Biotechnol Lett (2019) 41:79–90
https://doi.org/10.1007/s10529-018-2616-2

ORIGINAL RESEARCH PAPER

Clustered complementary amino acid pairing (CCAAP) for protein–protein interaction

Christina Kyung Eun Baek · Chang-Ho Baek

Received: 13 July 2018 / Accepted: 16 October 2018 / Published online: 24 October 2018

Abstract
Objectives Designing a polypeptide sequence to interact with a preselected target polypeptide sequence of a protein has long been of interest, yet remains an elusive goal.
Results Here, we propose a novel concept named "Clustered Complementary Amino Acid Pairing (CCAAP)," which plays an essential role in protein–protein interaction (PPI). Complementary amino acid pairing (CAAP) is a pairing between two amino acids encoded by a codon and its reverse complementary codon. CAAP interactions largely agree with the physicochemical and stereochemical requirements for probable amino acid pairings. Interestingly, 82 PPI structure data revealed that clusters of CAAP interactions (CCAAP boxes) are predominantly found in all PPI sites. Analysis of all amino acid pairings in the CCAAP boxes unveiled amino acid-pairing preferences and patterns for PPI that allowed us to develop a new method for designing an oligopeptide sequence to bind to a chosen polypeptide sequence of any target protein.
Conclusions Discoveries in the present study provide proof of the CCAAP principle.

Keywords Clustered complementary amino acid pairing · Protein detection · Protein–protein interaction · Protein targeting · Recombinant antibody · Synthetic antibody · Synthetic biology

Introduction

Specific targeting of a protein by a selected polypeptide sequence would be extremely useful in many branches of biotechnological sciences including disease prevention, diagnostics, and therapeutics. A number of approaches for predicting or identifying polypeptide sequences for said protein–protein interactions (PPI) have been developed. These approaches can be categorized into three different groups: computational prediction, massive library screening using a display system to identify protein interaction partners, and protein chip or microarray. However, none of these approaches provide a general pairing rule for protein–protein, protein-peptide, or peptide–peptide interaction.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10529-018-2616-2) contains supplementary material, which is available to authorized users.

C. K. E. Baek
Department of Bioengineering, University of California-Berkeley, Berkeley, CA, USA

C.-H. Baek (✉)
Peption LLC, San Diego, CA, USA
e-mail: contact@peptionllc.com

The existence of amino acid complementarity would provide an important insight into protein folding and PPI. There currently are three approaches for formulating amino acid complementarity: (1) Blalock's approach, which uses the hydropathic complementarity principle (molecular recognition theory) (Blalock and Smith 1984); (2) Root-Bernstein's approach, where peptides complementary to a given sequence are encoded by the antisense strand read in parallel to the sense strand (Root-Bernstein 1982); and (3) Siemion's approach, based on the periodicity of the genetic code (Siemion and Stefanowicz 1992). For all three approaches, a number of successful instances of the complementary peptide-antipeptide interactions have been reported. However, these results have been controversial due to logical contradictions and the inability to replicate some of these studies (Guillemette et al. 1989; Eberle et al. 1989; Kluczyk et al. 2004). Therefore, it is currently impossible to conclude which of the three approaches outlined above is most effective in predicting peptide-antipeptide interactions. In this study, however, we will focus on Blalock's approach using amino acid complementarity because it has been extensively confirmed in various applications (Siemion et al. 2004; Hardison and Blalock 2012).

More importantly, there is no published literature for a highly reliable algorithm for designing a complementary peptide sequence that can interact with a preselected target peptide sequence with high affinity and specificity, comparable to traditional animal-sourced antibodies. Here, we demonstrate a method using the clustered complementary amino acid pairing (CCAAP) concept to provide a general amino acid pairing rule for designing polypeptide synthetic antibody (sAb) sequences to interact with a chosen polypeptide sequence in any target protein.

Materials and methods

Materials

Synthetic peptides were purchased from Peptide 2.0 and are listed in Supplementary Table 1. Synthetic DNA fragments and oligonucleotides are listed in Supplementary Table 2. *E. coli* strain DH10B T1 [Thermo Fisher Scientific, catalog # 12331013] was used as a cloning host. *E. coli* strain BL21 Star (DE3) [Thermo Fisher Scientific, catalog # C601003] was used for the production of the recombinant proteins.

Construction of vectors

The bacterial expression vector, pET-21b, was obtained from EMD Millipore (catalog # 69741-3). The pET-21b vector was digested with SwaI/XhoI, and assembled with a linear 143 bp synthetic DNA fragment, 92_6HNLS or 93_6HNLS, using a seamless DNA assembly method following the manufacturer's protocol [Thermo Fisher Scientific, GeneArt™ Seamless Cloning and Assembly Enzyme Mix, catalog # A14606] to produce vector pC9-813-92 and vector pC9-813-93, respectively. The pC9-813-92 and pC9-813-93 vectors were digested with BamHI, and assembled with a PCR-amplified 1501 bp DNA fragment 92P [primer set: CH1424 and CH1425ART-R] or 93P [primer set: CH1425 and CH1425ART-R] from the *E. coli* MG1655 genome, corresponding to the *E. coli* alkaline phosphatase (AP) fusion, to generate pC9-813-92P and pC9-813-93P, respectively (Supplementary Fig. 1). The pC9-813-92P vector was digested with BglII, assembled with a 204 bp synthetic DNA fragment Sp-C9_813-821_CAA, corresponding to the CCAAP box tetramer recombinant antibody (rAb) against Cas9, to generate vector pC9-813-CAA4 (Supplementary Fig. 1). The pC9-813-CAA4 vector was digested with BglII, and self-ligated to remove 117 bp DNA fragment encoding two CCAAP boxes, producing pC9-813-CAA2 (Supplementary Fig. 1) which corresponds to the CCAAP box dimer antibody used to detect Cas9. To introduce two mutations, D153G and D330N, into the *E. coli* AP protein, we PCR-amplified three DNA fragments, P957-1 [primer set: CH1483ART-F and CH1486], P957-2 [primer set: CH1487 and CH1492], and P957-3 [primer set: CH1493 and CH1494] and assembled to produce a 1,473-bp DNA fragment corresponding to the mutant AP (or P957). This PCR product was digested with BamHI and XhoI, and ligated into BglII/XhoI digested pC9-813-CAA2, to generate p813C2-P957dB. For the production of the recombinant antibodies (rAbs), two synthetic DNA fragments, Anti-Bace1 (130 bp) and Anti-PDGFR (130 bp) (Supplementary Table 2), were digested with SwaI/BglII and ligated into the same enzyme site of the pC9-813-CAA2, to generate pAnti-Bace1-P and pAnti-PDGFR-P, respectively. Four synthetic DNA

