Protein Structures

DOI: 10.1002/anie.200603246

A Network of Coiled-Coil Associations Derived from Synthetic GCN4 Leucine-Zipper Arrays**

*Michael Portwich, Sandro Keller, Holger M. Strauss, Carsten C. Mahrenholz, Ines Kretzschmar, Achim Kramer, and Rudolf Volkmer**

The α-helical coiled coil was one of the first protein motifs whose structure was elucidated in detail.[1] First described by Crick in 1953,[2] the coiled coil is composed of at least two right-handed amphipathic α helices that wrap around each other into a left-handed supercoil such that their hydrophobic surfaces are in continuous contact (Figure 1a). Coiled-coil structures can associate up to pentamers, form homomeric and heteromeric complexes at different stoichiometries, and be aligned in parallel and antiparallel topologies.[1] A characteristic of all coiled coils is the presence of heptad repeat sequences [**abcdefg**]$_i$, where i denotes the heptad number (Figure 1a). Hydrophobic amino acids are usually required at core positions **a** and **d** and are crucial for folding.[3] Positions **b**, **c**, **e**, **f**, and **g** face the outside of the assembly and are generally occupied by hydrophilic residues. Furthermore, inter- and intrahelical salt bridges formed by charged residues frequently found at positions **b**, **c**, **e**, and **g** may stabilize the structure.[4] The "peptide velcro" hypothesis[5] summarizes these findings and outlines that 1) **a** and **d** are hydrophobic, 2) **e** and **g** are charged, and 3) **b**, **c**, and **f** are hydrophilic. Although coiled-coil motifs can be predicted with a high degree of confidence,[6] predicting their association states and topologies is still a great challenge.

Different selection systems have been established for screening recombinantly expressed coiled-coil complexes.[7] An impressive library-versus-library approach for identifying pairs of interacting coiled-coil-forming polypeptides was developed by Plückthun, Michnick, and co-workers.[8]

Recently, a glass-chip experiment has been used to monitor a network of bZIP transcription factors at the protein level.[9] Herein we report on the development of optimized synthetic peptide arrays[10] useful for analyzing and characterizing coiled-coil associations at the amino acid level. In contrast to rational design, which mostly depends on short model peptides, our approach relies on the full-length homodimeric GCN4 leucine-zipper coiled-coil domain (Figure 1a).[11] The influence of amino acid substitutions on the association is tested and the stoichiometry is examined by biophysical methods. We expect that the results will be helpful for a better predictability of coiled-coil interactions.

First, we synthesized a peptide array comprising 589 single-substitution variants of the GCN4 leucine-zipper sequence on cellulose membranes[12] and probed it for binding to the native GCN4 sequence (wild type (**wt**); Table 1, Figure 1b). Individual spot signal intensities of the pattern were measured and evaluated as described.[13,14] The replacement variability[14] of each sequence position was calculated (Figure 1c) and classified as low ($V \leq 20\,\%$), medium ($20\,\% \leq V \leq 50\,\%$), or high ($V \geq 50\,\%$). As expected, leucine at the core positions $\mathbf{d}_{\mathrm{I}}$–$\mathbf{d}_{\mathrm{IV}}$ is classified as having low variability, with the exception of Leu12 (d_{II}), which is replaceable by Ala. In contrast, core positions $\mathbf{a}_{\mathrm{I}}$–$\mathbf{a}_{\mathrm{V}}$ are of intermediate variability; synonymous residues within this class are characterized by an expanded physicochemical similarity[14] to the exchanged amino acids. The solvent-exposed positions **f**, **b**, and **c** are all classified as highly variable, whereas the pattern emerging for the core-flanking positions **e** and **g** is more complex. As the **e** positions are highly variable, we anticipated a similar behavior for the **g** positions. However, the variability of the **g** sites depends on the sequence position, ranging from low ($\mathbf{g}_{\mathrm{III}}$ = Glu22) to medium ($\mathbf{g}_{\mathrm{I}}$ = Lys8, $\mathbf{g}_{\mathrm{IV}}$ = Leu29) to high ($\mathbf{g}_{\mathrm{II}}$ = Lys15). This discrepancy between the two core-flanking

