions. The values of binding energies can be close (within experimental error) for certain sets of peptide pairs, which is consistent with the polymorphisms of inter-amino-acid interactions. Such polymorphisms in binding free energies are indicative of the complementarity and synergy of amino acids with significant chemical and structural differences. We posit that the polymorphisms at the level of individual amino acids can be associated with the polymorphisms of protein structures and related biological functions. The above-discussed specificity, selectivity, and polymorphism of glycine can be generalized to all amino acids and constitute the core elements of inter-amino-acid recognition.

The above discussion of the G8 ↔ A/A8 series demonstrates the feasibility of identifying representative side-chain-related and main-chain-related interactions based on comparative analyses of interpeptide binding energies. Importantly, the above comparative analysis on the glycine series suggests that contributions from side chains can be recognized distinctly in the strong interaction regime (particularly at or below −12 kcal/mol binding energy, or 10⁻⁸ M in binding affinity) between homopeptides. The inter-amino-acid interactions in these strong-binding regimes can be attributed to side-chain-enhanced binding energies between octapeptide pairs with the magnitudes of ca. −3 kcal/mol (at or below −12 kcal/mol).

The above comparative analysis illustrates that inter-amino-acid recognition is based on understandable combinatorial interactions of side chains with differences in size, shape, hydrophobicity, acidity, and hydrogen-bonding potential. Hydrophobic interactions are naturally associated with nonpolar side chains, and are of critical importance for protein folding and protein–protein interactions. The electrostatic interactions between similarly or oppositely charged side chains are associated with acidic and basic amino acids. In addition, solvent molecules are important in stabilizing the interactions through competitive bonding and entropic contributions. Full-fledged analyses of side-chain–side-chain, side-chain–main-chain, and main-chain–main-chain interactions must take into account contributions including ionic, hydrophobic, aromatic, hydrogen bonding, sulfuric, and stereochemical interactions. The approach used here effectively circumvents and also provides key quantitative data for such analyses. It is apparent that multiple, rather than single, types of inter-amino-acid interactions are involved. We attribute the significant differences in the magnitudes of the free energies in the literature and in this work to disparities in local solvent environments, and cooperativity in protein structures. We anticipate that extended studies on peptides with heterogeneous sequences will provide useful routes for elucidating the effects of heterogeneity and cooperativity in the sequence dependence of protein structures.

On the basis of the above discussion of the strengths of inter-amino-acid interactions, four categories can be defined based on contributions from main-chain-related and side-chain-related effects:

1. **Main-Chain–Main-Chain Attraction (e.g., G ↔ G):** These amino acid pairs are characterized by interaction strengths comparable to that of glycine–glycine.

2. **Side-Chain–Main-Chain Reduced Attraction (e.g., G ↔ I):** These amino acid pairs are characterized by identifiable reductions in interaction strength compared to that of glycine–glycine. The reduced inter-amino-acid interactions may originate from disfavored side-chain–main-chain interactions (such as shielding effects due to side-chain desolvation).

3. **Side-Chain–Main-Chain Enhanced Attraction (e.g., G ↔ W):** These amino acid pairs are associated with pronounced side-chain–main-chain interactions. Side-chain–side-chain interactions may also contribute.

4. **Side-Chain–Side-Chain Enhanced Attraction (e.g., R ↔ K, R ↔ D):** These amino acid pairs reflect significant contributions from side-chain–side-chain interactions in the strong interaction regime (equivalent to the octapeptide pairs with at or below −12 kcal/mol in...
binding energy, or $10^{-9}$ M binding affinity). The magnitude of side chain interactions in this regime suggest that side-chain–side-chain interactions may be comparable to an individual amide hydrogen bond $\text{N}^{-}\text{H} \cdots \text{O} = \text{C}$.

On the basis of the experimental results, the majority of side-chain-related inter-amino-acid interactions are found to be in side-chain–main-chain enhanced attraction and side-chain–main-chain reduced attraction due to comparable contributions from side-chain–main-chain and main-chain–main-chain interactions. The distribution of interaction strengths is shown in Figure 4 and the majority of the values for $\Delta G_f^{\text{AA8}}(\text{AA8} \leftrightarrow \text{AA8})$ range between $-12$ and $-8$ kcal/mol. Pronounced selectivity, as well as specificity and polymorphism, can be observed from this work for the inter-amino-acid interactions of the selected amino acid pairs originates from their chemical structures, conformations, etc. Selectivity can be demonstrated in preferential interactions between amino acids, particularly in the strong interaction regimes, such as $G \leftrightarrow W$, etc. Specificity can be seen in the interaction strengths shown in Figure 3.