fragments, Anti-Brca1 (124 bp), Anti-Hsp90 (124 bp), Anti-EstR (124 bp), and Anti-Xiap (124 bp) (Supplementary Table 2), were digested with SwaI/BglII and ligated into the SwaI/BamHI sites of the p813C2-P957dB, to generate pAnti-Brca1-P957, pAnti-Hsp90-P957, pAnti-EstR-P957, and pAnti-Xiap-P957, respectively. To produce the recombinant Cas9 protein, pET-Spy-Cas9_d6H vectors were constructed by assembling five parts with overlapping DNA ends using the seamless DNA assembly kit. Briefly, four insert parts [a 1000 bp Spy-Cas9_1, a 1030 bp Spy-Cas9_2, a 1030 bp Spy-Cas9_3, and a 1303 bp Spy-Cas9_5, corresponding to the tagless Cas9] (Supplementary Table 2) and the SwaI/XhoI-digested pET-21b were assembled, to create pET-Spy-Cas9_d6H.

Protein production and purification

For recombinant protein production, BL21 Star (DE3) cells harboring an expression vector were grown to mid-log phase (optical density at 600 nm [OD600] of 0.6) in LB medium [ampicillin (Amp), 100 μg/ml] at 28 °C and induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 5 h. Cells were harvested by centrifugation at 3000×*g* for 10 min. Harvested cells were disrupted using a chemical lysis method following the manufacturer's protocol [Thermo Fisher Scientific, B-PER™ Complete Bacterial Protein Extraction Reagent, catalog # 89821]. Cell debris and insoluble proteins in the lysate were separated by centrifugation at 16,000×*g* for 5 min. His-tagged recombinant proteins were purified via metal-affinity chromatography using Dynabeads™ His-Tag Isolation and Pulldown beads following the manufacturer's protocol [Thermo Fisher Scientific, catalog # 10103D]. Recombinant Cas9 proteins were purified using the HiTrap Heparin HP column [GE Healthcare, catalog # 17-0406-01] as previously described (Karvelis et al. 2015).

Dot blot and western blot analyses

For dot blot analysis, 2 μl (5 μg) of samples were spotted onto a nitrocellulose (NC) membrane and dried completely. Then, non-specific sites were blocked by soaking the membrane in blocking solution [Thermo Fisher Scientific, WesternBreeze™ Blocker/Diluent (Part A and B), catalog # WB7050] for 1 h at room temperature (or up to 72 h at 4 °C). The membrane was washed twice with water (1 ml per cm^2 of membrane), and incubated with the 1st antibody (Ab) in a binding/wash (BW) buffer [50 mM sodium phosphate, pH 8.0, 300 mM NaCl, and 0.01% Tween 20] or TBS-T [25 mM Tris, pH 7.5, 0.15 M NaCl, 0.05% Tween 20] for 1 h at room temperature. The membrane was washed 4 times (2 min per wash) with wash buffer [Thermo Fisher Scientific, WesternBreeze™ Wash Solution, catalog # WB7003] or TBS-T. If the 1st Ab was Anti-Cas9 Ab-HRP conjugate [Thermo Fisher Scientific, catalog # MAC133P] or the peptide-AP fusions (2nd Ab not required), the membrane was washed twice with water, and incubated with a chromogenic substrate: Chromogenic Substrate (TMB) [Thermo Fisher Scientific, catalog # WP20004] for HRP and NBT/BCIP substrate solution for AP [Thermo Fisher Scientific, catalog # 34042]. Otherwise, the membrane was incubated with 2nd Ab in the blocking solution for 1 h. To detect His-tagged peptide and proteins, the Anti-6His Ab-HRP conjugate [Thermo Fisher Scientific, catalog # 46-0707] was used as 2nd Ab. Then the membrane was washed four times with the wash buffer and two times with water. Finally, the blot was incubated with the chromogenic substrates. For the western blot analysis, the protein samples were resolved in 4–20% gradient SDS-PAGE gel, transferred to an NC membrane, and analyzed using the same method for the dot blot analysis.

Digital image processing and analysis

For image processing, we used Adobe Photoshop 7.0. Quantitative image analysis of the digital images was carried out using measuring tools in imaging software ImageJ (Schneider et al. 2012). Image analysis results were calculated by averaging data from three independent experiments.

Statistical analysis

Statistical analyses were performed using a one-way analysis of variance (ANOVA) and confirmed by Student's *t* test [two tails, two-sample equal variance (homoscedastic)]. *p* values < 0.05 are considered statistically significant. Results are categorized into 5 levels based on their *p* value: $^{\blacklozenge}p < 0.05$; $^{\blacklozenge\blacklozenge}p < 0.01$;

$^{\blacklozenge\blacklozenge\blacklozenge}p < 0.001$; $^{\blacklozenge\blacklozenge\blacklozenge\blacklozenge}p < 0.0001$; and; $^{\blacklozenge\blacklozenge\blacklozenge\blacklozenge\blacklozenge}p < 0.00001$. All graphs display mean ± SD.

Results and discussion

Physicochemical and stereochemical features of the complementary amino acid pairing (CAAP)

To deliver a new concept for PPI, we first demonstrate that the pairing between two amino acids encoded by a codon and the reverse complementary codon (c-codon) is favored in PPI. We name this pairing the "Complementary Amino Acid Pairing (CAAP)." We summarize all possible CAAPs in Fig. 1. Based on the side chain hydrophobicity and polarity (Eisenberg et al. 1984), we categorize CAAP interactions (↔) into the following groups: ①, hydrophobic (nonpolar/neutral) ↔ hydrophobic (nonpolar/neutral) [6.9%]; ②, hydrophobic (nonpolar/neutral) ↔ hydrophilic (polar/positively charged) [17.2%]; ③, hydrophobic (nonpolar/neutral) ↔ hydrophilic (polar/neutral) [27.6%]; ④, hydrophobic (nonpolar/neutral) ↔ hydrophilic (polar/negatively charged) [13.8%]; ⑤, hydrophobic (nonpolar/neutral) ↔ hydrophilic (nonpolar/neutral) [6.9%]; ⑥, hydrophobic (nonpolar/neutral) ↔ hydrophobic (polar/neutral) [6.9%]; ⑦, hydrophilic (nonpolar/neutral) ↔ hydrophilic (polar/positively charged) [6.9%]; ⑧, hydrophilic (polar/neutral) ↔ hydrophilic (polar/positively charged) [7.9%]; ⑨, hydrophilic (nonpolar/neutral) ↔ hydrophilic (polar/neutral) [3.4%]. According to our categorization, group ① and ⑥ pairings (A-C, A-G, I-Y, and V-Y) possess hydrophobic interactions, while group ⑧ and ⑨ pairings (2 R-S, R-T, and S-T) may form hydrogen bonds (Root-Bernstein 1982). Some of the group ② and ③ pairings involve charge transfer complexing (F-K) and hydrogen bonding (A-R and C-T) (Root-Bernstein 1982). However, most of the group ② and ③ (2 L-Q, A-S, D-I, D-V, E-F, G-S, G-T, H-M, I-N, L-K, and N-V) and group ⑦ (2 P-R) pairings have not been systematically evaluated for intermolecular interactions before. Interestingly, 38% of CAAP interactions in Fig. 1 (√ group) belong to the