Table 1: Sequences of selected GCN4 variants.

	g abcdefg abcdefg abcdefg abcdefg ab
	I II III IV
wt[a]	RMKQLEDK VEELLSK NYHLENE VARLKKL VG
1	RMKQLEDK VEELLSK NYHLENE K ARLE KL VG
2	RMKQLEDK VEELLSK YYH T ENE VARL K KL VG
3	RMKQLEDK VEELLSK I YH N ENE VARLKKL VG

[a] The sequence corresponds to residues 249–279 of the GCN4 protein of *Saccharomyces cerevisiae*. The underlined residues highlight the substitutions. For the synthesis see S1 and S2 in the Supporting Information .

[*] Dr. M. Portwich,[$] C. C. Mahrenholz, I. Kretzschmar, Prof. Dr. A. Kramer, Dr. R. Volkmer
Institut für Medizinische Immunologie
Charité-Universitätsmedizin Berlin
Campus Charité Mitte
Schumannstrasse 20–21, 10117 Berlin (Germany)
Fax: (+49) 30-450-524942
E-mail: rve@charite.de
Homepage: http://www.charite.de/immunologie/research/agsm/index.html
Dr. S. Keller,[$] Dr. H. M. Strauss[+] [$]
Leibniz Institute of Molecular Pharmacology FMP
Robert-Rössle-Strasse 10, 13125 Berlin (Germany)
[+] Present address:
Max Planck Institute of Colloids and Interfaces
Am Mühlenberg 1, 14424 Golm (Germany)
[$] These authors have contributed equally.

[**] This work was supported by the Deutsche Forschungsgemeinschaft (SFB 449 to R.V. and SFB 498 to H.S.), the Ernst Schering Research Foundation, the Jürgen Manchot Foundation, and the Universitätsklinikum Charité Berlin.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