In Table 1, different amino acid pairs can have similar binding affinities, such as $G \leftrightarrow H$ and $G \leftrightarrow R$, $G \leftrightarrow D$, $G \leftrightarrow C$, $G \leftrightarrow N$, and $G \leftrightarrow S$ in Figure 3 and Table 1, as well as in the histogram distribution in Figure 4. This observed multiplicity in interaction strengths among hetero-amino-acid pairs with different chemical structures can be the basis for peptide interaction polymorphisms, particularly for heterogeneous sequences of peptides. Furthermore, the binding affinity measurements for homo-amino-acid pairs show that in addition to the mean value of equilibrium, the standard deviation specifies the characteristic distribution of the measurement results. The multiplicity in binding affinity for homo-amino-acid pairs is a reflection of contributions from the intrinsic population distributions of thermodynamic states from the conformation, configuration, etc., of the amino acids in the peptide pairs and the polymorphisms in inter-amino-acid interactions.

These results also enable compositional analyses of free energy with contributions from main chains and side chains. The physicochemical determinants of the binding affinities of peptide–peptide complexes include length, composition, sequence, environmental conditions, etc. that are intrinsically associated with contributions from side-chain interactions, structural and conformational effects in peptides. The measured binding affinities of homo-octapeptides here should facilitate the effort to unravel these determinants in future studies, as well as to pursue the ultimate goal of predicting peptide–peptide binding affinities.

A biologically relevant peptide length effect may be found in polyglutamate (polyQ) and polyasparagine (polyN) that have been associated with a number of diseases, including neurodegenerative diseases. The reported studies of diseases related to polyQ and polyN have shown significant length dependences. For example, a threshold of 35–60 glutamines was identified for disease symptoms. PolyQ peptides with lengths below 35 have no disease manifestations. The length effects on conformation and aggregation kinetics have also been explored. In addition, the presence of flanking sequences and cooperative effects in aggregation of polyQ peptides have been identified in a number of studies.

Note that the typical lengths of reported polyQ and polyN peptides are much longer than the octapeptides in our work. Definitive analyses of the binding affinities of polyQ and polyN require further insight into the length-dependence effects on binding affinities, which will be pursued in future studies using the methodology described here and complementary approaches. Furthermore, the composition and sequence effects on the binding affinities can be used to help elucidate the molecular mechanisms relating to the flanking sequences and cooperative effects on polyQ aggregation propensity.

The methodology of inter-amino-acid recognition described here can be adapted for studying interactions between the common naturally occurring amino acids and other categories of amino acids, such as non-natural amino acids, as well as glycans and glycopolymers. One could further explore the molecular-level effects of post-translational modification (e.g., phosphorylation and glycosylation) on interprotein interactions, which have broad biological and pharmaceutical importance. The obtained inter-amino-acid interaction strengths would also contribute to the pursuit of novel peptidomimetic polymers in which main-chain-related interactions can be neglected and peptidomimetic libraries are developed for modulating protein–protein interactions.

Preliminary efforts in the rational design of peptides based on these data and the principles of inter-amino-acid recognition have been explored targeting metastasis-associated protein for stroma-secreted chemokine (CXCL12) and its receptor 4 (CXCR4) in multiple acute myelocytic leukemia (AML) cells, epithelial cell adhesion molecule (EpCAM) and human epidermal growth factor receptor-2 (HER2) for capturing circulating tumor (CTC) cells in blood for breast cancer diagnosis and prognosis, respectively. Optimization and maximization of inter-amino-acid interactions are vital for further insight into the complexity of diverse protein structures and functions in biological systems and in deliberately designed, tailored, and predictable interaction strengths in synthetic peptides.

These principles will also be useful in systematic investigations of side-chain-related interactions, such as solvent dependence due to water-mediated hydrogen bonds between hydrophilic and charged side chains previously identified as important in protein stability. Ultimately, understanding interactions will also require systematic analyses of cooperativity in heterogeneous peptide sequences. Note that the presented inter-amino-acid interactions derived from unaggregated peptides could also be helpful in analyzing amyloid-forming propensities of the 20 common amino acids.
The importance of inter-amino-acid recognition can be found in the protein−protein interaction literature. On the basis of the selectivity of the amino acid interactions observed, we can assess previously reported conserved amino acid sequences that dominate protein−protein interactions. From Table 1, distinctively high propensities for the strongest inter-amino-acid interactions can be found for tryptophan (W), proline (P), arginine (R), tyrosine (Y), and methionine (M). This observation is qualitatively consistent with the reports of conserved amino acids in the high binding affinity regimes (hot spots for protein−protein interfaces) of tryptophan (W), arginine (R), and tyrosine (Y).

The inter-amino-acid interactions are inherently relevant to the interactions at different levels of protein complexity. Protein−protein, ligand−receptor interactions result from interpeptide interactions based on inter-amino-acid interactions. The binding energies between amino acids and the contributions from side-chain−side-chain interactions and side-chain−main-chain interactions can also guide the design of ligands or peptides for targeted proteins. The significant differences in inter-amino-acid interactions provide a basis for the sequence dependence of interpeptide and interprotein interactions at the level of individual amino acids. On the other hand, the apparent multiplicities of interaction strengths between amino acid combinations indicate that polyamorphism can conserve or tune protein−protein interaction strengths.

