Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design

Jody M. Mason, Mark A. Schmitz, Kristian M. Müller, and Katja M. Arndt*

Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved April 18, 2006 (received for review November 14, 2005)

Activator protein-1 (AP-1) is a crucial transcription factor implicated in numerous cancers. For this reason, nine homologues of the AP-1 leucine zipper region have been characterized: Fos (c-Fos, FosB, Fra1, and Fra2), Jun (c-Jun, JunB, and JunD), and semirational library-designed winning peptides FosW and JunW. The latter two were designed to specifically target c-Fos or c-Jun. They have been identified by using protein-fragment complementation assays combined with growth competition. This assay removes nonspecific, unstable, and protease susceptible library members from the pool, leaving winners with excellent drug potential. Thermal melts of all 45 possible dimeric interactions have been surveyed, with the FosW–c-Jun complex displaying a melting temperature (T_m) of 63°C, compared to only 16°C for wild-type c-Fos–c-Jun interaction. This impressive 70,000-fold K_D decrease is largely due to optimized core packing, α-helical propensity, and electrostatics. Contrastingly, due to a poor c-Fos core, c-Fos–JunW dimerizes with lower affinity. However the T_m far exceeds wild-type c-Fos–c-Jun and averaged JunW and c-Fos, indicating a preference over either homodimer. Finally, and with wider implications, we have compiled a method for predicting interaction of parallel, dimeric coiled coils, using our T_m data as a training set, and applying it to 59 bZIP proteins previously reported. Our algorithm, unlike others to date, accounts for helix propensity, which is found to be integral in coiled coil stability. Indeed, in applying the algorithm to these 59^2 bZIP interactions, we were able to correctly identify 92% of all strong interactions and 92% of all noninteracting pairs.

bioinformatics | protein design | protein stability | protein–protein interaction | dominant negative

The dimeric transcription factor activator protein-1 (AP-1) comprises Jun, Fos, activating transcription factor, and musculoaponeurotic fibrosarcoma families. Chief mammalian cell AP-1 constituents, Jun and Fos, contain a transactivation domain, a basic region for recognizing a DNA consensus sequence and a leucine zipper [coiled coil (CC)] region (see Fig. 1). The latter, in dimerizing, permits the two basic domains to bind to their consensus sequence. Such transcription factors, known as basic-zipper or bZIP proteins, are found at the closing stages of mitogen-activated protein kinase signaling cascades (e.g., the RAS pathway). AP-1 is implicated in various cancers where it can become up-regulated or overexpressed (1), and has been shown to be important in cell growth initiation, with c-Jun and c-Fos identified as cellular counterparts to viral oncoproteins v-Jun and v-Fos, thus establishing their role in tumorigenesis (2). Indeed, they are central in numerous oncogenic pathways and could therefore be prime candidates in anticancer drug design. However, AP-1 can also have antiproliferative properties, depending on subunit composition, transcription level, posttranslational modification (e.g., phosphorylation), and interaction with other proteins (e.g., Jun N-terminal kinase); this is manifested by the preponderance of different AP-1 family members in different types of tissue cancer. For example, c-Jun is central in skin and liver tumors, whereas JunB and JunD have very poor transactivation domains, weak transforming activities, and may have an alternative role to play. Fra1 and Fra2 have weak transactivation domains, but are found in lung and epithelial tumors, possibly by dimerizing with other family members

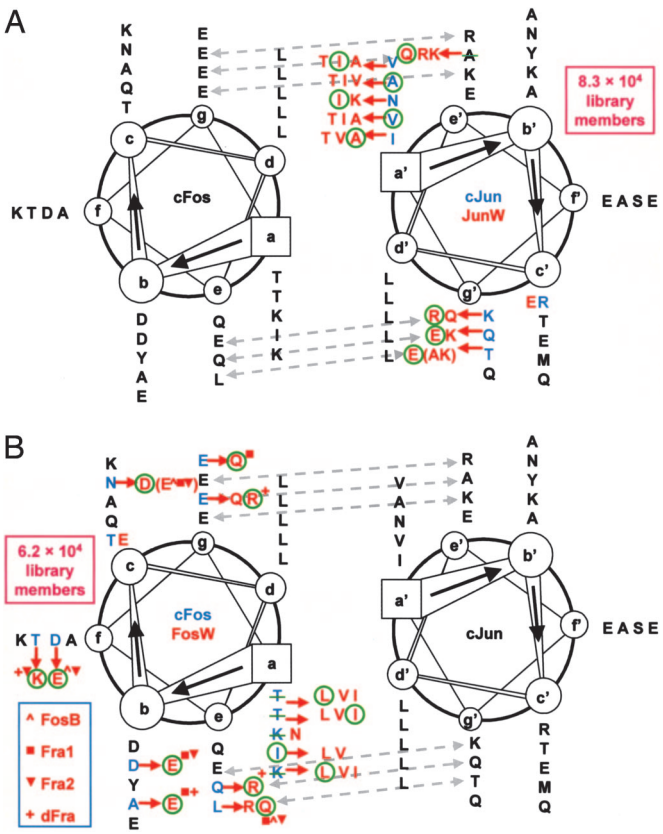

Fig. 1. Helical wheel representations of the AP-1 parallel CC. Shown are Jun (*A*) and Fos (*B*) libraries. Many changes reside in the core a positions (all *L* **d** positions were left untouched). Alternative residue options were included at **e** and **g** positions, and were aimed at varying electrostatic attraction levels at the dimeric interface. Other alternative residues, largely taken from homologues (see key), were introduced to the library pool at the solvent exposed **b**, **c**, and **f** positions. Wild-type residues removed from libraries are struck through in green, newly introduced amino acids are marked red, and wild-type positions left in the library are marked blue. Library winner selections are circled green. Position **c1** E (marked red) is for additional N-cap stability, whereas **b3** Y assists with concentration determination. Capping motifs have also been added.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AP-1, activator protein-1; CC, coiled coil.