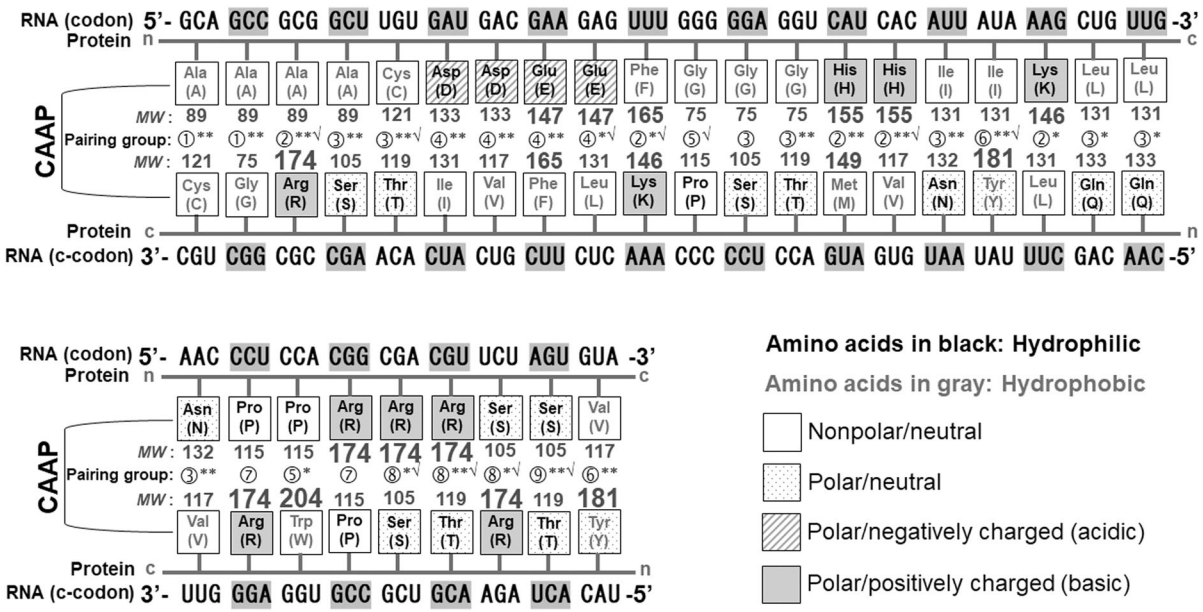


Fig. 1 Complementary amino acid pairing (CAAP) for 20 amino acids. The codon-complementary codon (c-codon) pairings for all possible CAAP interactions are shown at the top and bottom of the corresponding amino acids. Physicochemical properties of amino acids are shown in gray (hydrophobic), black (hydrophilic), white box (nonpolar/neutral), dotted box (polar/neutral), striped box (polar/negatively charged, acidic), and gray box (polar/positively charged, basic). CAAP interactions (↔) between two amino acids are categorized by side chain hydrophobicity and polarity: ①–⑨ (Eisenberg et al. 1984); asterisk(s), amino acid pairings favored in the antiparallel alignment only (*) or both parallel/antiparallel alignments (**) (Zhang et al. 2010); and √, probable amino acid pairings consistent with the bonding rules (Root-Bernstein 1982). *MW*, molecular weight

group of 26 probable amino acid pairings that can be formed (Root-Bernstein 1982). In addition, we found that 65% of the CAAP interactions are favored amino acid pairs [Relative Frequency (RF) > 1.0] in parallel β-strand interactions and 88% are favored in antiparallel strands (Zhang et al. 2010). Moreover, CAAP interactions have been shown to possess favorable stereochemistry. Notably, all high molecular weight (large) residues with bulky side chains such as Arg (R), Tyr (Y), and Trp (W) tend to pair with low molecular weight (small) residues with small side chains, while there is no CAAP interaction between two high molecular weight residues (Fig. 1). Therefore, the CAAP interactions may possess spatial flexibility at the PPI interface. These observations lead us to postulate that the physicochemical and stereochemical natures of the CAAP relationships between two polypeptide chains may provide an attractive environment for PPI.

The CAAP interactions are clustered in all PPI sites

To address the CAAP hypothesis for PPI, we first focused on finding CAAP interactions in the PPI structure database from the Protein Data Bank (PDB). We examined the well-known leucine zipper proteins (Landschulz et al. 1988): *Saccharomyces cerevisiae* GCN4/GCN4 homodimer [PDB_2ZTA], *Mus musculus* NF-k-B essential modulator (NEMO) homodimer [PDB_4OWF], *Homo sapiens* c-Jun/c-Fos heterodimer [PDB_1FOS], and *Rattus norvegicus* C/EBPA homodimer [PDB_1NWQ] (Fig. 2). We also examined five non-leucine-zipper proteins which include three helix-helix (Fig. 3a) and two β-sheet-β-sheet (Fig. 3b) interactions: *Saccharomyces cerevisiae* Put3 homodimer [PDB_1AJY], *Salmonella enterica* serovar Typhimurium TarH homodimer [PDB_1VLT], *Mus musculus* E47-NeuroD1 heterodimer [PDB_2QL2], *Arenicola marina* (lugworm) Arenicin-2 homodimer [PDB_2L8X], and *Laticauda semifasciata* Erabutoxin homodimer [PDB_1QKD]. We first determined the linear sequence representation of the dimers' protein sequences (Figs. 2, 3). In the global alignment for parallel interactions, the dimer molecules are aligned to obtain optimal homology matching (Needleman and Wunsch 1970). For the antiparallel interaction, however, global alignment is not applicable (Fig. 3b). In CAAP alignment, dimer molecules are aligned such that CAAP interactions correspond to the PDB PPI structure data, which we confirmed was when the dimers were shifted by one amino acid from each other in the global alignments (Figs. 2, 3). Clusters of CAAP residues are enclosed by a gray box called "CCAAP box" where at least 37.5% are CAAPs. Additionally, the CCAAP box is defined to enclose eight or more amino acid pairings for helix/helix, helix/coil, and coil/coil interactions. However, for β-sheet/β-sheet and β-sheet/coil interactions, the CCAAP box is defined to enclose five or more amino acid pairings. We set these CCAAP box criteria after discovering that a CCAAP box with 37.5% or higher CAAP content does not randomly occur in the non-PPI areas (Figs. 2, 3). In the CAAP alignments of the nine dimer proteins (Figs. 2, 3), we found 21 CCAAP boxes. Interestingly, 20 out of 21 CCAAP boxes are found in the PPI sites (Figs. 2, 3). In addition, all PPI sites corresponded to at least one CCAAP box (Figs. 2, 3). Conversely, we found only one CCAAP box in the non-PPI area of the TarH Homodimer [PDB_1VLT] (Fig. 3). Importantly, the clustered appearance of the CAAP interactions in the PPI sites is statistically significant (Supplementary Fig. 2, Supplementary Table 3). We then translated the linear sequence representation to its helical wheel representation to simulate the hypothesized α-helix structural configuration of the residues (Figs. 2, 3a) (McLachlan and Stewart 1975). The rotational angle (topology) of the two interacting molecules in the helical wheel representation was adjusted by comparing it with the PDB structure data to build a realistic simulation. In the helical wheel representation, we found that 50% of CAAP interactions in the linear representation are clearly aligned at the interface of the two interacting helices (Figs. 2, 3b). The helical wheel representation also revealed new CAAP interactions (underline) that could not be identified in the linear representations (Figs. 2, 3a). Conversely, 50% (dotted underline) of the CAAP interactions in the linear representation were lost in the helical wheel representations (Figs. 2, 3a). While we do not presently understand the meaning of the different CAAP configurations between the linear and the helical wheel representations, the PDB PPI structure data revealed that the CCAAP boxes in the linear representation are mostly linked with the actual PPI sites (Figs. 2, 3a). Rarely, however, we found that one out of 21 CCAAP boxes in the linear representation is