**CONCLUSIONS AND PROSPECTS**

The principles of inter-amino-acid recognition, as presented in this work, can be summarized as follows:

1. Inter-amino-acid interactions are exclusively attractive with synergistic contributions from both main chains and side chains.
2. Side-chain-related interactions can be identified and categorized according to the measured interaction strengths.
3. The specificity of inter-amino-acid recognition can be defined by binding energy profiles (or equivalent binding affinity profiles) reflecting synergistic contributions from side chains and main chains of the 20 common amino acids.
4. The selectivity of side-chain-related interactions is pronounced for the strongest inter-amino-acid interactions and provides the basis for inter-amino-acid recognition.
5. The multiplicity of inter-amino-acid recognition provides the basis for polyamorphism in interpeptide and inter-amino-acid interactions.
6. The specificity, selectivity, multiplicity, and polymorphism of inter-amino-acid recognition provide a basis for tuning/tailoring inter-amino-acid interactions in biological systems.

Among many long-standing challenges in studying protein structures and properties, the sequence dependence of structures and binding interactions between proteins are universal and pivotal in biological systems. We anticipate that understanding the principles of inter-amino-acid interactions will advance our ability to address these profound challenges and understanding of the roles of contacting amino acids in antibody−antigen, ligand−receptor, and other protein−protein and protein-peptide interactions.

**EXPERIMENTAL METHODS AND MATERIALS**

The detailed descriptions of the experiment protocols for the full set of binding affinity measurements are provided in the Supporting Information (including Tables S3−S5). The following brief descriptions are provided here for convenience. No unexpected or unusually high safety hazards were encountered.

**Materials.** All the peptides used in our experiments were purchased from Shanghai Science Peptide Biological Technology Co., Ltd. They are all homo-octapeptides. The full set of analytical data sheets including high-performance liquid chromatography (HPLC) and mass spectrometry (MS) results are provided for biotinylated peptides (Figures S420−S439) and the peptides with FITC and a linker unit of 6-aminoacapric acid (Acp) (Figures S440−S459).

Streptavidin (SA)-coated PS microbeads were purchased from Bangs Laboratories, Inc. with a mean diameter of 4.95 μm. The binding capacity is 0.057 μg of biotin-FITC per mg microspheres. The number of particles is 1.432 × 10^10 beads per gram of dry microbeads. The biotinylated peptides bound onto SA-coated PS microbeads are defined as the stationary-state molecules represented in the form of AAS, and the free peptides with FITC with a linker unit of Acp in solution are defined as the mobile-state molecules represented in the form of A/As.

Phosphate buffer solution was purchased from GE Healthcare Life Sciences Hyclone Laboratories. Dimethyl sulfoxide (DMSO) was purchased from Sigma-Aldrich Co., Ltd. All reagents were used without further purification.

**Preparation of Peptide Solutions.** One milligram of octapeptide (AA8 or A/As) was dispersed in 50 μL of DMSO and then sonicated for 5 s with several times for complete dissolution. The solution was saved in the dark for 1 h. 950 μL of PBS buffer was added to AA8 solution to obtain the peptide concentrations of 1 mg/mL. Calculated volumes of PBS buffer were added to A/As solution to obtain the stock solutions at peptide concentrations of 1 mM.

We have examined the fluorescence-labeled peptides deposited on mica surfaces by atomic force microscopy (AFM). The results suggest little or no oligomeric effects for hydrophilic, basic, and acidic amino acids in the range of low concentrations (typically 10^−10 to 10^−6 M or below). As an example of representative hydrophobic peptides, tryptophan homo-octapeptide was examined in the AFM observations. The topographic images indicate that the deposited tryptophan peptides are overwhelmingly monomeric, while a small fraction of peptide aggregates might be observed at relatively high concentrations (Figure S460).

The impact of the observed oligomeric state of hydrophobic peptides on the outcome of binding affinity measurements is unlikely to be dominant due to the following considerations. First, the magnitude of the measured binding affinity is mainly determined by the slope of the isotherm at low concentrations. The observed populations of oligomeric aggregates for hydrophobic peptides by AFM are not pronounced in this range of concentrations (typically 10^−6 to 10^−10 M or below). Second, the fluorescence intensity distributions in flow cytometry results are consistent with the dominant contributions being from monomeric peptides, since oligomeric peptides would likely lead to abnormal fluorescence intensities, which are not observed in the measurements.

**Binding of A/As with AA8-Modified Microbeads.** Streptavidin-coated PS microbeads suspensions were stored at...
4 °C and were vortexed for approximately 30 s prior to use to ensure dispersity. The 115 μL microbead suspensions were washed with PBS buffer followed by centrifugation at 10 krpm for 3 min, and then the supernatant was discarded. After three washes and centrifugation cycles, the centrifuged pellet was resuspended in 500 μL PBS with addition of 25 μL of octopeptide AA8 solution at a concentration of 1 mg/mL. The solutions were incubated at room temperature for 3 h, and then washed and centrifuged three times. The centrifuged pellet was resuspended in 5.0 mL of PBS and dispatched into centrifuge tubes, numbered as 0—9. Then, 450 μL of PBS was added into sample number 0 tube to bring the final volume of 500 μL. Next, corresponding volumes of PBS and a series of AA8 solutions were added into tube numbers 1—9 to obtain 500 μL microbead suspensions. In detail, after sequential dilution with PBS buffer, a series of peptide solutions was obtained with final concentrations of AA8 at 0.5 nM, 1.5 nM, 5 nM, 0.015 μM, 0.05 μM, 0.15 μM, 0.5 μM, 1.5 μM, and 5 μM, except for W8, which was prepared with final concentrations of 0.05 nM, 0.1 nM, 0.15 nM, 0.5 nM, 1.5 nM, 5 nM, 0.015 μM, 0.05 μM, and 0.15 μM. All of the samples were repeated 10 times. The samples were incubated overnight at 37 °C in a dark table concentrator at a speed of 10 rpm prior to use in FCM experiments.