*To whom correspondence should be addressed. E-mail: katja@biologie.uni-freiburg.de.

possessing intact transactivation domains (2). In inhibiting inappropriate AP-1 formation, or permitting correct AP-1 pairings by sequestering potential partners, specific AP-1 dimers can be targeted. In our opinion, the most efficient as yet untested way to do this is via the dimerization driving CC domain.

Found within 3–5% of all encoded amino acids (3), CCs are, despite their versatility, highly specific. In bZIP proteins, the CC consists of two parallel α-helices that wrap around one another in a left-handed supercoil. Characterized by a repeat of seven amino acids, denoted **a-g**, residues **a** and **d** consist largely of hydrophobic residues, forming a stripe which associates with respective partners on the other helix. Core flanking charged residues at **e** and **g** positions form interhelical ion pairs with **g′** and **e′** residues in the neighboring helix. Core region proximity means these residues are partially shielded from the solvent (4). For recent reviews of CC structures in general see refs. 5–7.

In this work, we describe the generation of high-affinity peptides able to bind their intended target (with higher affinity than wild-type). The beauty of this semirational approach lies in adding both intuitive and unintuitive library members, from which truly unique inhibitor sequences can arise. The possibility of broader interpretations has led us to include data sets taken from homologue variations. Diverse changes permit more universal conclusions, whereas differences in binding strengths and stabilities can also permit an understanding of how partners constituting AP-1 exert their effects. This simplification [differences in DNA binding domains, modifications (such as phosphorylation), and concentrations *in vivo* also play major roles] provides useful information regarding CC roles in conferring stability and specificity. Despite this, understanding CC contribution to stability should paint a clearer picture in the tumorigenic prevalence of particular AP-1 pairings.

Fos and Jun family based libraries have been designed with optimized residues crucial for dimerization and stability using numerous potential residues options (including most wild-type residues). Importantly, unintuitive residue selections, arising from retained wild-type amino acids and those appearing to contribute poorly to overall stability (from homologues), were included as well. Often far from the core, these can fulfill poorly understood roles in intramolecular interactions, helical propensity, and solubility, generating improved overall stability. Interacting helices from gene libraries were assayed by using a protein-fragment complementation assay (8, 9), generating soluble, nonaggregating, protease-resistant, stable inhibitors, binding their targets with maximum efficacy. Sequestering one half of AP-1 with such potency should strongly and indefinitely inhibit transcription of the target gene. Changes in the helix sequences will also consent improved understanding of CC stability, progressing beyond "Peptide Velcro" simplicity (10, 11).

Results

Targeting c-Fos: Design of Jun Library and Selection of JunW. Despite being a reasonable CC, with good core and electrostatic properties, the Jun interaction pattern with Fos is still taxing to understand. Consequently, in library design, wild-type core **a** residues have been retained, with similar options introduced (Fig. 1*A*). Specifically, three β-branched residues (I, V, T), and A at **a1**, **a2**, **a4**, and **a5**, were added (12, 13). Mutation of **a1** to L has previously yielded remarkably increased stability for a dominant negative Jun–Fos (14). However, in the interest of library size, and additional undesirable library options, L was excluded. Interestingly, of these four varied **a** positions, A arose twice in the selection, at **a1** and **a4**, V at **a2** was unchanged, whereas **a5** switched from V to I. Selection of A at positions **a1** and **a4** were late in growth competitions, indicating low preference. Indeed, the JunW core is little improved over parent molecule or homologues.

The N–K **a3** pair observed in the Jun-Fos heterodimer was kept, along with I and K options in the library. Despite instability incurred from N–N and K–K pairings, hydrogen bonding plays a role in specificity determination (15). This has been disputed on the grounds that both **a–g′** (parallel) and **a–e′** (antiparallel) K–E interactions can occur, with K side chain internal methylene groups maintaining core integrity (16). Additionally, K–I pairs are stable compared to N–K, although specificity conferred by such pairings are unknown. Regardless, the option is a good test-bed within our experimental confines, as either c-Fos–c-Fos or JunW–JunW homodimers forming preferentially over c-Fos–JunW will slow bacterial growth and be removed from the pool. Interestingly, in the selection process N and K (opposite c-Fos K) were rejected from **a3** in favor of I.

In parallel dimeric CCs, a $\mathbf{g_i}$ residue can form a coulombic interaction with an $\mathbf{e'_{i+1}}$ residue of the next heptad on the opposite helix (6). In the JunW, the **e3** position was changed from A to Q with both R and K rejected (Fig. 1*A*). All replacements were predicted to pair well with **g2** E in c-Fos (17). All c-Jun **g** positions have been retained in the library, with alternatives aiming for equivalent or better electrostatic interactions with c-Fos **e** residues. However, no wild-type residues were selected and, excepting the **c1** N-cap motif, no other changes were made to the molecule owing to library size and a lack of suitable homologue variations.

Positions **g1**, **g2**, and **a2**, settled fastest, and are perhaps pertinent in heterodimer formation or stability. **g1** R was predicted to pair well with c-Fos **e2** E while shielding the core better than alternatives Q or K. Both **a1** and **a4** fluctuated between V and A, with the latter selected last, suggesting no major preference between these residues at this position.

