Journal of Structural Biology 186 (2014) 380–385

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Your personalized protein structure: Andrei N. Lupas fused to GCN4 adaptors

Silvia Deiss [1], Birte Hernandez Alvarez [1], Kerstin Bär, Carolin P. Ewers, Murray Coles, Reinhard Albrecht, Marcus D. Hartmann *

Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstraße 35, D-72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form 22 January 2014
Accepted 23 January 2014
Available online 29 January 2014

Keywords:
Fusion protein
Chimera
Expression system

ABSTRACT

This work presents a protein structure that has been designed purely for aesthetic reasons, symbolizing decades of coiled-coil research and praising its most fundamental model system, the GCN4 leucine zipper. The GCN4 leucine zipper is a highly stable coiled coil which can be tuned to adopt different oligomeric states via mutation of its core residues. For these reasons it is used in structural studies as a stabilizing fusion adaptor. On the occasion of the 50th birthday of Andrei N. Lupas, we used it to create the first personalized protein structure: we fused the sequence ANDREI-N-LVPAS in heptad register to trimeric GCN4 adaptors and determined its structure by X-ray crystallography. The structure demonstrates the robustness and versatility of GCN4 as a fusion adaptor. We learn how proline can be accommodated in trimeric coiled coils, and put the structure into the context of the other GCN4-fusion structures known to date.

1. Introduction

Coiled coils are ubiquitous protein domains that are found in a vast range of functional contexts. They are bundles of α-helices that are wound around each other into superhelical structures and held together by a mostly hydrophobic interface in the core of the bundle. Predominantly, they have a heptad periodicity – they are built upon a simple, seven-residue sequence repeat, where the seven positions are labeled *a–g*, with positions *a* and *d* in the core and therefore mostly hydrophobic. Due to their simple, repetitive nature, the coiled coil is the only protein structure that can be modeled from sequence using simple parametric equations (Lupas and Gruber, 2005). Most naturally occurring coiled coils are dimeric, trimeric or tetrameric bundles, whereby the oligomerization specificity is governed to a large extent by the composition of the core residues.

Much of the basic research on coiled-coil structure and oligomerization specificity has been done using the coiled-coil domain of the yeast transcription factor GCN4 as a model system (Gonzalez et al., 1996; Harbury et al., 1993, 1994; O'Shea et al., 1991). This domain, known as the GCN4 leucine zipper, is a dimer by nature, but can be switched to higher oligomeric states by mutations in the *a*- and *d*-positions (Harbury et al., 1993). As the GCN4 leucine zipper is highly stable, it can serve as a stabilizing fusion partner for structural studies on otherwise hard to tame oligomeric proteins. The first reported crystal structures of proteins stabilized by such fusions were those of the trimeric viral glycoproteins HIV-1 GP41 (Weissenhorn et al., 1997) and Ebola virus GP2 (Weissenhorn et al., 1998). Here an N-terminal GCN4pII adaptor was employed, a trimeric GCN4 leucine zipper variant with the *a*- and *d*-positions changed to isoleucine. The first dimeric fusion structures were those of fragments of alpha-tropomyosin (Li et al., 2002) and vimentin (Strelkov et al., 2002), both with an N-terminal GCN4 adaptor based on the wild-type sequence.

About the time when the first crystal structure of the GCN4 leucine zipper was reported (O'Shea et al., 1991), Andrei N. Lupas published a landmark paper on the prediction of coiled coils from sequences (Lupas et al., 1991). Ever since he has steadily made significant contributions to the field, eventually adding a line of experimental coiled-coil research in his own department, amongst others with the authors of this paper. One major focus in the lab are Trimeric Autotransporter Adhesins (TAA), fibrous proteins in the outer membrane of gram-negative pathogens, a protein family with a seemingly inexhaustible diversity of trimeric coiled-coil motifs (Linke et al., 2006). As an important tool for studying TAA fragments, we developed a special variant of an expression vector for the construction of N- and C-terminal fusions to GCN4pII adaptors (Hernandez Alvarez et al., 2008). To date, we have solved more

* Corresponding author. Fax: +49 7071 601 349.
E-mail address: marcus.hartmann@tuebingen.mpg.de (M.D. Hartmann).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jsb.2014.01.013
1047-8477/

than 30 crystal structures of trimeric constructs using this vector, as well as several dimeric constructs with a modified version of the vector.

Now, for the occasion of Andrei's 50th birthday, which happened shortly before the 6th Alpbach workshop on coiled coils in 2013 – the year of the 60th anniversary of the structural model of the coiled coil proposed by Crick (1953) and by Pauling and Corey (1953) – we aimed to demonstrate the aesthetic potential of GCN4 fusions: we fused the sequence ANDREI-N-LVPAS to Andrei's own GCN4 adaptors and solved its (his) structure by X-ray crystallography.