Note that the interactions between mobile peptide and streptavidin could potentially contribute to the measured binding affinities. However, we expect such contributions should be insignificant due to the following considerations, based on the adsorption structure of the peptides. As to the structural effects of the immobilized peptides, it has been established for surface-immobilized linear molecules, such as long-chain alkanethiolates and oligonucleotides, that the molecular backbone is typically tilted from the surface normal direction with a cant angle, accompanied by twists of the main molecular backbone is typically tilted from the surface normal established for surface-immobilized linear molecules, such as structural elements based on the adsorption structure of the peptides. As to the speed of 10 rpm prior to use in FCM experiments. Incubation overnight at 37 °C.

A flow cytometry. An Accuri C6 flow cytometer (BD Biosciences) was used at an emission wavelength of 488 nm, a detection wavelength of 530 ± 15 nm, and a flow rate of 14 μL/min. The gate was set when the FCM measurements are performed without FITC-octopeptides in the mobile phase (sample number 0, referred as control). The fluorescence threshold Iₘ was set where the background signal above this fluorescent intensity gate was below pₛ = 2%. At the same fluorescence threshold, sample numbers 1—9 with FITC-labeled octopeptides AA8 in the mobile phase (FITC-AA8) were measured, and the corresponding detection values, i.e., the percentages of PS microbeads with fluorescence above the threshold Iₛ, pₛ were recorded. Each set of data consists of nine concentrations and 5000 events (each event corresponds to one PS microbead counted in flow cytometry) for each concentration. All experiments were repeated 10 times, thus one Kₒ value of an AA8—AA8 pair was obtained from as many as ca. 450 000 events. Such sample numbers are important in testing and validating the statistical methods for the microbead-based fluorescence methodology and data analyses used.

Data Analysis. Data analyses were performed by fitting the data to the probability of fluorescence intensity in the above expression. We conduct Bayesian analysis for the determination of the parameters in this expression. A nearly noninformative Gamma(2,0.01) distribution (with variance 20 000) is used as the prior distributions for all parameters. The posterior probability distribution is sampled using the Markov Chain Monte Carlo method.51 We draw 10 000 Monte Carlo samples for inference after 5000 burn-in samples. The lower and upper credible bounds are estimated using the 5% and 95% quantile points of the Monte Carlo samples.

Validation of the Microbead-Based Binding Affinity Measurements. Orthogonal Binding Assays. (G8 ↔ Q8, G8 ↔ W8) are presented in the Supporting Information of the manuscript. The detailed experimental protocol for performing the orthogonal tests is provided in the Supporting Information (including Tables S6—S11). The factors and levels for the orthogonal test are listed in Table S12. These assays are representative for side chain moieties with different structures and hydrophobicity/hydrophilicity.

The orthogonal test design L₉(2⁷) with interaction effects was performed to include four factors and two levels, which are relevant to the measurements of peptide—peptide binding affinities. The four factors are peptide incubation time on the microbead surface (factor C), preparation time of the peptide solutions (factor B), and incubation time and vibration speed for peptide—peptide binding interactions (factor A and factor D, respectively). In order to explore the possible impact of preparation conditions of mobile and stationary peptides on the binding affinity measurements, the possible interactions between factors A and B, and A and C are also included in the orthogonal test.

The intuitive analysis of the orthogonal experiments is provided in Tables S13 and S14 for binding assays G8 ↔ Q8 and G8 ↔ W8. The measured binding affinities with different experimental factors showed reasonable consistency in each assay with different binding strengths. The range of the orthogonal test is provided in the last rows in Tables S13 and S14 reflecting the impact of level variations of factors on the measured binding affinities for different peptide—peptide interactions.
interaction strengths, structures, and hydrophobicity/hydrophilicity.

According to the $R$ values in Tables S13 and S14, the significance of the factors on the peptide–peptide binding affinity measurements decreases in the following order:

- $G8 \leftrightarrow W8: A \times B > A \times C > D > C > B > A$
- $G8 \leftrightarrow Q8: C > A \times B > B > D > A > A \times C$

The above results indicate that affinity measurements can be optimized using orthogonal test design. Analyses of the above results also indicate that the experimental factors may have peptide-dependent effects on the binding affinity. Optimized experimental conditions can be selected for individual peptide–peptide pairs based on further orthogonal analysis. The assessment of the feasibility of developing rigorously optimized general experimental conditions for the full set of binding affinity measurements will be carried out in the near future.