Targeting c-Jun: Fos Library Design and Selection of FosW. In core design, Fos has much scope for improvement comparative to Jun. Consequently, FosW–c-Jun will always be more stable than a corresponding JunW–c-Fos. The instability of c-Fos (it cannot homodimerize) is the principal heterodimeric driving force for c-Jun–c-Fos formation, rather than heterodimeric preference over c-Jun–c-Jun homodimers (18). Repulsive **g**/**e′** interactions and poor core **a** residues largely explain this phenomenon (Fig. 1*B*). For example, K in the Fos core (and E at **g1**/**e2**) helps prevent Fos homodimerization (19). Two **a** position K → Nor-Leu exchanges indeed renders c-Fos a stable homodimer (20). **a2** T → I forms homodimers (21), with homo- and heterodimer stability increasing in line with additional I content. An exception is **a1** T → I, which destabilizes homodimers. Consequently, Fos library wild-type **a** positions were removed and replaced with β- and γ-branched options (L, V, I), for better packing and desolvation (Fig. 1*B*). In selection, L or I, but not V arose in all four instances. Position **a3**, opposing c-Jun N, was fixed from K to N; this was predicted to be the best pairing for specificity and to be favored over aliphatic–polar combinations (12).

$\mathbf{g/e'_{i+1}}$ interacting pairs, which complement c-Jun, were introduced according to Vinson's free energy values (17). **g1** and **e2** E residues, conserved in Fos homologues, were proposed to be central in accounting for most additional heterodimerization free energy with c-Jun (22), and were not varied in the library. Variations from Fos homologues were included in library design. In winning **e** and **g** positions, no wild-type residues were selected. Other options such as **e3** R were reasoned to form $\mathbf{g/e'_{i+1}}$ interactions with **g2** Q of c-Jun. Some changes (all from homologue variations, see Fig. 1*B*) are difficult to rationalize because they deviate from classical charged/polar pairings. For example, in FosW **g2**, R pairs with c-Jun **e3** A. Other changes, often subtle (e.g., **f2** D → E, found in FosB and Fra2; **c4** N → E, found in FosB, Fra2 and dFra, the latter being D in our library for coding reasons), appear to play no direct interfacial role. Strikingly, all homologue residues were selected in FosW, and collectively these changes introduce a higher extent of polar residues, with little overall change in pI.

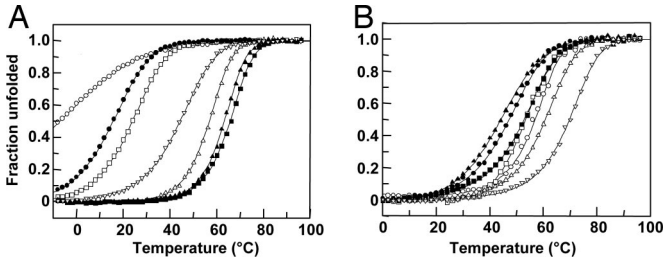

Fig. 2. Thermal stability of wild-type and winner peptide pairs measured by using temperature dependence of the CD signal at 222 nm. (*A*) Stability curves for c-Fos–c-Fos (open circles; $T_m = -1$°C), c-Jun–c-Jun (open squares; 24°C), c-Fos–c-Jun (filled circles; 16°C); FosW–FosW homodimer (open triangles; 57°C), JunW–JunW (filled squares; 66°C), FosW–c-Jun (filled triangles; 63°C) and c-Fos–JunW (open inverted triangles; 44°C). For both winners, the stability is greater than the average of the two homodimers, indicating that formation is real and not representative of two homodimers unfolding concomitantly. (*B*) Stability curves for c-Jun–JunW (open circles; 57°C), JunB–JunW (filled circles; 47°C), JunD–JunW (open squares; 53°C), FosB–JunW (filled triangles; 44°C), JunB–FosW (filled squares; 53°C), JunD–FosW (open triangles; 61°C), and JunW-FosW (inverted open triangles; 70°C).

I at **a4** settled fastest from all positions and could be the most extended option for packing against a small c-Jun **a4** A side chain, with V too short, and L's extra steric bulk less favored.

Thermal Melts of Winner and Homologues. CD spectrums revealed all dimers to be α-helical (data not shown), whereas 45 helical melts (Figs. 2 and 3) with T_m values, ranging from below −10°C to 70°C, demonstrate assay success in selecting those residues that stabilize the desired species to maximal effect. Indeed, excepting the nonnative FosW–JunW, FosW–c-Jun displays the highest T_m, indicating specificity in addition to stability. FosW–c-Jun (T_m = 63°C) was significantly more stable than wild-type (T_m = 16°C), with a ΔT_m of 23°C relative to FosW and c-Jun, corresponding to a $\Delta\Delta G_{U\rightarrow F(20°C)}$ of ≈3.4 kcal/mol (Figs. 2*A* and 3). In contrast, c-Fos–JunW displays lower stability (T_m = 44°C), but nevertheless clearly exceeds wild-type c-Jun–c-Fos as well as the average of c-Fos or JunW homodimers (ΔT_m = 11.5°C; Figs. 2*A* and 3). In general, homodimers of Jun (except c-Jun), and to a greater extent Fos, display lower stabilities compared to heterodimers. Wild-type heterodimers display modest stability, with FosW–c-Jun and c-Fos–JunW T_ms exceeding that of wild-type complexes.

Rationalization by Core, Electrostatic, and Propensity Ranking. Relating stability changes to peptide combinations is a daunting task involving a plethora of factors. Core changes affect hydrophobic burial, propensity, solubility, electrostatic attraction of flanking residues, and a number of intramolecular interactions. Ranking contributions, and building a relationship between residue changes and T_m, is consequently very difficult. However, three overriding factors (core, electrostatic and propensity) have been considered. Accordingly, we have devised an improved algorithm for CC prediction using core, electrostatic, as well as propensity. This algorithm can be found online at www.molbiotech.uni-freiburg.de/bCIPA and is known as the bZIP coiled coil interaction prediction algorithm (bCPIPA).

	cJun	JunB	JunD	cFos	FosB	Fra1	Fra2	JunW	FosW
cJun	24	-21	11	16	13	16	17	57	63
JunB		-28	-18	5	0	3	-3	47	53
JunD			-22	8	7	21	2	53	61
cFos				-1	-9	-9	-8	44	61
FosB					-6	-2	1	44	52*
Fra1						-6	-11	49*	54*
Fra2							19	43	56*
JunW								66	70
FosW									57

<75°C <50°C <25°C <0°C

Fig. 3. Forty-five interactions between Jun and Fos leucine zippers, and selected peptides. T_m values in °C and matching shading. Only data with a $T_m \geq$ 20°C gave a lower baseline reasonable enough to fit without restraint of lower baseline or slope. Temperatures marked with an asterisk show partial three-state denaturatation profiles; in these cases the two-state model is applied to the second (major) transition.