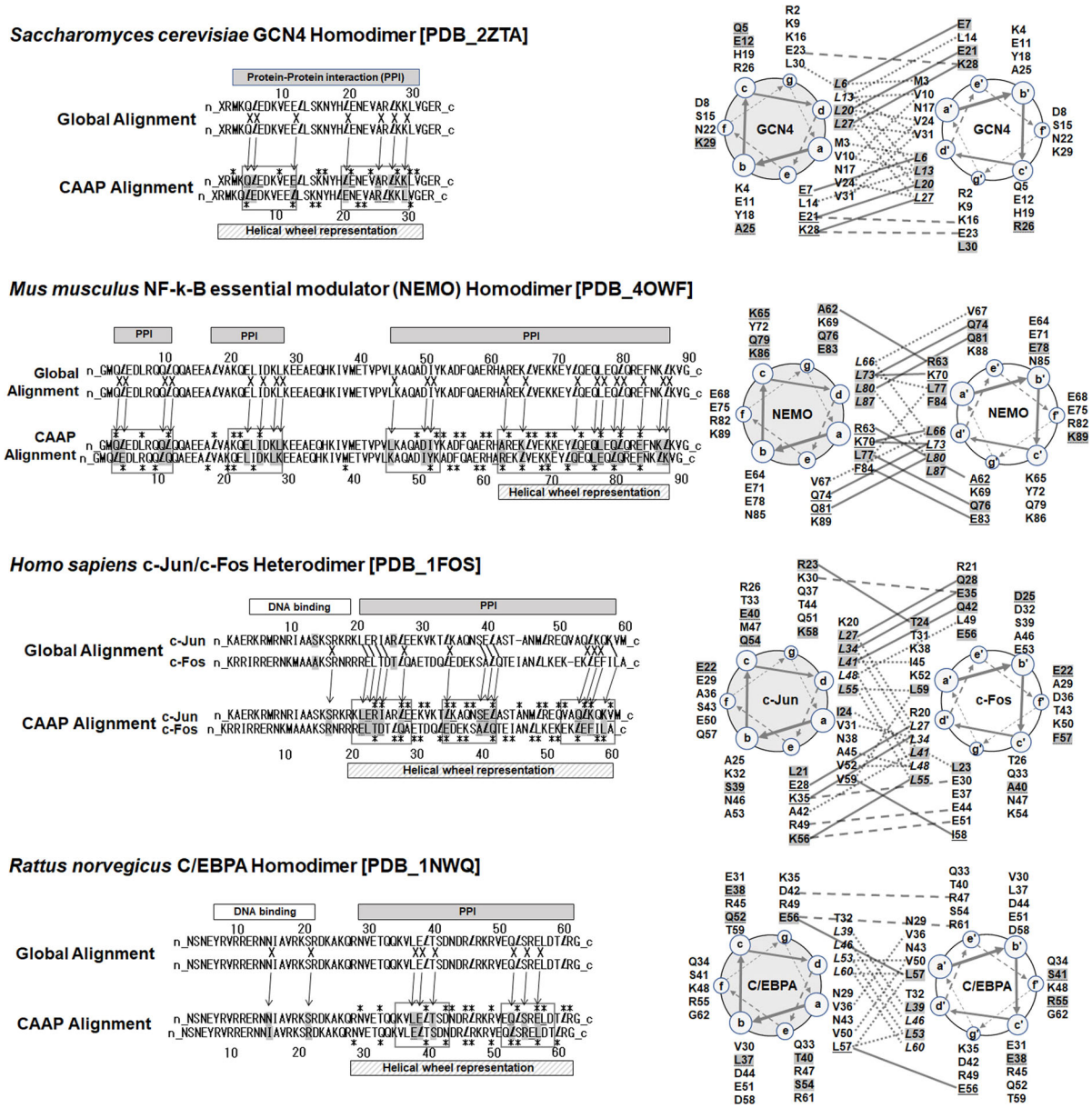


Fig. 2 The CCAAP boxes are found in the protein–protein interaction (PPI) site(s) of the leucine-zipper proteins. Global alignment and CAAP alignment in the linear representation of the four leucine-zipper proteins: *Saccharomyces cerevisiae* GCN4/GCN4 homodimer [PDB_2ZTA], *Mus musculus* NF-k-B essential modulator (NEMO) Homodimer [PDB_4OWF], *Homo sapiens* c-Jun/c-Fos heterodimer [PDB_1FOS], and *Rattus norvegicus* C/EBPA Homodimer [PDB_1NWQ]. Corresponding helical wheel representation is shown at the right-hand side of each CAAP alignment. In the linear representation, the leucine residues of the leucine-zipper are indicated by italic letters. The CAAP residues are highlighted in gray. The CCAAP boxes enclosing a cluster of the CAAP interactions are indicated by the gray boxes. The PPI sites are identified by a cluster of residues (asterisks) that have intermolecular interaction(s) of $<$ 3.6 Å distance, and indicated by gray bars on the top of the linear alignments. In the helical wheel representation, the new CAAP residues (that could not be identified in the linear representation) are underlined. Conversely, the CAAP residues (in the linear representation) losing the CAAP configuration in the helical wheel representation are indicated by dotted underline. The CAAP interactions in the helical wheel representation are indicated by gray lines. Hydrophobic and charged interactions are indicated by gray-dotted and gray-dashed lines, respectively. The possible CAAP interactions in the global alignments are indicated by letters (X, /, or \) between two molecules

not linked with the actual PPI site (Figs. 2, 3). Additionally, we took note that many amino acids in the PPI interface most likely interact with multiple amino acids < 4 Å away (Figs. 2, 3a).