Comparisons with Reported Independent Measurements.
The validation of the methodology described in this work was performed by comparing the measured binding affinities with the previously reported independent measurements (surface plasmon resonance, SPR; isothermal titration calorimetry, ITC; circular dichroism, CD) for protein A-IgG, streptavidin (SA)-biotin, a series of E. coli–K. coli coils, ranging from medium to strong interaction strength (for $K_D$ values between $10^{-7}$ and $10^{-10}$ M) (Table S15). Agreement can also been seen for the very strong binding interactions between SA and biotin ($K_D$ values around $10^{-16}$ M in Table S15). The difference in the quantitative values can be assigned to the variations in experimental assays. The qualitative consistency in the measured interaction strength can be identified from these independent results. This range of interaction strength fully covers the measurement regimes in our work. We note that different experimental approaches have been taken to measure medium to strong interaction strengths, as listed in Table S15 for surface SPR and ITC, etc., and are important to test the observed peptide–peptide interactions. Likewise, these measurements may be helpful in testing other experimental approaches to measuring peptide–peptide interactions.

Additional comparisons with independent microbead-based methods, such as mean fluorescence intensity (MFI), were performed. Mean fluorescence intensity is generally used as a binding reporter in microbead-based and cell-surface-based binding assays. The binding affinities determined by MFI for the data in the orthogonal tests are provided together with the fluorescence intensity probability method described here.

The comparison of affinity results for binding assays $G8 \leftrightarrow Q8$ and $G8 \leftrightarrow W8$ by using mean fluorescence intensity vs. fluorescent peptide concentration, in comparison with those from the fluorescence intensity probability formalism here, are provided in Figure S461 and Table S16. Specifically, the mean fluorescence intensity method appears to lead to affinities lower than or close to those of the fluorescence intensity probability method.

Note that for exact Gaussian-type fluorescence distributions, one expects that the use of median or mean fluorescence intensity should lead to similar results by using the formalism as rigorously derived in this work, which takes into account the distribution of peptide coverage on the PS microbead surface manifested by the fluorescence above the threshold intensity.

The observed differences in the binding affinities determined by fluorescence intensity probability method in this work and the median or mean fluorescence intensity methods may be attributed to the complementary fluorescence contributions from either below or above the threshold value defined here, particularly in the range of low peptide concentrations. Note that the magnitude of the measured binding affinity is predominantly determined by the slope of the isotherm at low concentrations. As the formalism presented in this work mainly takes into account the fluorescence intensity above the threshold, the outcome of the measured binding affinity has contributions from the peptide coverage on the PS microbead surface both at low and high concentrations.

In contrast, the median and mean fluorescence intensities at low peptide concentrations are predominantly distributed below the above-specified threshold intensity, which is within the range of detection limit of the cytometry detector. As a result, both median and mean fluorescence intensities measured at low peptide concentrations are more susceptible to intensity fluctuations. Such contributions from intensity fluctuations near the detection limit of the cytometry detector in the median or mean fluorescence intensity methods could lead to differences in the extracted affinity results in comparison to that of the fluorescence intensity probability method used here.

Preliminary Test of Length-Dependence in Interpeptide Interactions. There are notable increases from $10^{-7}$ to $10^{-11}$ M in interpeptide interactions ($K_D$) in Table S15 as the length of peptide in the stationary phase increases ($K_D$, 23 AA; $K_D$, 30 AA; $K_D$, 37 AA) while the peptide in the mobile phase remains the same ($E_8$, 30 AA). Our measurements and the reported values consistently illustrate the length dependence of binding interactions in the peptide pairs of E. coli and K. coil.

The effect of chain length in interaction strength is reproduced in the homopeptide pairs as illustrated in Table S17. We observe increasing binding affinity with increasing peptide length in the mobile phase interacting with different octapeptides in the stationary phase. For higher binding affinities, i.e., longer peptide chain length, the signal-to-noise ratio improved. Note that there is a technical limitation due to the synthetic difficulty of obtaining the full set of homopeptides beyond eight amino acids. The above observations are the rationale for our experimental design of utilizing homoctapeptides for the binding affinity measurements for the primary and complete set of experiments in this work.

Binding Affinities for Copeptides Containing Two Kinds of Amino Acids. To test the robustness of the experimental design, we measured copeptides containing two kinds of amino acids (glycine and tryptophan) in the form of single mutations in the sequences. As can be seen from the binding measurements of homologous octapeptides in Table S2, the affinity of $W8 \leftrightarrow W8$ ($5.6 \pm 1.5 \times 10^{-10}$ M) is much stronger than that of $G8 \leftrightarrow G8$ ($1.5 \pm 0.6 \times 10^{-7}$ M). Therefore, tryptophan and glycine can be considered as representative strongly and weakly interacting amino acids, respectively. The measured binding affinities of these copeptides in the mobile phase interacting with the octapeptide (G8) in the stationary phase revealed the determinant effect of strongly interacting amino acids in a heterogeneous peptide sequence.