A rudimentary core packing score has been assigned to all dimers to distinguish cores which make large contributions to stability from those which do not (Fig. 3) by scoring hydrophobic pairings highest, with **aa**′ and **dd**′ pairs treated the same for simplicity. LL (LL = −1.5) is ranked higher than all others owing to its exclusivity in **d** positions. Other hydrophobic pairings (VV≈II≈LV≈LI≈VI = −1) as well as KI and NN combinations (KI≈NN = −1) were favored over other lysine and alanine combinations (KL≈KV≈LA≈VA≈IA≈IT≈LT≈AA = −0.5) with a further subset being disfavored entirely (IN≈LN≈VN≈TV = +0.5). These rankings are based on ΔΔG energies from **aa**′ pairings relative to AA pairs (23), observance in other CC proteins, and our own unpublished results. No preference was given for homotypic over heterotypic hydrophobic core pairing. KI has been described to be more stable than KV and KL, for reasons that are not yet understood (23), and also KI was selected twice in c-Fos–JunW. Finally, NN was favored significantly over KK, possibly because of increased steric restraint and charge repulsion in the latter. Stability offered by NN pairing is relatively low. N has low propensity and high polarity for a core region, but confers specificity by limiting oligomeric states to dimers (12), this benefit outweighing lack of stability.

Our electrostatic parameters are based on opposing charge pairings and place energetic penalties on similar charge pairings (DD≈DE≈EE≈RR≈KK≈RK = +1; KD≈RD≈EQ = −0.5; KQ≈RQ = −1; QQ≈KE≈RE = −1.5) with $\mathbf{g}_i/\mathbf{e}'_{i+1}$ and $\mathbf{e}_{i+1}/\mathbf{g}_i'$ interactions treated the same for simplicity. Electrostatic interactions were related to free energy contributions based on data from a double mutant analysis (17). Consequently, although not included implicitly in the study, the scale indicates much improved electrostatic attractions.

Fong *et al.* (24) used "base optimized weighting" to predict CC interactions, and identified strong interactions based on $\mathbf{d}_i\mathbf{d}'_i$, $\mathbf{a}_i\mathbf{a}'_i$, $\mathbf{a}_i\mathbf{d}'_i$, $\mathbf{d}_i\mathbf{a}'_{i+1}$ $\mathbf{d}_i\mathbf{e}'_i$, $\mathbf{g}_i\mathbf{a}'_{i+1}$, and $\mathbf{g}_i\mathbf{e}'_{i+1}$ pairings (24), but did not consider α-helical stability as a direct contributing factor. However, we estimate helical propensity to be hugely important and a largely overlooked third parameter in CC stability (covered in depth in *Discussion*), precluding electrostatic and core considerations in forming a structure which is in a dimerization competent state. Surprisingly, we find that only two of the seven considerations made by Fong *et al.* (24) are strictly necessary (**ad**′, **da**′, **de**′, **ga**′ pairings are not required, and **aa**′ and **dd**′ count as one), and that propensity is a more important omission. Indeed, the role of surface residues has been probed for GCN4-p1, with helix propensity found to be a key factor in surface design (25). Additionally, intramolecular hydrogen bonding of high propensity residues such as Q, R, E, or K, which frequent these positions is also important, and if unsatisfied, can cause unfavorable effects. In combining these parameters with a least squares fit, stability can be rationalized and agrees well with actual T_m values (Fig. 4*A*). It should be noted that our algorithm fits T_m values that indicate pairing stabilities. For specificities (ΔT_m), the propensity term cancels out (see also supporting information, which is published on the PNAS web site).

In our fitting procedure, core pairings (**d**/**d**′ = **a**/**a**′), **g**/**e**′ electrostatic preferences and propensity scales from Williams *et al.* (26) were used to fit our T_m data (45 dimers) as well as selected and rationally designed peptides characterized previ-

BIOCHEMISTRY

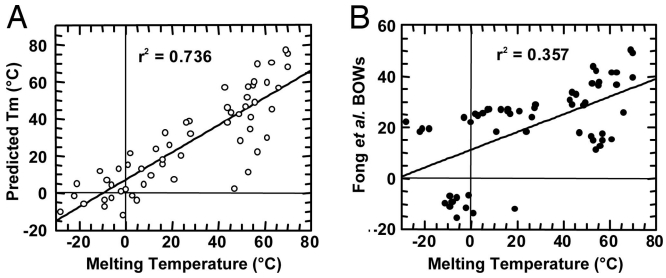

Fig. 4. Predictions of T_m based on 57 measured dimers used in our training set. Our fitting procedure (*A*) and the ''base optimized weights'' (BOWs) fitting procedure (24) (*B*) are plotted against measured T_m values. The r^2 of 0.736 reflects an improvement of our fit compared to the r^2 of 0.357 from the base optimized weights fitting procedure.

ously (12 dimers) (9, 11). A least squares fit to these 57 dimers (see Eq. **2**) yielded the coefficients a1 = 189.8, a2 = −11.8, a3 = −4.3, d = −299.6. The difference in magnitude results from the different scales used to score helix propensity, core, and electrostatics. Considering these, all three terms lie in the same range (see also supporting information). Our fit (www.molbiotech.uni-freiburg.de/bCIPA) yielded an r^2 of 0.74 (Fig. 4*A*) compared to 0.36 or 0.39 using base optimized weights (BOWs) or BOWs combined with data from human bZIPS, when using the Fong *et al.* (24) bZIP scoring form (Fig. 4*B* and http://compbio.cs.princeton.edu/bzip).