2. Materials and methods

2.1. Cloning, expression and purification

In this work we used an improved version of our original pIBA-GCN4tri vector (Hernandez Alvarez et al., 2008). Firstly, the aspartate in position 6 of the C-terminal GCN4pII adaptor was replaced by tryptophan to allow for an improved quantification of fusion inserts without tryptophan residues via UV absorption at 280 nm. Secondly, the corresponding aspartate of the N-terminal GCN4pII adaptor was replaced by methionine to allow for a more efficient seleno-methionine labeling of the fusion protein, if needed. Thirdly, a hexa-histidine tag followed by a TEV (tobacco etch virus) protease cleavage site and a GGGSG-linker now precedes the N-terminal GCN4pII adaptor.

The DNA fragment encoding the coiled-coil construct shown in Fig. 1A was synthesized (Eurofins MWG). Restriction sites BamHI and XhoI were inserted at the ends of the synthetic gene and used for directed cloning. The gene sequence was verified by sequencing.

The protein was expressed in *Escherichia coli* TOP 10 cells at 37 °C. Expression was induced with 0.2 μg/ml anhydrotetracycline in the logarithmic growth phase. After incubation for another 4 h cells were harvested by centrifugation.

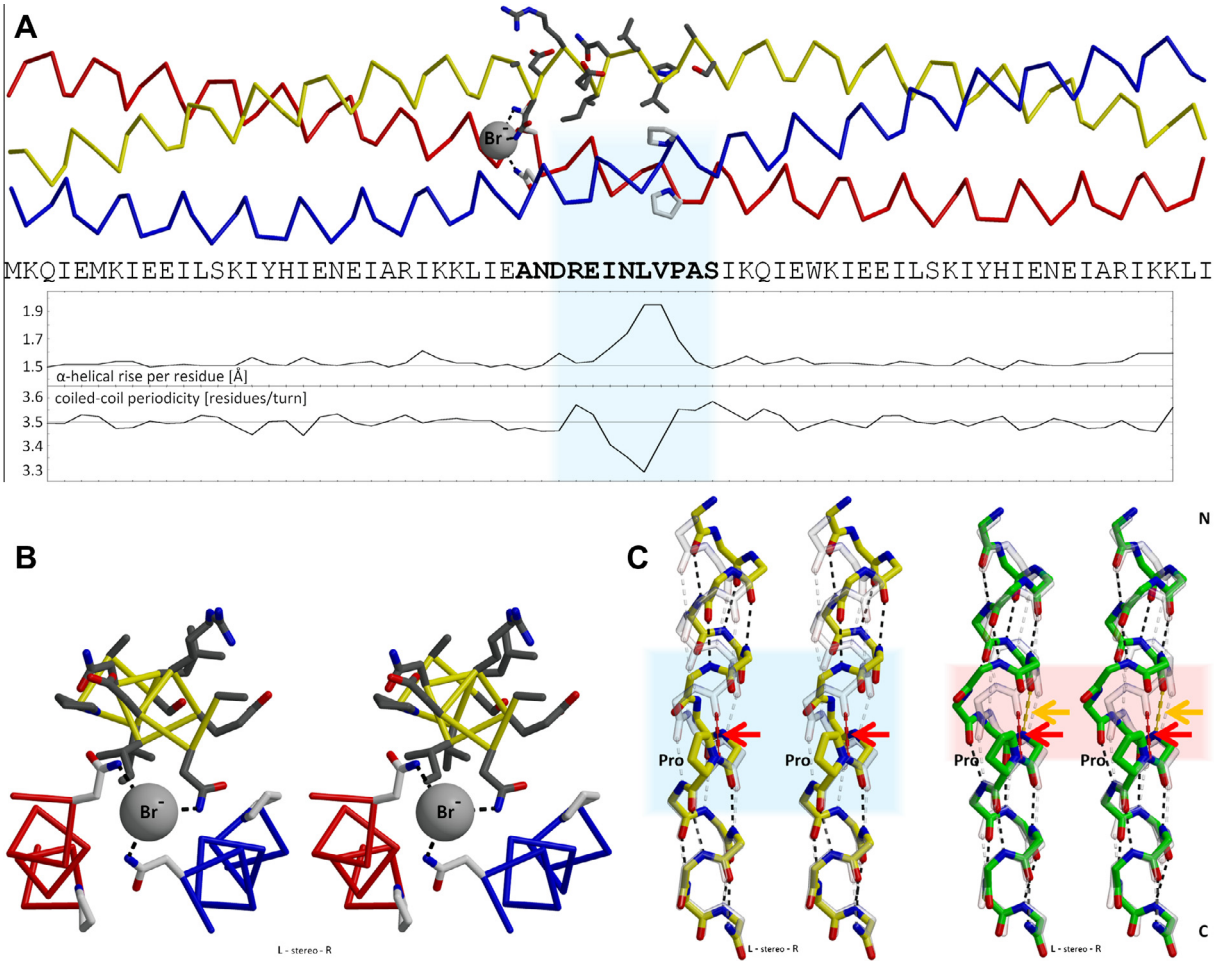

Fig.1. The structure of "ANDREI-N-LVPAS fused to GCN4 adaptors". The Cα traces of the three chains are colored individually. Residues of special interest have their side chains highlighted in grey: at the beginning of the insert, the asparagine of ANDREI in a *d*-position (N@*d*) is coordinating a bromide ion in the core of the bundle; at the end of the insert, the proline of LVPAS is causing major disturbances to the helices of the coiled coil. Additionally, in the yellow chain, the side chains of the insert sequence ANDREI-N-LVPAS are highlighted in dark grey. (A) The whole structure is aligned with its sequence and plots of its structural parameters. Throughout the GCN4 adaptors the α-helical rise per residue and the coiled-coil periodicity stay constant at values expected for heptad coiled coils (1.5 Å and 3.5 residues/turn, respectively). However, in the insert, a severe glitch of these parameters is observed: A drop in the periodicity to 3.3 residues/turn is accompanied by an increase in the rise per residue to almost 2 Å. This is the consequence of the accommodation of the proline, which disturbs the backbone hydrogen bonding pattern of the α-helices in the