The above results illustrate that it is feasible to observe the qualitative trend of binding affinities for binary peptides containing strong and weak interaction amino acids. Note that predicting binding affinities of copeptides essentially requires detailed insight of the sequence and length effects of interactions between heteropeptides in general. The capability of rigorous prediction of interaction strength affinity of different hetero-
peptides could be achieved by examining the effects in combinatory mutations based on the homopeptides. This is an extremely important effort that should be implemented in future efforts.

**ASSOCIATED CONTENT**

◆ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.8b00723.

Tdk binding affinity measurements and equilibrium dissociation constants of octapeptide pairs of the 20 common amino acids (PDF)

**AUTHOR INFORMATION**

Corresponding Authors

*(C.W.) E-mail: wangle@nanoctr.cn.
*(Y.Y.) E-mail: yangyl@nanoctr.cn.
*(P.S.W.) E-mail: psw@cnst.ucla.edu.

ORCID

Fuyang Qu: 0000-0002-4641-0289
Paul S. Weiss: 0000-0001-5527-6248
Yanlian Yang: 0000-0002-5607-5509

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The authors thank Prof. Anne Andrews, Dr. Linling Bai, Prof. David Eisenberg, Prof. Qiaojun Fang, Prof. Liming Xie, Dr. Lingling Geng, Prof. Ken Houk, Prof. David Tirrell, Dr. Chenxuan Wang, and Dr. Meng Xu for helpful discussions. The authors are also grateful to Prof. Xingjie Liang, Prof. Wei Liang, Dr. Juan Liu, and Dr. Ping Li for their support in providing technical assistance with the FCM facilities. We thank the support from National Natural Science Foundation of China (Grant Nos. 21673055, 21721002, and 21773042), and Key the support from National Natural Science Foundation of China providing technical assistance with the FCM facilities. We thank the support from National Natural Science Foundation of China (Grant Nos. 21673055, 21721002, and 21773042), and Key Research Program of Frontier Sciences, CAS, Grant No. QYZDJ-SSW-SLH048, and The National Key Research and Development Program of China (2016YFF0203803).