Having obtained these fits, 59^2 bZIP interactions (which contained none of the data within our training set) probed in a microarray analysis by Newman and Keating (27) were used to test the ability of our algorithm in correctly predicting strong and noninteractions. Using similar criteria as described by Fong *et al.* (24), we were able to correctly identify 92% of noninteractors and 92% of strong interactors, although only **e**/**g**, **a**/**a**, and **d**/**d** pairings and very basic weightings scales were used. Predictions by Fong *et al.* (24) reported similar values of 89% and 83%.

Discussion

Screening of *in vivo* peptide libraries, as well as binding studies to elucidate affinity for homologues, has given valuable insight into the mechanism of CC stability. Ranking the affinities of all possible AP-1 homologue pairings may also be relevant in the preponderance of oncogenic pairings *in vivo*.

Dominant Negatives (DN) Directed Against AP-1. The concept of DN to compete with wild-type bZIP or helix–loop–helix bZIP (HLH–bZIP) is not new. A peptide previously designed to bind the c-Jun CC, preventing functional c-Jun–c-Fos and c-Jun–c-Jun, had been constructed (14). Acidic, E-rich extended CCs, designed as DNA mimics, were able to bind and even sequester DNA bound AP-1 (28), a task requiring impressive K_D values in the pM range. A c-Jun transactivation domain deletion mutant, TAM67, is able to arrest AP-1 activity in normal and malignant breast cells (29). Further elucidation of rules governing CC stability and specificity will aid in potential drug development.

Stability of AP-1. AP-1 T_m variations depend on numerous factors. Previous experiments have been disulfide bridged (30, 31), had basic domains (28), been of varying length (27), and been conducted under various conditions (ionic strength, pH, etc.). Boysen *et al.* (30) 38-residue bridged CCs display T_ms of 59, 69, and 71°C for c-Fos–c-Fos, c-Jun–c-Jun, and c-Fos–c-Jun, respectively. Olive *et al.* (28) report 25, 30, and 50°C, for the same complexes unbridged but with acidic extensions, whereas O'Shea *et al.'s* (31) bridged peptides have T_ms of 30, 41, and 51°C. There are only some early qualitative studies that document existence of alternative Jun–Fos homologue dimers beyond c-Jun–c-Fos (32–36).

In the only comprehensive, semiquantitative, study to date (27), Jun–Fos homologues range from 84–107 residues, contain terminal subcloning sequences, a his-tag, an N-terminal basic region extension, and at least nine residues of C-terminal extension until the PAIRCOIL program (37) predicts the probability of CC to be <10%. In contrast, our peptides are all 37-mers of same register from the CC region, N- and C-capped, and are not disulfide bridged. Despite these differences, the core CC sequences (excepting a **b** position Y for absorbance) are identical in both studies, and display moderate agreement with our own T_m values. Using *Z* scoring (27) as an affinity measure, that earlier study showed in a microarray analysis heterodimers to yield a good interaction ($Z > 10$). Homodimeric Fos typically displayed low stabilities, and homodimeric Jun combinations (excepting c-Jun; $Z > 5$, T_m 40°C) were less stable than heterodimers. We found c-Jun–c-Jun to be more stable than c-Jun–c-Fos; however, our data do not dispute that c-Fos instability drives heterodimeric preference (18). Regardless, excluding electrostatic considerations, c-Jun–c-Jun has better α-helical propensity, and a significantly more stable core than c-Fos–c-Jun. Importantly, to our knowledge, our study is the first quantitative biochemical analysis of all possible Jun-Fos CC combinations from human (see Fig. 3).

Core Conclusions from Winners. Core comparison indicates that dimeric Fos homologue cores are disfavored compared to Fos–Jun, with Jun cores displaying greater hydrophobic burial and greater stability. Winners have similar (JunW) or enhanced (FosW) cores compared to homologues, resulting in optimized cores and improved stabilities.

In general, I is favored over V in winner **a** positions, has higher propensity, and is bulkier for core packing. It has been reported to be more stable and better than V and L at this position in conferring dimers (12, 13). FosW L and I were selected over V in all four instances, despite a documented β- over γ-branching preference (12, 38). V alone yields trimers (39) or a mixture of dimers and trimers (12), whereas I can specify dimers exclusively.

Surprisingly, no homotypic pairing preference was observed in FosW–c-Jun, with **aa′** LI, IV, IA, and LV selected. Nor were any **a1** or **a2** TT pairings observed despite speculation that generally homotypic pairings are energetically favored over similar hydrophobic pairings (23). In contrast, the bulk of core winner **aa′** pairings are heterotypic.

g/e′$_{i+1}$ Conclusions from Winners. Abundance of strong interacting pairs involving (with the exception of Q) terminal charge attractions such as KE, RE, QE, QQ, RQ, and KQ, suggest that hydrophobic bulk plays an additional role (40, 41). The side chain of E ($-(CH_2)_2COO^-$) can pair with K ($-(CH_2)_4$–NH_3^+) and R ($-(CH_2)_3$–NH–C(NH)(NH_3^+)) side chains, both having positively charged termini able to contact the negative carboxyl group of E. Q is also favorable, possibly because its side chain ($-(CH_2)_2CONH_2$) is of sufficient length to shield the core from the solvent. However, lack of terminal charge is predicted to lower the specificity of the interaction. Partial hydrophobic environment felt by the side chains may also improve the energetic contribution of these charge interactions, yielding a greater contribution in less aqueous surroundings.