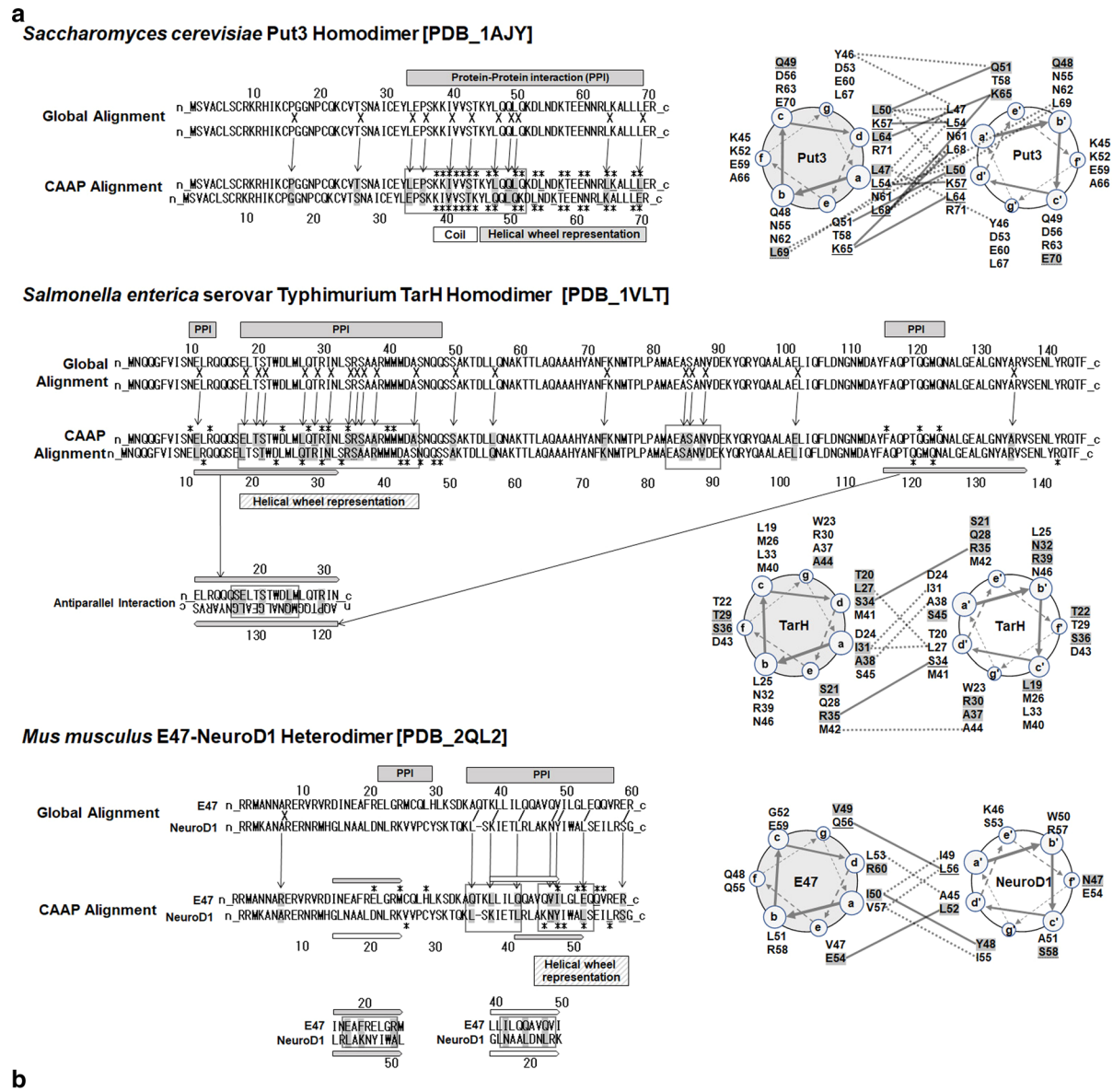
We also investigated 75 additional PPI structures for CCAAP interactions (Supplementary Table 4). A total of 84 protein structures were selected for their relatively simple PPI structures, which limit the effect of any other potential parameters. Protein structures were also categorized according to parallel or antiparallel alignment. We found CCAAP boxes in all PPI sites in the 82 structure data from PDB (Supplementary Table 4). However, we could not find any CCAAP box from PPI sites of two dimers: *Homo sapiens* ERBB2-EGFR heterodimer [PDB_2KS1] and *Bos taurus* IF1 homodimer [PDB_1GMJ]. Interestingly, the PPI sites of these two dimers have a high content of either charged amino acid pairings (D-K, E-H, E-K, E-R, and H-K) [PDB_2KS1] or hydrophobic amino acid pairings (A-I, A-M, F-I, F-L, G-G, G-V, I-L, I-M, L-L, L-V, and V-V) [PDB_1GMJ]. These non-CAAP electrostatic and hydrophobic interactions are not predictable by any amino acid complementarity model including the hydropathic complementarity principle (Blalock and Smith 1984). We found 79 CCAAP boxes in parallel (↓↓) interactions (76 helix/helix, 2 β-sheet/coil, and 1 β-sheet/β-sheet interactions) and 81 CCAAP boxes in antiparallel (↓↑) interactions (67 helix/helix and 14 β-sheet/β-sheet interactions) (Supplementary Table 4). Notably, 93% of the β-sheet/β-sheet interactions are antiparallel interactions.