**REFERENCES**

(1) Anfinsen, C. B. Principles That Govern Folding of Protein Chains. Science 1973, 181, 223–230.
(2) Anfinsen, C. B.; Scheraga, H. A. Experimental and Theoretical Aspects of Protein Folding. Adv. Protein Chem. 1975, 29, 205–300.
(3) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. Adv. Protein Chem. 1959, 14, 1–63.
(4) Dill, K. A. Dominant Forces in Protein Folding. Biochemistry 1990, 29, 7133–7155.
(5) Kastritis, P. L.; Bonvin, A. M. J. J. On the Binding Affinity of Macromolecular Interactions: Daring to Ask Why Proteins Interact. J. R. Soc., Interface 2013, 10, 20120835.
(6) Blazer, L. L.; Roman, D. L.; Muxlow, M. R.; Neubig, R. R. Use of Flow Cytometric Methods to Quantify Protein–Protein Interactions. Curr. Prot. Cytometry 2010, 51, 13.11.1–13.11.15.
(7) Van Antwerp, J. J.; Wittrup, K. D. Fine Affinity Discrimination by Yeast Surface Display and Flow Cytometry. Biotechnol. Prog. 2000, 16, 31–37.
(8) Gai, S. A.; Wittrup, K. D. Yeast Surface Display for Protein Engineering and Characterization. Curr. Opin. Struct. Biol. 2007, 17, 467–473.
(9) Simons, P. C.; Young, S. M.; Carter, M. B.; Waller, A.; Zhai, D. Y.; Reed, J. C.; Edwards, B. S.; Sklar, L. A. Simultaneous in Vitro Molecular Screening of Protein-Peptide Interactions by Flow Cytometry Using Six Bcl-2 Family Proteins and Examples. Nat. Protoc. 2011, 6, 943–952.
(10) Forment, J. V.; Jackson, S. P. A Flow-Cytometry-Based Method to Simplify the Analysis and Quantification of Protein Association to Chromatin in Mammalian Cells. Nat. Protoc. 2015, 10, 1297–1307.
(11) Mullen, D. G.; Fang, M.; Desai, A.; Baker, J. R., Jr.; Orr, B. G.; Holl, M. M. B. A Quantitative Assessment of Nanoparticle-Ligand Distributions: Implications for Targeted Drug and Imaging Delivery in Dendrimer Conjugates. ACS Nano 2010, 4, 657–670.
(12) Li, M. H.; Choi, S. K.; Leroueil, P. R.; Baker, J. R., Jr. Evaluating Binding Avidities of Populations of Heterogeneous Multivalent Ligand-Functionalized Nanoparticles. ACS Nano 2014, 8, 5600–5609.
(13) Epand, R. M.; Scheraga, H. A. The Influence of Long-Range Interactions on the Structure of Myoglobin. Biochemistry 1968, 7, 2864–2872.
(14) Taniuchi, H.; Anfinsen, C. B. An Experimental Approach to the Study of the Folding of Staphylococcal Nuclease. J. Biol. Chem. 1969, 244, 3864–3875.
(15) Dyson, H. J.; Rance, M.; Houghten, R. A.; Wright, P. E.; Lerner, R. A. Folding of Immunogenic Peptide Fragments of Proteins in Water Solution: II. The Nascent Helix. J. Mol. Biol. 1988, 201, 201–217.
(16) Wang, C.; Du, H. W.; Yang, Y. L. A Method to Investigate the Interactions between Biological Molecules and Its Applications, China Patent No. ZL201410232021.6 (2016).
(17) Pittman, J. Probability; Springer-Verlag: New York, 1993.
(18) Feller, W. An Introduction to Probability Theory and Its Applications, 3rd ed.; John Wiley & Sons: New York, 1968; Vol. 1.
(19) The magnitude of $N$ is defined in probability theory of the normal approximation to the binomial distribution (pp 98–104 in ref 17). By defining the standard deviation $\sigma = \sqrt{N}\theta(1-\theta)$, a satisfactory approximation to a binomial distribution can be achieved for $\sigma \geq 3$. In addition, for experimental conditions that meet the requirement of $\sigma \geq 3$, it is appropriate to approximate $\theta$ by a standard normal distribution rather than a Poisson distribution (page 104 in ref 17). In our study, with $\theta$ above 0.02, the estimated value of experimental number should be 300 or higher in order to satisfy $\sigma \geq 3$. Specifically, $N \geq 5000$ will lead to $\sigma \approx 10$, which meets the criteria and justifies the approximation by a standard normal distribution here.
(20) Serrano, L., Jr.; Kellis, J. T.; Cann, P.; Matouschek, A.; Fersht, A. R. The Folding of an Enzyme: II Substructure of Barnase and the Contribution of Different Interactions to Protein Stability. J. Mol. Biol. 1992, 224, 783–804.
(21) Mirsky, A. E.; Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. Proc. Natl. Acad. Sci. U. S. A. 1936, 22, 439–447.
(22) Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. Prog. Biophys. Mol. Biol. 1984, 44, 97–179.
(23) Sengupta, A.; Mahalakshmi, R.; Shamala, N.; Balam, P. Aromatic Interactions in Tryptophan-Containing Peptides: Crystal Structures of Model Tryptophan Peptides and Phenylalanine Analogues. J. Pept. Res. 2005, 65, 113–129.
(24) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interactions: A Mechanism of Protein Structure Stabilization. Science 1985, 229, 23–28.
(25) Burley, S. K.; Petsko, G. A. Amino-Aromatic Interactions in Proteins. FEBS Lett. 1986, 203, 139–143.
(26) Samanta, U.; Pal, D.; Chakrabarti, P. Environment of Tryptophan Side Chains in Proteins. Proteins: Struct., Function, Genet. 2000, 38, 288–300.
(27) Luo, P.; Baldwin, R. L. Interaction between Water and Polar Groups of the Helix Backbone: An Important Determinant of Helix Propensities. Proc. Natl. Acad. Sci. U. S. A. 1999, 96, 4930–4935.
(28) Bolen, D. W.; Rose, G. D. Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding. Annu. Rev. Biochem. 2008, 77, 339–362.
(29) Penney, J. B.; Vonsattel, J. P.; MacDonald, M. E.; Gusella, J. F.; Myers, R. H. CAG Repeat Number Governs the Development Rate of Huntington’s Disease. Ann. Neurol. 1997, 41, 689–692.
(30) Snell, R. G.; Macmillan, J. C.; Cheadle, J. P.; Fenton, I.; Lazarou, L. P.; Davies, P.; Macdonald, M. E.; Gusella, J. F.; Harper, P. S.; Shaw, D. J. Relationship between Trinucleotide Repeat Expansion and Phenotypic Variation in Huntington’s Disease. *Nat. Genet.* 1993, 4, 393–397.

(31) Scherzinger, E.; Lurz, R.; Turmaine, M.; Mangiarini, L.; Hollenbach, B.; Hasenbank, R.; Bates, G. P.; Davies, S. W.; Lehachi, H.; Wanker, E. E. Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-Like Protein Aggregates in Vitro and in Vivo. *Cell* 1997, 90, 549–558.

(32) Lu, X. M.; Murphy, R. M. Asparagine Repeat Peptides: Aggregation Kinetics and Comparison with Glutamine Repeats. *Biochemistry* 2015, 54, 4784–4794.

(33) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; Byeon, I. L.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; Gronenborn, A. M.; Wetzel, R. Polyglutamine Disruption of the Huntingtin Exon 1 N Terminus Triggers a Complex Aggregation Mechanism. *Nat. Struct. Mol. Biol.* 2009, 16, 380–389.

(34) Tam, S.; Spiess, C.; Auyeung, W.; Joachimiak, L.; Chen, B.; Poirier, M. A.; Frydman, J. The Chaperonin TRiC Blocks a Huntingtin Sequence Element That Promotes the Conformational Switch to Aggregation. *Nat. Struct. Mol. Biol.* 2009, 16, 1279–1286.

(35) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. Glutamine Repeats as Polar Zippers: Their Possible Role in Inherited Neurodegenerative Diseases. *Proc. Natl. Acad. Sci. U. S. A.* 1994, 91, 5535–5538.