shaded area. The N-terminal sequence GGGSG preceding the N-terminal adaptor was not resolved in the electron density and is omitted from the sequence in the figure. (B) The structure of the sequence ANDREI-N-LVPAS, viewed from the N-terminus, highlighting the coordination of the bromide ion by the N@*d* residues. (C) The accommodation of the proline in the present structure (left, yellow) compared to the proline in 1ZTM with a preceding π-turn (right, green). For both structures, the backbone of a single helical segment containing the proline is superimposed on a helix from an unperturbed trimeric canonical coiled coil (transparent grey). The backbone hydrogen bonds of the helices are represented by dashed lines – in the shaded areas the i + 4 → i hydrogen bonding pattern of α-helices is broken. Highlighted in red is the bond that would involve the proline nitrogen. The single i + 5 → i hydrogen bond of the π-turn in 1ZTM is highlighted yellow. The direct comparison of the two modes of accommodation shows how an over-winding of the helix is circumvented in 1ZTM by the π-turn, which allows the helix to solve part of the sterical problem by providing more space for the proline side chain.

The cell pellet was resuspended in lysis buffer containing 20 mM Tris/HCl, pH 7.6, 150 mM NaCl, 4 mM $MgCl_2$, DNase I and cOmplete Protease Inhibitor Cocktail (Roche). After lysis using a French press, guanidinium hydrochloride and NaCl were added to the lysate to final concentrations of 6 M and 500 mM, respectively, followed by stirring for 1 h at room temperature. The denatured cell extract was centrifuged and the supernatant was loaded on a NiNTA column equilibrated with 20 mM Tris/HCl, pH 8, 500 mM NaCl, 6 M guanidinium hydrochloride. The bound histidine tagged protein was eluted by a linear gradient from 0 to 500 mM imidazole. Protein containing fractions were pooled and dialyzed against 20 mM Tris/HCl, pH 7.7, 150 mM NaCl for refolding. The hexa-histidine tag was removed by TEV cleavage at 4 °C for 12 h. The cleaved tag and the histidine tagged TEV protease were removed by running a second NiNTA column in 20 mM Tris/HCl, pH 7.5, 150 mM NaCl. Although the untagged protein was found to be bound to the column it could be separated from the TEV protease using a gradient from 0 to 500 mM imidazole for elution. The protein was dialyzed against 50 mM sodium acetate buffer, pH 4.0, 50 mM NaCl and subsequently concentrated to 6.7 mg/ml.