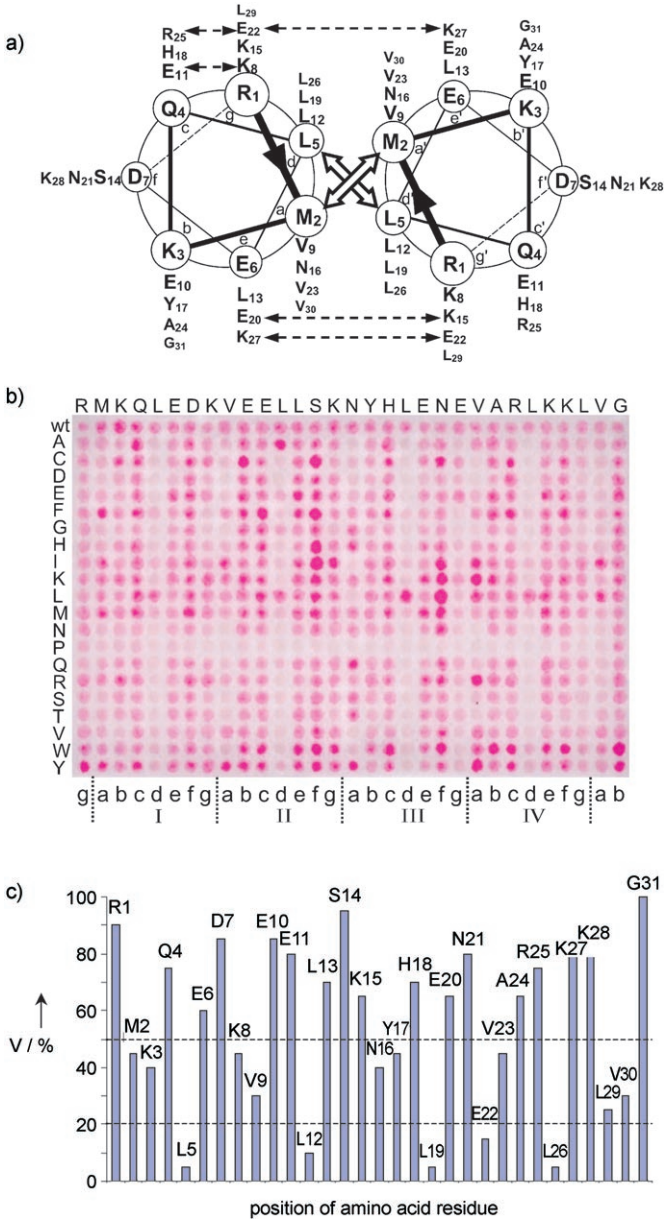

Figure 1. a) A helical-wheel diagram depicting the homodimeric leucine zipper (coiled coil) of the yeast bZIP transcription factor GCN4 as viewed from the N terminus. The residues (single-letter code for amino acids is used) next to the viewer are surrounded by circles. Crossed arrows in the center denote interactions in the hydrophobic interface termed the core and dashed arrows represent inter- and intrahelical salt bridges apparent in the structure.[11] b) Complete single-point substitution analysis of the GCN4 leucine-zipper sequence. Pink spots indicate interactions between cellulose-membrane-bound variants and a dye-labeled wild-type (**wt**) GCN4 leucine-zipper sequence that was synthesized by standard solid-phase peptide synthesis and labeled at the N terminal with tetramethylrhodamine (Tamra) (see S1 and S2 in the Supporting Information). Each spot corresponds to a variant in which one residue of the **wt** sequence given at the top was replaced by one of the 20 gene-encoded amino acids as specified on the left. Spots in the first row represent replicas of the **wt** sequence. c) Percentage of replacement variability (*V*) of each sequence position. All spot signals on the array shown in (b) were measured quantitatively and successful replacements (countable binding spots) were determined (see S4 and S5 in the Supporting Information). *V* was calculated as *V* = (number of binding spots/20) × 100[14] and plotted against the GCN4 leucine-zipper sequence.

positions has also been observed by Matthews, Vinson and coworkers.[15]

As multiple substitutions can entail a leap from dimeric to trimeric or tetrameric structures,[3a,16] we extended the peptide-array approach to an investigation of double substitutions, focusing on residues located close to and within the core (see Table S4.1 in the Supporting Information). Double substitutions at positions **a**/**d**, **a**/**e**, and **d**/**g** of the GCN4 leucine-zipper sequence resulted in a synthetic peptide array of 4320 double-substitution variants, which were then probed for binding the native GCN4 sequence wt (Figure 2 a,b). Altogether, 933 heteromeric associations were observed and are summarized in Figure 2 c. Heptad I is characterized by the highest tolerance toward substitution, whereas heptads II–IV show different substitution tolerances according to I > III > II > IV for **a**/**d**, I > III > IV > II for **a**/**e**, and I > II > III > IV for **d**/**g**. Given their classification as moderately and highly variable (Figure 1 c), it is not surprising that double substitutions at **a**/**e** also resulted in the highest number of associations. This implies that single-point substitutions at positions **a** and **e** act additively upon simultaneous exchange. Surprisingly, however, **a**/**d** and **d**/**g** substitutions were equally poorly tolerated, suggesting that, besides the well-known core positions **a** and **d**, the core-flanking position **g** plays an important role in coiled-coil formation. The particular status of position **g** implied by single-point substitutions was later confirmed by the double-substitution analysis.