However, from frequency in CCs (40, 41) and energetic rankings (17), we conclude that D and N (both of which are shorter, containing only one side chain methylene group) are disfavored and should be omitted in designed coils, despite possible charge complementarity (for D) with K or R again due to poor core shielding.

The extra length and polarity of the JunW **e3** A → Q change is likely to enable contact with the corresponding E at c-Fos **g2**.

Additionally, the extra hydrophobic bulk of Q's β- and γ-carbon methylene groups can stabilize the molecule by shielding the core from the solvent to a greater extent than can A (4), and these methylene groups may provide favorable interactions with the **d3** δ-carbon methylene groups of c-Fos L. Electrostatic attractions estimated according to Krylov *et al.* (17) in c-Fos–JunW ($\Delta\Delta G = -5.2$ kcal/mol), are improved compared to the c-Fos homodimer (−0.6 kcal/mol) and surpass the average of c-Fos and JunW homodimers (−2.4 kcal/mol), indicating an electrostatic preference for heterodimer formation.

Outer Positions: Intrahelical Stability and Solubility. Incorporating Fos homologue residue changes at solvent exposed regions interestingly resulted in acceptance of all proposed amino acids at all nine positions in the library winner (Fig. 1). Depending on the scale used, either five (26) or six (42) of these changes were found to have a higher helical propensity than wild-type c-Fos. FosW A → E change at position **b4** could be interacting favorably with Q at **e4** and K at **f4**, with the resulting increase in helix stability propagating to stabilize the CC. Other solvent exposed changes such as T → K, N → D, and D → E, may act by contributing to increased helical propensity, with charges increasing protein solubility, and aiding the driving force for folding by increasing the concentration of monomers in a dimerization competent state.

Helical Propensity Considerations. Propensity scales inform upon the frequency or preference with which a given residue occurs in a particular conformation. In our analysis we have used the scales devised by Williams *et al.* (26) as well as Gromiha and Parry (42). From these scales, averaged helical propensity predictions were assigned to each of the helices, discounting N- and C-caps. Williams *et al.* is similar to the Chou and Fasman scale (43) in that it is derived from statistical data, whereas Gromiha and Parrys is derived specifically from CCs, with the former scale giving a mildly better fit to our data set. We favored these scales over other experimental approaches (44–46) because they include not only substituted solvent exposed residues at the center of a helix, but partially and completely buried residues together with residues at the helix termini, both of which would differ in propensity in these contexts (47, 48). This finding is of particular significance in short helical CC motifs such as ours, where residues are completely buried, partially buried, or completely solvent exposed, depending on side chain and heptad position, or centrally or terminally located. Analysis of the homologues predicts winning peptides will display increased helicity. This is somewhat surprising given that it was not a criterion in library design. However, homologues are informative in this respect because they contain higher proportions of the destabilizing residue G and S compared to c-Jun and c-Fos, which is also reflected in their poor homodimeric and heterodimeric T_m values. Although G is universally accepted as a poor helical propensity residue, S is more contentious, but may account for improvement in fit to our data set when using the scale of Williams *et al.* (26). G is particularly destabilizing due to increased numbers of ϕ and ψ angles accessible to the backbone. This destabilization results in an unfavorable conformational entropy change upon helix formation. S may display poor helical propensity (26, 43) because its side chain hydroxyl donor competes for an interaction with the surrounding backbone NH and CO groups. Consistent with both helical scales used here is a similar scenario with acceptor groups at the terminus of N and D side chains. In designed helices, these residues should be avoided in favor of the longer Q (or E), which are also likely to display greater conformational entropy in the unfolded state than S, D, or N (48). Likewise, the role of electrostatic preferences clearly reaches beyond charge pairing preferences. Propensity predicts R, Q, E, and K (α-helical propensity $P_\alpha R = 1.21$, $P_\alpha Q = 1.27$, $P_\alpha E = 1.59$, $P_\alpha K = 1.23$), being favored, with both D and N ($P_\alpha D = 0.99$, $P_\alpha N = 0.76$) disfavored (26) (see also refs. 42–44). Pace and Scholtz (48) suggest that polar groups separated from the backbone by one methylene group (e.g., D) have a lower propensity than those with two (e.g., E) due to the cost of fixing an extra methylene group in random-coil state, thus favoring helix formation.

α-Helical propensity in the JunW core plays a key role additional to hydrophobic burial in selecting winning amino acids. Of five positions, V is selected only once over I or A, both of which have higher propensities for α-helix ($P_\alpha V = 0.98$, $P_\alpha I = 1.09$, $P_\alpha A = 1.41$). It is not fully understood why I has a greater propensity than V, but it has been suggested that the δ-carbon of L and I (absent in V) can increase propensity by burying nonpolar surface area against the helix (49), meaning that cores concomitantly improve helix propensity, further improving CC stability. L ($P_\alpha L = 1.34$), fixed for **d** position residues of AP-1, also contributes significantly to stability. Selection favoring higher propensity residues I and L, but not V, may reflect that our winner peptides are designed rather than native. This means that, although a designed inhibitor with the highest affinity possible is a top criterion, nature strives to evolve a balance between optimal and nonoptimal residues to suit the demands of the protein.

No specific sequences are identified that conform to a speculated "trigger sequence" (50), although propensity may play a crude role during folding, possibly acting to enforce α-helical topology, thus ensuring structures are driven thermodynamically and on a biologically realistic time scale. FosW and JunW contain no central G residues that are replaced with higher propensity residues. Closely connected, but much harder to predict, is the role of context in determining stability, or how interactions of side chains with surrounding side chains affect overall propensity.

In the future, it should be possible to factor in a greater negative design aspect where design for specificity, as well as stability, can play an increased role. More difficult will be designing to generate the lowest possible K_D, while retaining specificity so that requirements of the proteins are met. All these considerations must be accounted for as well as being incorporated into library designs, to design coils that are both stable and specific.