2.2. Crystallization, data collection and structure determination

Crystallization trials were performed at 295 K. Sitting drops containing 300 nl of protein solution and 300 nl of reservoir solution were prepared on 96-well plates with a reservoir volume of 50 µl. Crystals grew only in conditions containing bromide salts. The best diffracting crystal was harvested after a week from a condition containing 1.75 M sodium bromide and 100 mM HEPES, pH 7.5. It was transferred into a drop containing 30% (v/v) glycerol in addition to the reservoir solution, loop-mounted, and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K and a wavelength of 1 Å on a PILATUS 6 M detector at beamline PXII of the Swiss Light Source (PSI, Villigen, Switzerland). Data were indexed, integrated and scaled to a resolution of 1.95 Å in space group $P2_1$ using XDS (Kabsch, 1993). The unit cell parameters (Table 1) suggested a single trimer in the asymmetric unit with a solvent content of about 40%. Molecular replacement was carried out with MOLREP (Vagin and Teplyakov, 2000) and a trimeric GCN4pII adaptor cut from the structure 2YNY as a search model. Both the N- and C-terminal GCN4 adaptor of the single trimer were found in a single translation search using the pseudo translation vector (3.5, 0, 26.4 Å) of a very strong (36% of origin) Patterson peak. After initial rigid body refinement using REFMAC5 (Murshudov et al., 1999), the insert between the adaptors was clearly traceable in the electron density so that the structure could be completed by cyclic manual modeling with Coot (Emsley and Cowtan, 2004) and refinement with REFMAC5. Data collection and refinement statistics are summarized in Table 1. The structure was deposited in the Protein Data Bank (PDB) under accession code 4C46.

Table 1
Data collection and refinement statistics.

Data collection	
Space group	$P2_1$
Unit cell parameters	a = 42.00 Å, b = 41.60 Å, c = 64.14 Å, β = 95.84°
Resolution range (Å)	36.6–1.95 (2.07–1.95)
Completeness (%)	98.7 (97.5)
Redundancy	3.28 (3.33)
$I/\sigma(I)$	11.3 (1.94)
R_{merge} (%)	5.7 (57.5)
Refinement	
Resolution range (Å)	36.6–1.95 (2.0–1.95)
R_{cryst} (%)	23.8 (34.5)
R_{free} (%)	27.5 (44.8)
# atoms protein/bromide/water	1755/1/72
Mean B value protein/bromide/water ($Å^2$)	43.6/39.1/48.8
Bond length/angle RMSD (Å/°)	0.016/1.72
Residues in the core region of the Ramachandran plot	100%

Values in parenthesis refer to the highest resolution shell; Ramachandran statistics were determined by PROCHECK (Laskowski et al., 1993).

2.3. Bioinformatics

The coiled-coil and alpha-helical parameters shown in Fig. 1A were determined using TWISTER (Strelkov and Burkhard, 2002) and plotted using Gnuplot. All molecular depictions were prepared using MolScript (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

For a survey of coiled-coil structures with embedded proline residues we employed the CC+ database as of 26th April, 2013, a relational database of coiled coils of known structure compiled from the Protein Data Bank (Testa et al., 2009). CC+ was configured to return all structures with proline residues embedded in parallel, homo-oligomeric coiled-coil assignments of any length. Filtering at 50% sequence identity yielded 2 hits for trimers (1ZTM and 3RRT) and 6 hits for dimers (1LJ2, 1WZ8, 2R9A, 2V4H, 3DV8 and 3SR2). These structures were manually inspected to exclude instances where proline is found only at coiled-coil N-termini, which reduced the list for dimers to 4 instances (1LJ2, 2R9A, 2V4H and 3SR2).

To collect all known GCN4 fusion structures we did a BLAST search with the sequence MKQLEDKVEELLSKNYHLENEVARLKKL of the GCN4 dimerization domain against the PDB as of 30th October, 2013. The full-length sequences of hits with an *E*-value lower than 1.0 were collected and filtered manually to contain only true fusion proteins with a minimum length of 40 residues. As the only exception, the wild-type GCN4 structure 2DGC was kept for illustrative purposes. The resulting set contained 36 sequences including "ANDREI-N-LVPAS fused to GCN4 adaptors". This set was enriched by the sequences of 10 unreleased fusion structures that we presented at the 6th Alpbach meeting and which are currently to be published, yielding 46 sequences. These were clustered in 2D using CLANS (Frickey and Lupas, 2004) at a *P*-value cut off of 9.0*e*-20 with default settings. (A very illustrative example of this clustering approach is given in Alva et al. (2010).) Details of the sequences are described in Table 2. It should be noted that the actual number of GCN4 fusion structures is higher than the number of GCN4 fusion constructs, as for some constructs there are multiple structures deposited.

3. Results and discussion

3.1. The structure of Andrei N. Lupas fused to GCN4 adaptors

For the design of the fusion construct, we tried to find a register for the insert sequence that is most suitable for trimeric coiled coils. Therefore, the sequence ANDREI-N-LVPAS was fused such that the isoleucine occupies an *a*-position, and that the valine and the first asparagine occupy *d*-positions. The asparagine in *d* (N@*d*) seemed especially attractive, as it can serve as a trimerization determinant and coordinates an anion in the core of the trimer (Hartmann et al., 2009). With this assignment, the insert would start with a *c*-position and end with a *g*-position. To maintain a continuous heptad register of the fusion protein, two amino acids, isoleucine (in position *a*) and glutamate, were added to the N-terminus of the insert sequence, and then the whole 14-residue insert was fused to N- and C-terminal GCN4pII adaptors, yielding the

Table 2
Description of the 46 sequences in the cluster map and associated PDB codes. Only one PDB code is listed for sequences with more than one PDB entry. The column "N or C" details whether GCN4 was fused N- and/or C-terminally.