To analyze if pairs of favorable or unfavorable substitutions exist, we classified amino acids into two sets: a hydrophilic set (ξ) comprising Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, and Thr; and a hydrophobic set (Φ) comprising Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val, and Tyr. We then analyzed to what extent the GCN4 heptads tolerate combinations of Φ/Φ, ξ/Φ, Φ/ξ, and ξ/ξ substitutions, as depicted by the color-coded slices in Figure 2 c. All combinations are allowed for **a**/**e** substitutions, with the exception of the second heptad, where substitutions are restricted to Φ/Φ and Φ/ξ combinations. At positions **a**/**d** and **d**/**g**, however, combinations of hydrophilic (ξ/ξ) substitutions hardly occur and, if so, only in the first heptad. Astonishingly, we observed a significant number of Φ/ξ replacements at positions **a**/**d**. Except for heptad IV, the hydrophobic amino acid at position **d** can be replaced by a hydrophilic one if the residue at position **a** is simultaneously exchanged by a hydrophobic one. Such an effect was not found for **d**/**g** substitutions. With the exception of the first heptad, only a hydrophobic amino acid could replace a **d** residue independent of the **g** replacement. Interestingly, both **a**/**d** and **d**/**g** pairs in heptad IV seem to be crucial for association.

Overall, three substitution effects were observed: First, several combinations of two previously tolerated single substitutions (Figure 1 b) were also accepted in double-substitution variants (addition effect). For example, variant **1** included a basic Lys instead of Val at position $\mathbf{a}_{IV}$ and an acidic Glu instead of Lys at position $\mathbf{e}_{IV}$ (Table 1). Second, some previously untolerated single substitutions were accepted when they were combined with an already previously tolerated substitution (transition effect). In variant **2**, for instance, a hydrophilic Thr was accepted instead of the