Design synthetic antibodies (sAbs) using the CCAAP principle

We assessed the composition of all amino acid pairings in the CCAAP boxes (Supplementary Table 4) to obtain information on pairing preference and how the CAAPs were spaced out in the CCAAP box, which may be important factors for binding affinity, specificity, and stability. The raw abundance numbers of all amino acid pairings in the CCAAP boxes (Supplementary Table 4) are shown in Supplementary Table 5 and summarized in Supplementary Fig. 3. These data were then used for designing an oligopeptide synthetic antibody (sAb) sequence that can interact with a target polypeptide sequence of a protein. The general rule was to design the sAb sequence such that it forms a CCAAP box in the PPI with the target sequence. For the spacing, we tried to mimic some CCAAP box examples covering diverse spacing patterns (Supplementary Table 4): OXXOXOXOO [PDB_1YKH], OXOOOOXXX [PDB_3NMD], OXOOOOXO [PDB_4ZM8], OOXOOXOO [PDB_3VIR], OOXOOOXOO [PDB_4BWN], OOXXOOXO [PDB_3VMX], OOOXOXOOO [PDB_2WT7], and OOOOOXOOOO [PDB_4XA1] (O stands for a CAAP interaction residue, X stands for a non-CAAP interaction residue, and positions modified from that observed are underlined). These spacing formats with no or minor modifications allowed us to test many different sAb designs with a range of CAAP contents (55%–90%). We designed the CAAP content to be greater than 55%, since the median value of the natural range (between 37.5% and 75%) of the CAAP content in the 137 CCAAP boxes was 53.8%. For each designated CAAP or non-CAAP pairing, we generally selected the most frequent pairing partner according to the data in Supplementary Fig. 3b and Supplementary Table 4.

CAAP-based sAbs interact specifically with the preselected peptide sequence in the target protein

To test the sAb design tool based on the CCAAP principle, we selected a target sequence in the HNH domain of the *Streptococcus pyogenes* Cas9 protein [PDB_5B2R]. *S. pyogenes* CRISPR-Cas9 system has been broadly applied to edit the genome of bacterial and eukaryotic cells. The target sequence for Cas9 is n_EKLYLYYLQ_c (Helix: E813 to Q821). We designed two different types of synthetic antibody (sAb) molecules, sAb monomer (PTD13, Supplementary Table 1) and sAb dimer (PTD14, Supplementary Table 1), to detect the target protein sequences. The dimer constructs in this study are designed to have a U-shape linker (CYPEN or KTGEVNN) from the Lin-7/Lin2 heterodimer (PDB_1ZL8). The intention behind the U-shape (↓_↑) was to allow for CAAP interactions in both parallel and antiparallel orientations simultaneously. As shown in the dot blot experiment (Fig. 4a), the sAb monomer (PTD13) and sAb dimer (PTD14) could interact with the target peptide (PTD12, Supplementary Table 1), but there was no detectable interaction with the control peptide (PTD8, unrelated peptide, Supplementary Table 1).



No signal was detected from the no peptide control (Fig. 4a). Remarkably, the sAb dimer (PTD14) showed a stronger (twofold) interaction than that of the sAb monomer PTD13 (Fig. 4a).

To verify these results, we first produced three CCAAP-based recombinant antibody (rAb) constructs, C9-813-92P (monomer, parallel), C9-813-93P (monomer, antiparallel), and C9-813-CAA2 (dimer, antiparallel and parallel). As shown in Fig. 4b,

◀**Fig. 3** The CCAAP boxes are found in the protein–protein interaction (PPI) site(s) of the non-leucine-zipper proteins. Global alignment and CAAP alignments in the linear representation of the five non-leucine-zipper proteins, three helix-helix (**a**) and two β-sheet-β-sheet (**b**) interactions: *Saccharomyces cerevisiae* Put3 Homodimer [PDB_1AJY], *Salmonella enterica* serovar Typhimurium TarH Homodimer [PDB_1VLT], *Mus musculus* E47-NeuroD1 Heterodimer [PDB_2QL2], *Arenicola marina* (lugworm) Arenicin-2 Homodimer [PDB_2L8X], and *Laticauda semifasciata* Erabutoxin Homodimer [PDB_1QKD]. Corresponding helical wheel representation is shown at the right-hand side of each CAAP alignment. The CAAP residues are highlighted in gray. The CCAAP boxes enclosing a cluster of the CAAP interactions are indicated by the gray boxes. The PPI sites are identified by a cluster of residues (asterisks) that have intermolecular interaction(s) of < 3.6 Å distance, and indicated by gray bars on the top of the linear alignments. The CAAP interactions in the helical wheel representation are indicated by gray lines. Hydrophobic and charged interactions are indicated by gray-dotted and gray-dashed lines, respectively. Possible CAAP interactions in the global alignments are indicated by letters (X or /) between two molecules. The PDB structure data also revealed some regional interactions that do not appear in the linear alignments: gray-arrow bars in PDB_1VLT and gray- and white-arrow bars in PDB_2QL2

we confirmed that the rAb C9-813-CAA2 (dimer, antiparallel and parallel) has stronger (2.5-fold) interaction with the Cas9 target sequence (PTD12) than the rAb C9-813-92P (monomer, parallel) or rAb C9-813-93P (monomer, antiparallel). In addition, a semi-quantitative assay revealed that the rAb C9-813-CAA2 (dimer) can interact with an approximately fivefold lower concentration of the target peptide (PTD12) near the detection limit than either the rAb C9-813-92P (monomer) or the rAb C9-813-93P (monomer) does (Supplementary Fig. 4a). We confirmed this phenomenon in two additional cases: detecting alkaline phosphatase (AP) (Supplementary Fig. 6) and PDGF-B (Supplementary Fig. 7).

Finally, we further examined the performance of the CCAAP oligopeptides by testing whether they can detect the whole Cas9 protein in both non-denatured (dot blot) and denatured (western blot) conditions (Fig. 4c). We used a recombinant Cas9 protein. The purified Cas9 protein is shown in Fig. 4c (Coomassie stain). We used the sAb monomer (PTD13) and sAb dimer (PTD14) as the 1st Ab to detect Cas9 protein. The anti-Cas9 Ab-HRP conjugate was used as positive control 1st Ab in the western blot experiment (Fig. 4c). The sAb dimer (PTD14) was able to detect the Cas9 protein in both the dot blot and western blot, while the monomer and the no peptide (negative control) were unable to detect the Cas9 protein (Fig. 4c). Notably, although the sAb monomer (PTD13) detected the synthetic Cas9 target oligopeptide (PTD12) in the dot blot experiment (Fig. 4a), it failed to detect the whole Cas9 protein (Fig. 4c). This may reflect two probable factors: (1) the molecular weight difference between the target oligopeptide PTD12 (1 kDa,) and Cas9 (160 kDa), which caused the molar ratio (PTD12:Cas9) in the same amount (5 μg) of the samples used for the dot blots to be 160:1, and (2) the structural difference of the same target peptide sequence between that in a small oligopeptide and a whole Cas9 protein. Lastly, we confirmed that the rAb C9-813-CAA2 (dimer) can interact with the purified Cas9 protein (Supplementary Fig. 4b). Additionally, this semi-quantitative assay showed that the binding affinity of the rAb C9-813-CAA2 (dimer) to Cas9 protein is comparable to the conventional anti-Cas9 Ab-HRP conjugate (Supplementary Fig. 4b).