Seq#	PDB	N or C	Fused to:	References
Trimers				
1	4C46	N + C	ANDREI-N-LVPAS	This work
2–7	–	N + C	Short inserts (<10 aa)	Hartmann et al. (in preparation)
8	–	N + C	The α/β coiled coil	
9	–	N + C	TAA fragment from *actinobacillus*	
10	2XZR	N + C	*E. coli* immunoglobulin-binding protein EibD fragment 391–438	Leo et al. (2011)
11	2WPR	N + C	*Salmonella enterica* SadA fragment 483–523	Hartmann et al. (2009)
12	2WPQ	N + C	*Salmonella enterica* SadA fragment 479–519	
13	2YO3	N + C	*Salmonella enterica* SadA fragment 1185–1386	Hartmann et al. (2012)
14	2YO0	N*+C	*Salmonella enterica* SadA fragment 1049–1304	
15	2YNY	N + C	*Salmonella enterica* SadA fragment 255–302	
16	2YNZ	N	*Salmonella enterica* SadA fragment 823–947	
17	2YO2	N + C	*Salmonella enterica* SadA fragment 255–358	
18	3ZMF	N + C	*Salmonella enterica* SadA fragment 303–358	Hernandez Alvarez et al. (2008)
19	1CZQ	N	HIV-1 gp41 fragment	Eckert et al. (1999)
20	3L35	N	HIV-1 gp41 fragment	Welch et al. (2010)
21	1FAV	N	HIV-1 gp41 fragment	Zhou et al. (2000)
22	1ENV	N	HIV-1 gp41 fragment	Weissenhorn et al. (1997)
23	2VNL	C	Headbinding domain of phage P22 tailspike	To be published
24	2VKY	C	Headbinding domain of phage P22 tailspike	
25	3S6X	N *	Reovirus attachment protein sigma1 fragment	Reiter et al. (2011)
26	1EBO	N	Ebola virus GP2 ectodomain	Weissenhorn et al. (1998)
27	4G2K	N	Marburg virus GP2 ectodomain	Koellhoffer et al. (2012)
28	2B9B	C	Parainfluenza virus 5 F protein fragment	Yin et al. (2006)
29	4GIP	C*	Parainfluenza virus 5 F protein fragment	Welch et al. (2012)
30	2R32	N	Human GITRL variant	Chattopadhyay et al. (2007)
Tetramer				
31	3NAF	N*	Gating ring of human high-conductance Ca2 + gated K + Channel	Wu et al. (2010)
Dimers				
32	2D3E	N	C-Terminal fragment of rabbit skeletal alpha-tropomyosin	To be published
33	2Z5H	N/C	Head-to-tail junction of tropomyosin	Murakami et al. (2008)
34	1KQL	N	C-terminal region of striated muscle alpha-tropomyosin	Li et al. (2003)
35	2EFR	N + C	C-terminal tropomyosin fragment	To be published
36	2B9C	C	Tropomyosin's mid-region	Brown et al. (2005)
37	1GK6	N	Human vimentin coil 2b fragment	Strelkov et al. (2002)
38	1NKN	C	N-terminal segment of the scallop myosin rod	Li et al. (2003)
39	3P8M	C	In vitro evolved peptide	Rapali et al. (2011)
40	3I5C	N	GGDEF domain of WspR from *Pseudomonas aeruginosa*	De et al. (2009)
41	–	N	Poly-HAMP construct with 4 HAMP domains	Ewers et al. (in preparation)
42	–	N	Poly-HAMP construct with 6 HAMP domains	
43	2WG5	N	Proteasome-activating nucleotidase (PAN) N-domain fragment	Djuranovic et al. (2009)
44	2WG6	N	Proteasome-activating nucleotidase (PAN) N-domain fragment, mutant	
45	1LLM	C	Transcription factor Zif268 fragment	Wolfe et al. (2003)
46	2DGC	–	No fusion: GCN4 basic domain, leucine zipper complexed with DNA	Keller et al. (1995)

* The corresponding GCN4 adaptor is not resolved in the structure.

sequence shown in Fig. 1A. The construct was expressed in *E. coli*, purified, and set up for crystallization. After one week we obtained well-diffracting crystals that led to a dataset with a resolution of 1.95 Å. We could solve the structure via molecular replacement using a GCN4pII adaptor as a search model.