CC Prediction. We have rationalized our winning peptides based upon well understood principles. The algorithm of Fong *et al.* (24) falls noticeably short in predicting the T_m values for Fos–Fos homologues (Fig. 4*B*). Our own algorithm is much improved with respect to this subset. Additionally, the Fong algorithm is less able to predict T_m values of heterodimeric winner complexes, whereas ours underestimates only three winner complexes, notably c-Jun–JunW, JunD–JunW, and JunB–JunW, but makes reasonable estimates of the remainder.

We have used a combination of simplistic core and electrostatic parameters, combined with well documented but little implemented (in CC stability prediction) helical propensity scales. In combining these parameters, we have devised a prediction algorithm. Although this has further potential for optimization, it is (at least in the context of bZIPs) on par with and certainly more simplistic than that of Fong *et al.* (24). A corollary of this work is to design more robust CC pairs and dominant negatives with improved therapeutic value, and as a potential use as building blocks in nanobiotechnological design.

Materials and Methods

Library Design and Cloning. Mega-primers were synthesized including relevant degenerate codons for residue options (for libraries), and a fill-in reaction was performed, resulting in 111-bp double-stranded oligonucleotides. These were cloned via

NheI and AscI sites into a pQE16 derivative (Qiagen) containing a G/S linker tagged to fragment 1 (pAR200d; c-Jun and Jun library; ampicillin resistance; K.M.A., unpublished data) or fragment 2 (pAR300d; c-Fos and Fos library; chloramphenicol resistance; K.M.A., unpublished data) of murine dihydrofolate reductase (mDHFR), respectively. Library plasmids were transformed into BL21 gold cells (Stratagene) containing target plasmid and pREP4 (Qiagen; for lac repression). To assess library quality, we sequenced pools and single clones and found approximately equal distributions of varied amino acids. Pooled colonies exceeded the library size 5- to 10-fold.

Selection of Winner Peptides. The protein-fragment complementation assay has been described (8, 9, 11). Briefly, CCs are tagged to either half of murine dihydrofolate reductase. Only two interacting helices will bring the two halves of the enzyme into close proximity, render the enzyme active, and result in colony formation on M9 minimal medium plates with trimethoprim (1 ng/ml) to inhibit bacterial dihydrofolate reductase. Surviving colonies were pooled, grown, and serially diluted under selective conditions. Fastest growth, and hence the highest-affinity interacting partner, will dominate the pool.

Peptide Synthesis and Purification. Peptides (see supporting information for sequences) were synthesized by Protein Peptide Research and subsequently purified to >98% purity by using RP-HPLC with a Jupiter Proteo column (4-μm particle size, 90 Å pore size, 250 × 10 mm; Phenomenex) and a gradient of 5–50% acetonitrile (0.1% TFA) in 50 min at 1.5 ml/min. Correct masses were verified by electrospray mass spectrometry. Peptide concentrations were determined in water by using absorbance at 280 nm with an extinction coefficient of 1209 $M^{-1}\cdot cm^{-1}$ (51) corresponding to a Tyr residue inserted into a solvent exposed **b3** position.

CD Measurements. Spectra and thermal melts were performed at 150 μM total peptide concentration in 10 mM K-phosphate/100 mM KF (pH 7) using a Jasco J-810 CD instrument. The temperature was ramped at a rate of 0.5°C per min. Melting profiles were ≥94% reversible with equilibrium denaturation curves fitted to a two-state model to yield the melting temperature (T_m)

$$\Delta G = \Delta H - (T_A/T_m) \times (\Delta H + R \times T_m \times \ln(P_t)) + \Delta C_p \times (T_A - T_m - T_A \times \ln(T_A/T_m)), \quad [1]$$

where ΔH is the change in enthalpy, T_A is the reference temperature; R is the ideal gas constant; P_t is the total peptide concentration; and ΔC_p the change in heat capacity. Additionally, ΔT_m of a heterodimer AB is calculated by using $\Delta T_{m(AB)} = T_{m(AB)} - 0.5(T_{m(A)} + T_{m(B)})$.

Stability Prediction. Helix propensity (HP) is calculated as an average over the whole helix, i.e., the individual residues are summed and divided by the total number of residues. Electrostatics (ES) and core (C) are calculated by using a simple weighting scheme (see *Results*) and summed over the whole peptide to account for increased stability in longer helices. Scores for measured T_m values were fitted as follows

$$T_m = a1 \times HP + a2 \times C + a3 \times ES + d, \quad [2]$$

where *a1*, *a2*, and *a3* are weighting factors for the three parameters, and *d* is an offset factor. Temperatures were fitted in Kelvin (see www.molbiotech.uni-freiburg.de/bCIPA).

More detailed descriptions can be found in supporting information.

This work was funded by the Emmy Noether program of the Deutsche Forschungsgemeinschaft (Grant Ar373).