In order to assess the specificity of the rAb C9-813-CAA2 (dimer), we carried out a western blot experiment to detect an unpurified Cas9 protein in *E. coli* crude extract. We found that the rAb C9-813-CAA2 (dimer) showed highly specific interaction with the Cas9 protein while this rAb dimer displayed no detectable non-specific interaction with any of the *E. coli* proteins (Supplementary Fig. 5).

To generalize the CCAAP principle for protein targeting, we have designed a synthetic antibody (sAb) construct and 6 recombinant antibody (rAb) constructs to detect 7 additional clinically important proteins: Anti-PDGF sAb (PTD18, Supplementary Table 1) for Human Platelet-Derived Growth Factor B (PDGF-B) [PDB_3MJG]; Anti-Bace1 rAb for Human Bace1 [PDB_4B05]; Anti-Brca1 rAb for Human Brca1 [PDB_3PXE]; Anti-Hsp90 rAb for Human Hsp90 [PDB_2VCI]; Anti-EstR rAb for Human Estrogen Receptor [PDB_1A52]; Anti-Xiap rAb for Human Xiap [PDB_2KNA]; and Anti-PDGFR rAb for PDGF Receptor (PDGFR) [PDB_3MJG] (Fig. 4d). These proteins are important clinical target candidates for the treatment of cancer and Alzheimer diseases. The dot blot analysis showed that all sAbs and rAbs can specifically interact with their target oligopeptides, while they have no or very weak interaction with the unrelated target oligopeptides, which cannot form a CCAAP box (Fig. 4d). However, the binding affinities of these interactions appeared to be varied as described

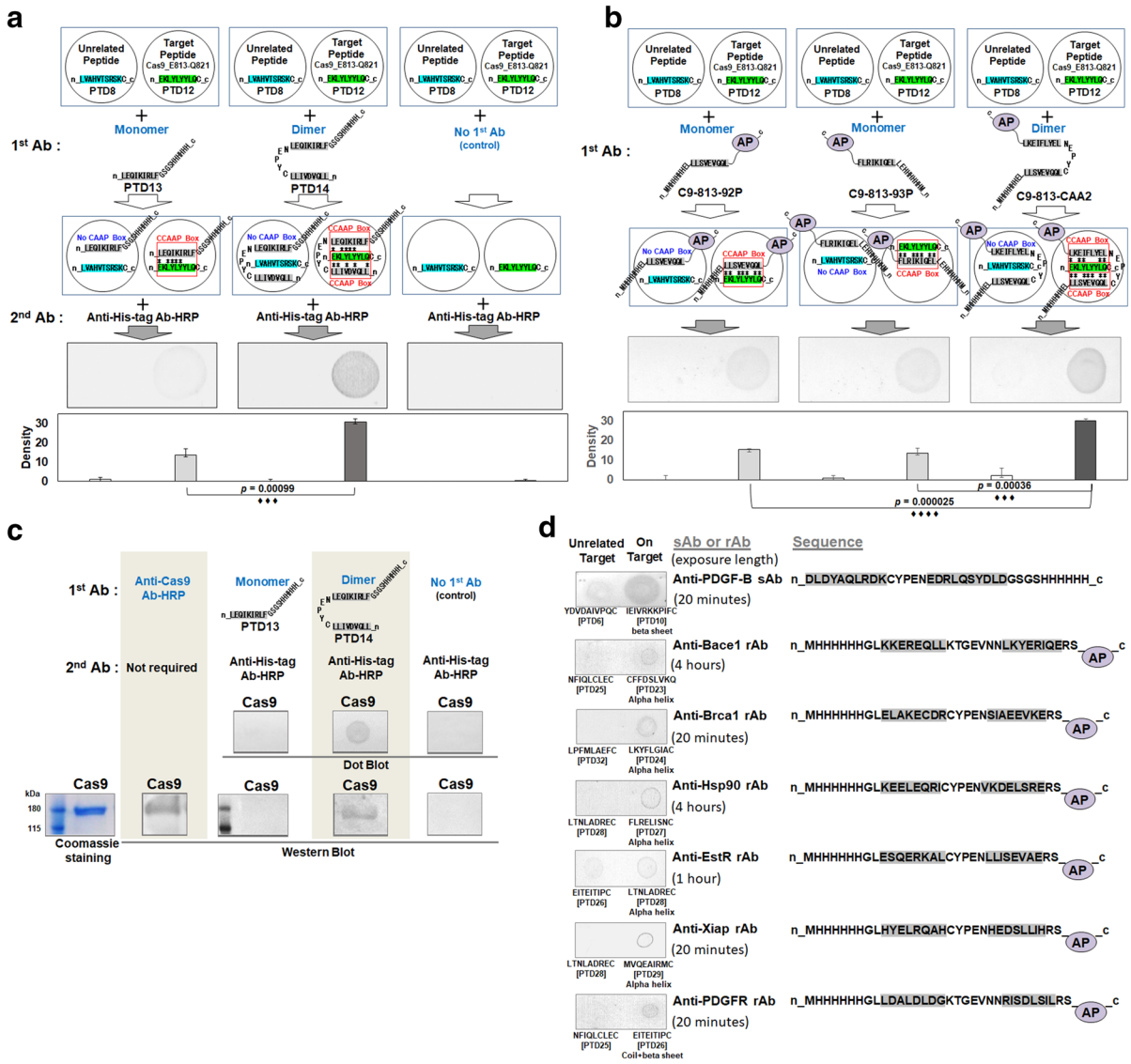


Fig. 4 CCAAP-based sAbs and rAbs can interact with the preselected peptide sequences of the target proteins. **a** Dot blot analysis to detect the Cas9 target sequence using the His-tagged synthetic CCAAP oligopeptides (sAbs) as 1st Abs: synthetic His-tagged CCAAP sAb monomer (PTD13) and synthetic His-tagged CCAAP sAb dimer (PTD14). No peptide used for the negative control. CAAP interactions are shown in asterisks. **b** Dot blot analysis to detect the Cas9 target sequence using the recombinant CCAAP oligopeptides-alkaline phosphatase (AP) fusion proteins (rAbs) as 1st Abs: C9-813-92P (monomer, parallel), C9-813-93P (monomer, antiparallel), and C9-813-CAA2 (dimer, parallel-linker-antiparallel). CAAP interactions are shown in asterisks. **c** Dot blot and western blot analyses to detect the whole Cas9 proteins using the His-tagged CCAAP oligopeptide synthetic antibodies (sAbs). The CCAAP sAb monomer (PTD13) and dimer (PTD14) were used as 1st Abs. No 1st Ab was used for the negative control. The Anti-Cas9 Ab-HRP conjugate was used as positive control 1st Ab to detect Cas9 protein. **d** Dot blot analysis to detect preselected target sequences in 7 additional target proteins using synthetic and recombinant antibodies (sAbs and rAbs). The blots in **a**, **b**, and **c** were incubated with the chromogenic substrates for 15 min to visualize the CCAAP sAb-Cas9 interaction. The dot blots in **d** were incubated with the chromogenic substrates for various lengths of incubation time (expose length) to obtain a sufficient intensity of the blot images. The Selected images are representing similar results from three independent experiments. The *p* values for the densitometry data were obtained using a one-way ANOVA