As anticipated, the overall fold of the structure appears to be a continuous trimeric coiled coil with heptad periodicity (Fig. 1A). The helices of the GCN4 adaptors pass seamlessly into the helices of the insert, and at the beginning of the insert, the N@*d* layer coordinates an anion in the core of the bundle. The anion is enclosed in a cavity bounded by the N@*d* layer on one side and by the isoleucines in the preceding *a*-position on the other side. In the majority of the known instances, N@*d* layers coordinate chloride ions (Hartmann et al., 2009). Here, due to a large excess of sodium bromide in the crystallization condition, the N@*d* layer sequesters a bromide ion instead (Fig. 1B). In fact, we did not obtain crystals in conditions without bromide salts. This might be due to the fact that bromide-coordinating N@*d* layers have a larger coiled coil radius, which might have been crucial for the present crystal packing. However, the most distinctive feature of the structure becomes manifest C-terminally to the N@*d* layer: the remainder of the insert is not strictly α-helical because the proline residue causes major perturbations.

3.2. The accommodation of proline in the heptad repeat

Proline is unable to participate in the backbone hydrogen bonding interactions of α-helices and poses a sterical problem for the helices. In the present structure it occupies an *e*-position and Fig. 1C illustrates how its side chain is literally stuck in the helical backbone, severely distorting the geometry of the preceding helical turn. Strictly speaking, this turn is no longer α-helical; it is effectively over-wound, with a helical rise per residue corresponding to a 3_{10} helix. This is especially evident in sharp peaks in the plots of the α-helical rise per residue and the coiled-coil periodicity in Fig. 1A. Interestingly, we have observed a comparable behavior for the insertion of 3 residues (a stammer) into a heptad repeat coiled coil, which leads to a similar over-winding and steeper rise per residue of the helices (Hartmann et al., 2009). All together, the proline in the present structure impairs the formation of 5 backbone hydrogen bonds within the helix (Fig. 1C).

A search for other instances of proline embedded in coiled coils of known structure using CC+ (Testa et al., 2009) yielded 2 more structures of trimeric coiled coils, 1ZTM (Yin et al., 2005) and 3RRT (McLellan et al., 2011), both of viral fusion proteins. In these two proteins the proline is found in an *f*-position and

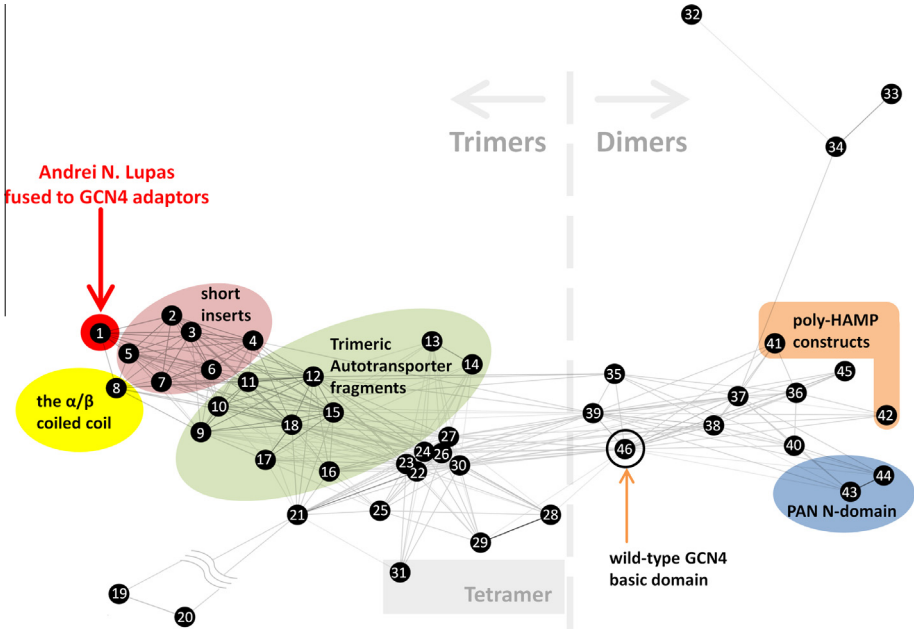

Fig.2. Cluster map of all reported GCN4-fusion structures with non-identical sequence, created with CLANS (Frickey and Lupas, 2004). The map contains 45 sequences including "ANDREI-N-LVPAS fused to GCN4 adaptors", plus the wild-type GCN4 structure 2DGC. It shows an equilibrated state after a force-directed clustering procedure in which the sequences are treated as point masses in a virtual 2D space that attract or repel each other according to their pairwise sequence similarity. The 22 encircled and annotated constructs originate from Andrei's lab. Of these, "ANDREI-N-LVPAS fused to GCN4 adaptors" was not conceived by Andrei.

an over-winding of the helices is avoided by the formation of a π-turn N-terminal to the proline, as exemplified for 1ZTM in Fig. 1C. A π-turn results from the incorporation of an additional residue into a helix, resulting in a wider turn. This provides more room for the accommodation of the proline by departing from the heptad repeat.