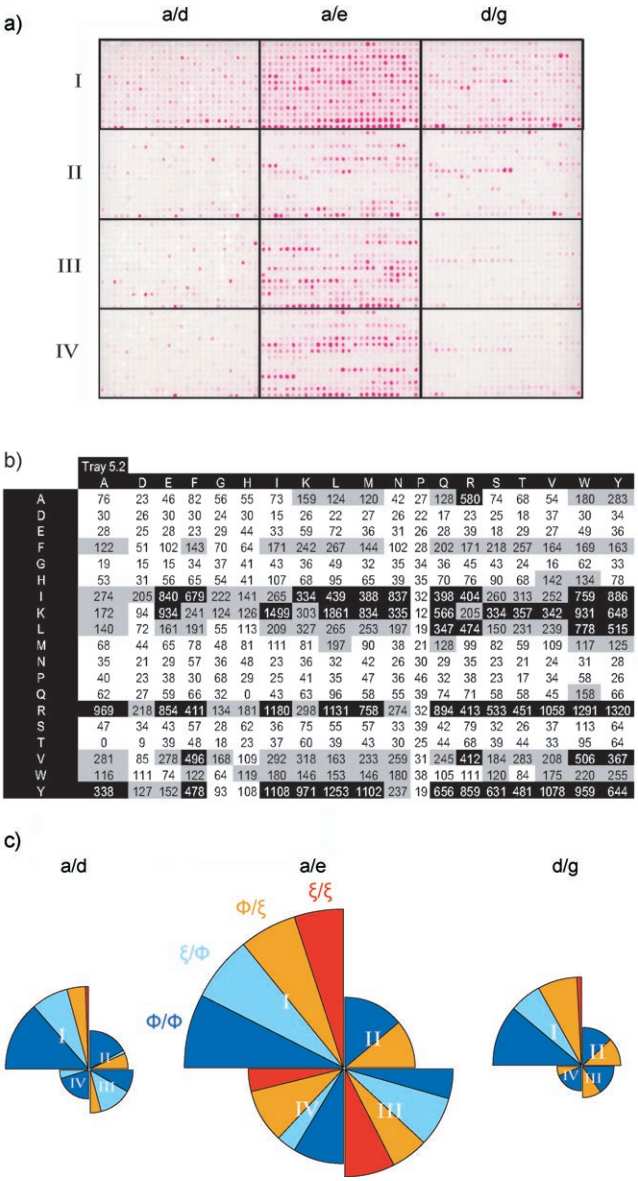

Figure 2. a) Synthetic peptide arrays displaying the complete set of GCN4 leucine-zipper sequences with double substitutions at heptad positions **a**/**d**, **a**/**e**, and **d**/**g** in heptads I–IV by using all gene-encoded amino acids except for Cys. For practical reasons, each assembly comprised 26×14 synthesis sites, thus enabling the implementation of controls. Each colored spot represents a variant that is heterospecifically associated with a wild-type peptide that is marked with Tamra on the N terminus (see S2 in the Supporting Information). b) Double substitutions of $\mathbf{a}_{IV}/\mathbf{e}_{IV}$ (V23/K27). Arrays were analyzed with a Lumi-Imager and the signal intensities were then translated into Boehringer light units (BLUs), charted, and grayscaled for better evaluation (see S4 and S5 in the Supporting Information). Rows represent the $\mathbf{a}_{IV}$ position and columns the $\mathbf{e}_{IV}$ position, both of which were scanned by using the 19 amino acids shown. Gray and black cells denote associating and strongly associating variants, respectively. See S5 in the Supporting Information for the complete set of tables. c) Pie charts standing for **a**/**d**, **a**/**e**, and **d**/**g** substitutions and summarizing heterospecific associations observed in (a). Each chart represents one type of double substitution in all four heptads, shown as quarters I–IV. The radius of a quarter correlates with the number of possible substitutions in the corresponding heptad. The surface of each slice scales with the number of substitutions classified as hydrophobic–hydrophobic (Φ/Φ; blue), hydrophilic–hydrophobic (ξ/Φ; turquoise), hydrophobic–hydrophilic (Φ/ξ; orange), or hydrophilic–hydrophilic (ξ/ξ; red).

usually highly sensitive Leu at position $\mathbf{d}_{III}$ when a Tyr was located close in space to position $\mathbf{a}_{III}$. Third, several combinations of two previously untolerated single substitutions were allowed when they were combined in the GCN4 sequence (saltation effect). For example, variant **3** included Ile at position $\mathbf{a}_{III}$ and Asn at $\mathbf{d}_{III}$, neither of which were tolerated as a single substitution (Figure 1b).

As the peptide-array experiments used were designed to detect heterospecific associations, homospecific interactions of GCN4 variants were studied by a different approach. The synthetic GCN4 variants **1**, **2**, and **3** (see S3 in the Supporting Information) were immobilized in triplicate on a cellulose membrane.[17] Each array was then probed with one of the variants and revealed a pattern of homo- and heteromeric associations (Figure 3). In contrast to **2** and **3**, **1** did not show homospecific association. Moreover, heterospecific association was observed between **2** and **3**, whereas **1** associated heterospecifically only with **3**. CD spectroscopy affirmed the coiled-coil character of the associations, which became apparent in an increase in the helicity upon mixing two associating variants (see S6 in the Supporting Information).

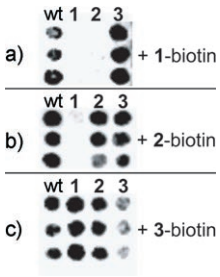

Figure 3. Array analysis of homo- and heterospecific associations of GCN4 leucine-zipper variants **1**–**3** (Table 1). Three identical arrays a)–c) were generated by immobilizing the synthetic peptides wt, **1**, **2**, and **3** on a cellulose support (see S1 and S3 in the Supporting Information). Each array contained variants **1**–**3** and the native GCN4 leucine-zipper wt as a control in triplicate. Arrays were incubated with biotinylated derivatives of **1**–**3** as indicated on the right. The associations were visualized by subsequent incubation with peroxidase-labeled streptavidin (see S2 and S4 in the Supporting Information).[17b]

Analytical ultracentrifugation was employed to characterize the homospecific association processes. Although **2** (Figure 4a) and **3** both folded into stable homotrimeric structures (termed $\mathbf{2}_3$ and $\mathbf{3}_3$, respectively), **1** was mainly monomeric (see S7 in the Supporting Information). Isothermal titration calorimetry allowed a deeper view into the heterospecific association processes. Upon mixing, $\mathbf{2}_3$ and $\mathbf{3}_3$ rearranged to form a heteromeric coiled coil with a 1:1 stoichiometry (Figure 4b and S7 in the Supporting Information). Monomeric **1** also formed a 1:1 complex when mixed with $\mathbf{3}_3$, whereas no heterospecific interaction was observed between **1** and $\mathbf{2}_3$. Ultracentrifugation demonstrated that all heteromeric com-