1. Shaulian, E. & Karin, M. (2002) *Nat. Cell Biol.* **4,** E131–E136.
2. Eferl, R. & Wagner, E. F. (2003) *Nat. Rev. Cancer* **3,** 859–868.
3. Wolf, E., Kim, P. S. & Berger, B. (1997) *Protein Sci.* **6,** 1179–1189.
4. Mason, J. M., Müller, K. M. & Arndt, K. M. (2006) *Methods Mol. Biol.*, in press.
5. Woolfson, D. N. (2005) *Adv. Protein Chem.* **70,** 79–112.
6. Mason, J. M. & Arndt, K. M. (2004) *Chem. Biochem.* **5,** 170–176.
7. Lupas, A. N. & Gruber, M. (2005) *Adv. Protein Chem.* **70,** 37–78.
8. Pelletier, J. N., Arndt, K. M., Plückthun, A. & Michnick, S. W. (1999) *Nat. Biotechnol.* **17,** 683–690.
9. Arndt, K. M., Pelletier, J. N., Müller, K. M., Alber, T., Michnick, S. W. & Plückthun, A. (2000) *J. Mol. Biol.* **295,** 627–639.
10. O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993) *Curr. Biol.* **3,** 658–667.
11. Arndt, K. M., Pelletier, J. N., Müller, K. M., Plückthun, A. & Alber, T. (2002) *Structure (Cambridge, U.K.)* **10,** 1235–1248.
12. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262,** 1401–1407.
13. Zhu, B. Y., Zhou, N. E., Kay, C. M. & Hodges, R. S. (1993) *Protein Sci.* **2,** 383–394.
14. Bains, N. P. S., Wilce, J. A., Heuer, K. H., Turnstall, M., Mackay, J. P., Bennet, M. R., Weiss, A. S. & King, G. F. (1997) *Lett. Pept. Sci.* **4,** 67–77.
15. Gonzalez, L., Jr., Woolfson, D. N. & Alber, T. (1996) *Nat. Struct. Biol.* **3,** 1011–1018.
16. Campbell, K. M. & Lumb, K. J. (2002) *Biochemistry* **41,** 7169–7175.
17. Krylov, D., Barchi, J. & Vinson, C. (1998) *J. Mol. Biol.* **279,** 959–972.
18. O'Shea, E. K., Rutkowski, R., Stafford, W. F., III, & Kim, P. S. (1989) *Science* **245,** 646–648.
19. Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R. P., Bravo, R. & Müller, R. (1989) *Cell* **56,** 507–516.
20. Campbell, K. M., Sholders, A. J. & Lumb, K. J. (2002) *Biochemistry* **41,** 4866–4871.
21. Porte, D., Oertel-Buchheit, P., John, M., Granger-Schnarr, M. & Schnarr, M. (1997) *Nucleic Acids Res.* **25,** 3026–3033.
22. John, M., Briand, J. P., Granger-Schnarr, M. & Schnarr, M. (1994) *J. Biol. Chem.* **269,** 16247–16253.
23. Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R. & Vinson, C. (2002) *Biochemistry* **41,** 14122–14131.
24. Fong, J. H., Keating, A. E. & Singh, M. (2004) *Genome Biol.* **5,** R11.
25. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. (1997) *Protein Sci.* **6,** 1333–1337.
26. Williams, R. W., Chang, A., Juretic, D. & Loughran, S. (1987) *Biochim. Biophys. Acta* **916,** 200–204.
27. Newman, J. R. & Keating, A. E. (2003) *Science* **300,** 2097–2101.
28. Olive, M., Krylov, D., Echlin, D. R., Gardner, K., Taparowsky, E. & Vinson, C. (1997) *J. Biol. Chem.* **272,** 18586–18594.
29. Ludes-Meyers, J. H., Liu, Y., Munoz-Medellin, D., Hilsenbeck, S. G. & Brown, P. H. (2001) *Oncogene* **20,** 2771–2780.
30. Boysen, R. I., Jong, A. J., Wilce, J. A., King, G. F. & Hearn, M. T. (2002) *J. Biol. Chem.* **277,** 23–31.
31. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992) *Cell* **68,** 699–708.
32. Nakabeppu, Y., Ryder, K. & Nathans, D. (1988) *Cell* **55,** 907–915.
33. Zerial, M., Toschi, L., Ryseck, R. P., Schuermann, M., Muller, R. & Bravo, R. (1989) *EMBO J.* **8,** 805–813.
34. Cohen, D. R., Ferreira, P. C., Gentz, R., Franza, B. R., Jr., & Curran, T. (1989) *Genes Dev.* **3,** 173–184.
35. Curran, T., Van Beveren, C. & Verma, I. M. (1985) *Mol. Cell. Biol.* **5,** 167–172.
36. Suzuki, T., Okuno, H., Yoshida, T., Endo, T., Nishina, H. & Iba, H. (1991) *Nucleic Acids Res.* **19,** 5537–5542.
37. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8259–8263.
38. Betz, S. F., Bryson, J. W. & DeGrado, W. F. (1995) *Curr. Opin. Struct. Biol.* **5,** 457–463.
39. Potekhin, S. A., Medvedkin, V. N., Kashparov, I. A. & Venyaminov, S. (1994) *Protein Eng.* **7,** 1097–1101.
40. Vinson, C. R., Hai, T. & Boyd, S. M. (1993) *Genes Dev.* **7,** 1047–1058.
41. Woolfson, D. N. & Alber, T. (1995) *Protein Sci.* **4,** 1596–1607.
42. Gromiha, M. M. & Parry, D. A. (2004) *Biophys. Chem.* **111,** 95–103.
43. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 211–222.
44. O'Neil, K. T. & DeGrado, W. F. (1990) *Science* **250,** 646–651.
45. Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994) *Protein Sci.* **3,** 843–852.
46. Litowski, J. R. & Hodges, R. S. (2002) *J. Biol. Chem.* **277,** 37272–37279.
47. Doig, A. J. & Baldwin, R. L. (1995) *Protein Sci.* **4,** 1325–1336.
48. Pace, C. N. & Scholtz, J. M. (1998) *Biophys. J.* **75,** 422–427.
49. Chakrabartty, A. & Baldwin, R. L. (1995) *Adv. Protein Chem.* **46,** 141–176.
50. Kammerer, R. A., Schulthess, T., Landwehr, R., Lustig, A., Engel, J., Aebi, U. & Steinmetz, M. O. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 13419–13424.
51. Du, H., Fuh, R. A., Li, J., Corkan, A. & Lindsey, J. S. (1998) *Photochem. Photobiol.* **68,** 141–142.