For dimeric coiled coils we found 4 structures. One of them, 1LJ2 (Groft and Burley, 2002), has the proline (in position *b*) embedded in the heptad repeat and a similar mode of accommodation as in the present structure. In two others, 2R9A (Andres et al., 2007) and 3SR2 (Hammel et al., 2011), the proline (in position *c*) is accompanied by a π-turn as in the trimers described above. However, the fourth example, 2V4H (Grubisha et al., 2010), shows an interesting peculiarity: The proline is found right after an insertion of 3 residues (a stammer) into the heptad repeat. It occupies a *b*-position when the resulting decad is written in heptad notation as *[abc]ab**b**cdefg* (where the brackets delimit the stammer), which is structurally equivalent to the proline in the *e*-position in the present structure. As mentioned above, the effect of a stammer is similar to that of a proline in heptad repeat. What is special about this structure is that it combines a stammer and a proline such, that their effects are not cumulative: depending on the point of view, one could either argue that the proline sits in a position where it is structurally tolerated due to the over-winding and steeper rise per residue caused by the stammer, or that the stammer is located where an over-winding and steeper rise per residue is required by the proline.

The available set of instances is very small, so that general conclusions are hardly possible. However, one interesting observation is that an accommodation via π-turns is only found for prolines in position *f* or *c*. These are, together with position *b*, the positions on the outer face of the coiled coil. As the π-turns bulge out the helices on the face on which the proline is located, it is conceivable that this accommodation mode is generally restricted to prolines in these outer positions. On the contrary, an accommodation within the heptad repeat as in the present structure should be possible in any position.

3.3. Andrei N. Lupas and the rest of the GCN4-fusion world

This paper presents a worthwhile opportunity to compile an overview of all GCN4 fusion structures that have been reported to date. For this purpose we did a BLAST search with the GCN4 leucine zipper sequence against the Protein Data Bank. After filtering the hits as detailed in the methods section we counted 35 structures with non-identical sequence. To these we added the unreleased fusion structures that we presented at the 6th Alpbach meeting in 2013: 2 dimeric constructs of consecutive HAMP domains (Ewers et al., in preparation) and 8 trimeric constructs including the α/β coiled coil (Hartmann et al., in preparation). We then calculated a sequence cluster map for the resulting 45 structures, plus the wild-type GCN4 reference structure 2DGC (Fig. 2 and Table 2).

The 45 sequences of fusion proteins in the map contain 14 dimers, 30 trimers and 1 tetramer. Although many of the fusion proteins have nothing in common apart from their GCN4 adaptors, the map shows a clear separation into dimers and trimers. Highlighted and annotated are the structures originating from the Lupas lab: 4 dimers and 18 trimers. When these are omitted, the map becomes much sparser: the trimeric half, with one exception, consists only of viral (9) and phage (2) proteins, and the dimeric half is dominated by muscle and filamental proteins (7 out of 10). In contributing a large number of TAA fragments and other trimeric coiled-coil motifs, as well as some dimeric constructs as esoteric as GCN4 fused to the N-domain of the proteasomal ATPase PAN, Andrei has clearly brought a lot more diversity into the GCN4 fusion world. With "ANDREI-N-LVPAS fused to GCN4 adaptors" he extends this contribution, literally. Depending on your name, You too could contribute to the world of GCN4 fusions with your own personalized protein structure.

Acknowledgments

We would like to thank Andrei N. Lupas for sharing his highly inspiring enthusiasm for coiled coils with us. We thank Moritz

Ammelburg for the critical reading of the manuscript and suggesting the profound coloring scheme for the trimer structure. We are very grateful to the staff of beamline PXII at the Swiss Light Source and to Kotaro Koiwai for excellent technical support. The work was supported by institutional funds from the Max Planck Society.

References

Alva, V., Remmert, M., Biegert, A., Lupas, A.N., Söding, J., 2010. A galaxy of folds. Protein Sci. 19, 124–130.

Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., Junop, M.S., 2007. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. Mol. Cell 28, 1093–1101.

Brown, J.H., Zhou, Z., Reshetnikova, L., Robinson, H., Yammani, R.D., Tobacman, L.S., Cohen, C., 2005. Structure of the mid-region of tropomyosin: bending and binding sites for actin. Proc. Natl. Acad. Sci. USA 102, 18878–18883.

Chattopadhyay, K., Ramagopal, U.A., Mukhopadhaya, A., Malashkevich, V.N., Dilorenzo, T.P., Brenowitz, M., Nathenson, S.G., Almo, S.C., 2007. Assembly and structural properties of glucocorticoid-induced TNF receptor ligand: implications for function. Proc. Natl. Acad. Sci. USA 104, 19452–19457.