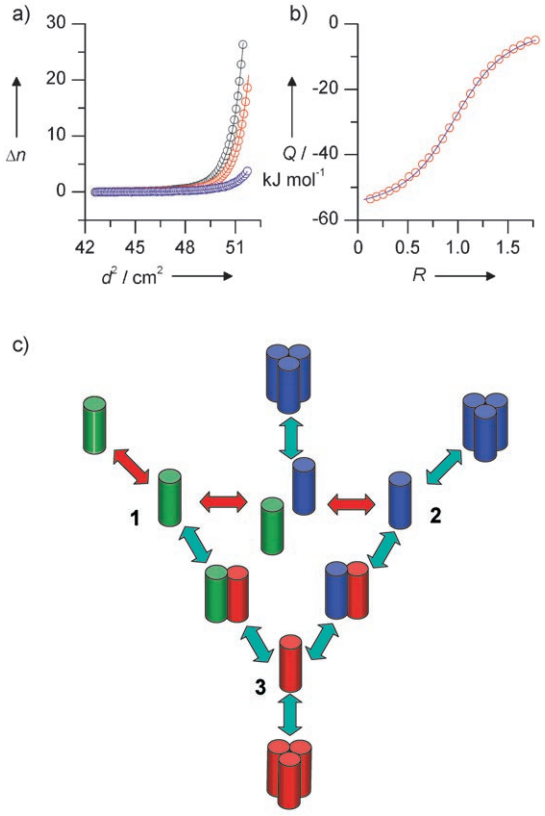

Figure 4. a) Homoassociation of variant **2** as followed by analytical ultracentrifugation. Peptide samples at concentrations of 3.0 mg mL^{-1} (black circles), 1.5 mg mL^{-1} (red circles), and 0.3 mg mL^{-1} (blue circles) were spun at 50 krpm to give the local refractive index difference, Δn, as a function of the squared distance from the rotor axis, d^2, at sedimentation and chemical equilibrium. For clarity, only every 10th data point is shown. The best fits (lines) were obtained using a nonideal monomer-dimer-trimer model (see Supporting Information S7). b) Heteroassociation of variants **2** and **3** as monitored by isothermal titration calorimetry. Aliquots (10 μL) of a 1.08 mM solution of **3** were injected into 124 μM **2** at 25 °C. The best fit (blue line) to the normalized heats of reaction, Q (red circles), yielded an association constant of 1.2×10^5 M^{-1} and a stoichiometry of 1:1. R is the molar ratio of **3**:**2**. c) Diagram of mutual associations between nonimmobilized variants **1**, **2**, and **3** as derived from the biophysical experiments shown in a) and b) (see also S7 in the Supporting Information).

plexes formed were dimers and confirmed the absence of heteroassociation between variants **1** and **2** (see S7 in the Supporting Information). Figure 4c summarizes the interactions and stoichiometries of all coiled coils formed by the three variants.

In conclusion, the presented synthetic peptide arrays can be applied to the study of coiled-coil associations, without, however, providing information about the stoichiometry and topology of the coiled coils. Our array analyses lead to a new appraisal of the heptad position **g**, whose replacement sensitivity was found to be significantly higher than that of the formally similar position **e**. Double-point substitutions can entail addition, transition, or saltation effects, which allowed us to create an interaction network of three GCN4 leucine-zipper sequence variants including a monomer as well as two homotrimeric and two heterodimeric coiled coils. Positions $\mathbf{a}_{III}$ and $\mathbf{d}_{III}$ act as a switch for the transition from the native homodimeric coiled coil to homotrimeric structures. Hence, the native function of the GCN4 leucine zipper, that is, the formation of a homodimeric protein, can be fundamentally altered by only two substitutions. Conceivably, this supramolecular flexibility of the coiled-coil motif is the quality that makes it such a versatile actor in nature.