in Fig. 4d (different expose time lengths). Although the target polypeptide sequence is a key determinant for the binding affinity, we believe that designing an ideal binding sequence for a sAb may reduce the range of variation in the binding strengths.

In the present study, we have developed a novel CCAAP principle and obtained experimental evidence that CCAAP box is a critical driving force for PPI. Therefore, we conclude that the CCAAP concept can be applied to design sAb or rAb that can specifically interact with a preselected oligopeptide sequence (8–10 amino acids) in the target protein. However, we believe that there are opportunities to improve the CCAAP principle for affinity and specificity.

Acknowledgements The authors acknowledge that the full financial support for this research received directly from Peption LLC (San Diego, CA, USA).

Supporting information Supplementary Fig—1 Expression vectors for the production of the recombinant antibodies (rAbs).

Supplementary Fig—2 The clustered appearance of the CAAP interactions in the PPI sites is statistically significant ($^{\blacklozenge\blacklozenge\blacklozenge\blacklozenge\blacklozenge}p < 0.00001$). The abundance of the CAAP interactions in the PPI and non-PPI sites was calculated by averaging the % of CAAP interactions from the CAAP alignment samples in Fig. 2 and 3 (Supplementary Table 3). The p value was obtained using a one-way ANOVA.

Supplementary Fig—3 Composition (a) and pairing frequencies (b) of amino acids in the CCAAP boxes from the exemplary 82 crystal structure data. The data for parallel interactions and the antiparallel interactions are shown in dark bars and light bars, respectively. The bar graphs for cysteine, histidine, proline, and tryptophan are not included since they rarely appeared (<32 times).

Supplementary Fig—4 Semi-quantitative assays to test binding affinities. (a) Dot blot analysis to determine the detection limits of the CCAAP-based rAbs, C9-813-92P (monomer) and C9-813-CAA2 (dimer) against the target peptide (PTD12) at various concentrations. (b) Dot blot analysis to compare the detection limits of rAb C9-813-CAA2 (dimer) and the conventional anti-Cas9 Ab-HRP conjugate against the purified Cas9 protein at various concentrations.

Supplementary Fig—5 The CCAAP-based rAb C9-813-CAA2 interacts specifically with the target Cas9 protein in the *E. coli* BL21 Star (DE3) crude extract. Western blot analysis was carried out using rAb C9-813-CAA2 (dimer) to detect the whole Cas9 protein which is produced in *E. coli* BL21 Star (DE3) cells harboring pET-Spy-Cas9-dH6. The *E. coli* BL21 Star (DE3) strain harboring pET-21b was used as negative control.

Supplementary Fig—6 Dot blot analysis to detect the alkaline phosphatase (AP) target sequence (PTD8) using the CCAAP-based oligopeptide synthetic antibodies (sAbs) as 1st Abs: sAb monomer (PTD15) and sAb dimer (PTD16). Synthetic linker-His-tag oligopeptide (PTD20) was used as a negative control sAb. The synthetic oligopeptide PTD7 was used as an unrelated target.

Supplementary Fig—7 Dot blot analysis to detect the PDGF-B target sequence (PTD10) using the CCAAP-based oligopeptide synthetic antibodies (sAbs) as 1st Abs: sAb monomer (PTD17) and sAb dimer (PTD18). Synthetic linker-His-tag oligopeptide (PTD20) was used as a negative control sAb. The synthetic oligopeptide PTD6 was used as an unrelated target.

Supplementary Table—1 Synthetic antibodies (sAbs) and target peptides used in this study.

Supplementary Table—2 Synthetic DNA fragments and oligonucleotides used in this study.

Supplementary Table—3 Appearance of the CAAP interactions in the PPI and non-PPI sites.

Supplementary Table—4 Clustered complementary amino acid pairing (CCAAP) for protein-protein interaction.

Supplementary Table—5 Abundance of the amino acid pairings in the CCAAP boxes from 82 PPI structure data.

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Blalock JE, Smith EM (1984) Hydropathic anti-complementarity of amino acids based on the genetic code. Biochem Biophys Res Commun 121:203–207

Eberle AN, Drozdz R, Baumann JB, Girard J (1989) Receptor-specific antibodies by immunization with "antisense" peptides? Pept Res 2:213–220

Eisenberg D, Weiss RM, Terwilliger TC (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci USA 81:140–144

Guillemette G, Boulay G, Gagnon S, Bosse R, Escher E (1989) The peptide encoded by angiotensin II complementary RNA does not interfere with angiotensin II action. Biochem J 26:309

Hardison MT, Blalock JE (2012) Molecular recognition theory and sense-antisense interaction: therapeutic applications in autoimmunity. Front Biosci (Elite Ed) 4:1864–1870

Karvelis T, Gasiunas G, Young J, Bigelyte G, Silanskas A, Cigan M, Siksnys V (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol 16:253

Kluczyk A, Cebrat M, Zbozień-Pacamaj R, Lisowski M, Stefanowicz P, Wieczorek Z, Siemion IZ (2004) On the peptide-antipeptide interactions in interleukin-1 receptor system. Acta Biochim Pol 51:57–66

Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764

McLachlan AD, Stewart M (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol 98:293–304

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Root-Bernstein RS (1982) Amino acid pairing. J Theor Biol 94:885–894

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675

Siemion IZ, Stefanowicz P (1992) Periodical changes of amino acid reactivity within the genetic code. Biosystems 27:77–84

Siemion IZ, Cebrat M, Kluczyk A (2004) The problem of amino acid complementarity and antisense peptides. Curr Protein Pept Sci 5:507–527

Zhang N, Duan G, Gao S, Ruan J, Zhang T (2010) Prediction of the parallel/antiparallel orientation of beta-strands using amino acid pairing preferences and support vector machines. J Theor Biol 263:360–368