(36) van Hest, J. C. M.; Tirrell, D. A. Protein-Based Materials, Toward a New Level of Structural Control. *Chem. Commun.* 2001, 1897–1904.

(37) Sievers, S. A.; Karanicolas, J.; Chang, H. W.; Zhao, A.; Jiang, L.; Zirafi, O.; Stevens, J. T.; Münch, J.; Baker, D.; Eisenberg, D. Structure-Based Design of Non-Natural Amino-Acid Inhibitors of Amyloid Fibril Formation. *Nature* 2011, 475, 96–100.

(38) Broecker, F.; Areta, J.; Yang, Y.; Hanske, J.; Guo, X. Q.; Reinhardt, A.; Wahlinb, A.; Rademaker, C.; Anish, C.; Seeberger, P. H. Epitope Recognition of Antibodies against a Yersinia pestis Lipopolysaccharide Trisaccharide Component. *ACS Chem. Biol.* 2014, 9, 867–873.

(39) Rabuka, D.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. Noncovalent Cell Surface Engineering: Incorporation of Bioactive Synthetic Glycopolymers into Cellular Membranes. *J. Am. Chem. Soc.* 2008, 130, 5947–5953.

(40) Kiesling, L. L.; Grim, J. C. Glycopolymer Probes of Signal Transduction. *Chem. Soc. Rev.* 2013, 42, 4476–4491.

(41) Sun, J.; Zuckermann, R. N. Peptoid Polymers: A Highly Designable Bioinspired Material. *ACS Nano* 2013, 7, 4715–4732.

(42) Whitby, L. R.; Boger, D. L. Comprehensive Peptidomimetic Libraries T-Protein Interactions. *Acc. Chem. Res.* 2012, 45, 1698–1709.

(43) Li, X. J.; Guo, H.; Yang, Y. L.; Meng, J.; Liu, J.; Wang, C.; Xu, H. Y. A Designed Peptide Targeting CXCR4 Displays Anti-Acute Myelocytic Leukemia Activity in Vitro and in Vivo. *Sci. Rep.* 2015, 4, 4610.

(44) Wang, C.; Duan, H. Y.; Yang, Y. L.; Xie, H. Y.; Xu, H. Y.; Li, X. J. A Peptide for Metastasis and Its Applications. China Patent No. CN201610907839.2 (filed 2016).

(45) Bai, L. L.; Shao, B.; Li, H. P.; Wang, C.; Yang, Y. L.; Wang, C. Peptide-Based Isolation of Circulating Tumor Cells by Magnetic Nanoparticles. *J. Mater. Chem. B* 2014, 2, 4080–4088.

(46) Peng, J. X.; Zhao, Q.; Zheng, W. S.; Li, W. Z.; Li, P.; Zhu, L.; Liu, X. R.; Shao, B.; Li, H. P.; Wang, C.; Yang, Y. L. Peptide-Functionalized Nanomaterials for the Efficient Isolation of HER2-Positive Circulating Tumor Cells. *ACS Appl. Mater. Interfaces* 2017, 9, 18423–18428.

(47) Smith, C. K.; Regan, L. Guidelines for Protein Design: The Energetics of β-Sheet Side Chain Interactions. *Science* 1995, 270, 980–982.

(48) Chothia, C.; Janin, J. Principles of Protein-Protein Recognition. *Nature* 1975, 256, 705–708.

(49) Clarkson, T.; Wells, J. A. A Hot Spot of Binding Energy in a Hormone-Receptor Interface. *Science* 1995, 267, 383–386.

(50) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot-Spots: A Review of the Protein-Protein Interface Determinant Amino-Acid Residues. *Proteins: Struct., Funct., Genet.* 2007, 68, 803–812.

(51) Bogan, A. J.; Thorn, K. S. Anatomy of Hot Spots in Protein Interfaces. *J. Mol. Biol.* 1998, 280, 1–9.

(52) Laibinis, P. E.; Whitesides, G. M.; Allara, D. L.; Tao, Y. T.; Panikh, A. N.; Nuzzo, R. G. Comparison of the Structures and Wetting Properties of Self-Assembled Monolayers of n-Alkanethiols on the Coinage Metal Surfaces, Copper, Silver, and Gold. *J. Am. Chem. Soc.* 1993, 115, 7152–7167.

(53) Han, P.; Kurland, A. R.; Giordano, A. N.; Nanayakkara, S. U.; Blake, M. M.; Pochas, C. M.; Weiss, P. S. Heads and Tails: Simultaneous Exposed and Buried Interface Imaging of Monolayers. *ACS Nano* 2009, 3, 3115–3121.

(54) Hill, H. D.; Millstone, J. E.; Banholzer, M. J.; Mirkin, C. A. The Role of Curvature Plays in Thiolated Oligonucleotide Loading on Gold Nanoparticles. *ACS Nano* 2009, 3, 418–424.

(55) Liu, J. S. Monte Carlo Strategies in Scientific Computing; Springer-Verlag: New York, 2001.