Received: August 9, 2006
Published online: January 9, 2007

Keywords: helical structures · peptide arrays · peptides · protein structures · spot synthesis

[1] a) J. M. Mason, K. M. Arndt, *ChemBioChem* **2004**, *5*, 170; b) A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37.
[2] F. H. S. Crick, *Acta Crystallogr.* **1953**, *6*, 689.
[3] a) P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **1993**, *262*, 1401; b) P. B. Harbury, P. S. Kim, T. Alber, *Nature* **1994**, *371*, 80; c) D. L. Akey, V. N. Malashkevich, P. S. Kim, *Biochemistry* **2001**, *40*, 6352; d) B. Tripet, K. Wagschal, P. Lavigne, C. T. Mant, R. S. Hodges, *J. Mol. Biol.* **2000**, *300*, 377.
[4] a) N. E. Zhou, C. M. Kay, R. S. Hodges, *J. Mol. Biol.* **1994**, *237*, 500; b) D. Krylov, I. Mikhailenko, C. Vinson, *EMBO J.* **1994**, *13*, 2849; c) W. D. Kohn, C. M. Kay R. S. Hodges, *J. Mol. Biol.* **1998**, *283*, 993; d) R. A. Kammerer, D. Kostrewa, P. Progias, S. Honnappa, D. Avila, A. Lustig, F. K. Winkler, J. Pieters, M. O. Steinmetz, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13891.
[5] E. K. O'Shea, K. J. Lumb, P. S. Kim, *Curr. Biol.* **1993**, *3*, 658.
[6] a) Multicoil: E. Wolf, P. S. Kim, B. Berger, *Protein Sci.* **1997**, *6*, 1179; b) Socket: J. Walshaw, D. N. Woolfson, *J. Mol. Biol.* **2001**, *307*, 1427; c) Coils: A. Lupas, M. van Dyke, J. Stock, *Science* **1991**, *252*, 1162; d) Paircoil: B. Berger, D. B. Wilson, E. Wolf, T. Tonchev, M. Milla, P. S. Kim, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8259; e) J. M. Mason, M. A. Schmitz, K. M. Müller, K. M. Arndt, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8989.
[7] J. C. Hu, E. K. O'Shea, P. S. Kim, R. T. Sauer, *Science* **1990**, *250*, 1400.
[8] J. N. Pelletier, K. M. Arndt, A. Plückthun, S. W. Michnick, *Nat. Biotechnol.* **1999**, *17*, 683.
[9] J. R. Newman, A. E. Keating, *Science* **2003**, *300*, 2097.
[10] R. Frank, *J. Immunol. Methods* **2002**, *267*, 13.
[11] E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Albert, *Science* **1991**, *254*, 539.
[12] For details about the synthesis and purity of spot-synthesized cellulose-membrane-bound peptide arrays please refer to S1 in the Supporting Information.
[13] A. A. Weiser, M. Or-Guil, V. Tapia, A. Leichsenring, J. Schuchhardt, C. Frömmel, R. Volkmer-Engert, *Anal. Biochem.* **2005**, *342*, 300.
[14] J. Przdzezdiak, S. Tremmel, I. Kretzschmar, M. Beyermann, M. Bienert, R. Volkmer-Engert, *ChemBioChem* **2006**, *7*, 780.
[15] a) B. Ibarra-Molero, J. A. Zitzewitz, C. R. Matthews, *J. Mol. Biol.* **2004**, *336*, 989; b) C. Vinson, A. Acharya, E. J. Taparowsky, *Biochim. Biophys. Acta* **2006**, *1759*, 4.
[16] X. Zeng, H. Zhu, H. A. Lashuel, J. C. Hu, *Protein Sci.* **1997**, *6*, 2218.
[17] H. K. Rau, N. DeJonge, W. Haehnel, *Angew. Chem.* **2000**, *112*, 256; *Angew. Chem. Int. Ed.* **2000**, *39*, 250.