Crick, F.H., 1953. The packing of alpha-helices: simple coiled-coils. Acta Cryst. 6, 669–689.

De, N., Navarro, M.V., Raghavan, R.V., Sondermann, H., 2009. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J. Mol. Biol. 393, 619–633.

Djuranovic, S., Hartmann, M.D., Habeck, M., Ursinus, A., Zwickl, P., Martin, J., Lupas, A.N., Zeth, K., 2009. Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. Mol. Cell 34, 580–590.

Eckert, D.M., Malashkevich, V.N., Hong, L.H., Carr, P.A., Kim, P.S., 1999. Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell 99, 103–115.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Frickey, T., Lupas, A., 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics (Oxford, England) 20, 3702–3704.

Gonzalez Jr., L., Woolfson, D.N., Alber, T., 1996. Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat. Struct. Biol. 3, 1011–1018.

Groft, C.M., Burley, S.K., 2002. Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization. Mol. Cell 9, 1273–1283.

Grubisha, O., Kaminska, M., Duquerroy, S., Fontan, E., Cordier, F., Haouz, A., Raynal, B., Chiaravalli, J., Delepierre, M., Israel, A., Veron, M., Agou, F., 2010. DARPin-assisted crystallography of the CC2-LZ domain of NEMO reveals a coupling between dimerization and ubiquitin binding. J. Mol. Biol. 395, 89–104.

Hammel, M., Rey, M., Yu, Y., Mani, R.S., Classen, S., Liu, M., Pique, M.E., Fang, S., Mahaney, B.L., Weinfeld, M., Schriemer, D.C., Lees-Miller, S.P., Tainer, J.A., 2011. XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. J. Biol. Chem. 286, 32638–32650.

Harbury, P.B., Zhang, T., Kim, P.S., Alber, T., 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262, 1401–1407.

Harbury, P.B., Kim, P.S., Alber, T., 1994. Crystal structure of an isoleucine-zipper trimer. Nature 371, 80–83.

Hartmann, M.D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D.N., Sauer, G., Dunin-Horkawicz, S., Lupas, A.N., Alvarez, B.H., 2009. A coiled-coil motif that sequesters ions to the hydrophobic core. Proc. Natl. Acad. Sci. USA 106, 16950–16955.

Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N., Hernandez Alvarez, B., 2012. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. Proc. Natl. Acad. Sci. USA 109, 20907–20912.

Hernandez Alvarez, B., Hartmann, M.D., Albrecht, R., Lupas, A.N., Zeth, K., Linke, D., 2008. A new expression system for protein crystallization using trimeric coiled-coil adaptors. Protein Eng. Des. Sel. 21, 11–18.

Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 26, 795–800.

Keller, W., Konig, P., Richmond, T.J., 1995. Crystal structure of a bZIP/DNA complex at 2.2 A: determinants of DNA specific recognition. J. Mol. Biol. 254, 657–667.

Koellhoffer, J.F., Malashkevich, V.N., Harrison, J.S., Toro, R., Bhosle, R.C., Chandran, K., Almo, S.C., Lai, J.R., 2012. Crystal structure of the Marburg virus GP2 core domain in its postfusion conformation. Biochemistry 51, 7665–7675.

Kraulis, P.J., 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK – a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

Leo, J.C., Lyskowski, A., Hattula, K., Hartmann, M.D., Schwarz, H., Butcher, S.J., Linke, D., Lupas, A.N., Goldman, A., 2011. The structure of *E. coli* IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. Structure 19, 1021–1030.

Li, Y., Mui, S., Brown, J.H., Strand, J., Reshetnikova, L., Tobacman, L.S., Cohen, C., 2002. The crystal structure of the C-terminal fragment of striated-muscle alpha-tropomyosin reveals a key troponin T recognition site. Proc. Natl. Acad. Sci. USA 99, 7378–7383.

Li, Y., Brown, J.H., Reshetnikova, L., Blazsek, A., Farkas, L., Nyitray, L., Cohen, C., 2003. Visualization of an unstable coiled coil from the scallop myosin rod. Nature 424, 341–345.

Linke, D., Riess, T., Autenrieth, I.B., Lupas, A., Kempf, V.A., 2006. Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol. 14, 264–270.

Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences. Science 252, 1162–1164.

Lupas, A.N., Gruber, M., 2005. The structure of alpha-helical coiled coils. Adv. Protein Chem. 70, 37–78.

McLellan, J.S., Yang, Y., Graham, B.S., Kwong, P.D., 2011. Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J. Virol. 85, 7788–7796.

Merritt, E.A., Bacon, D.J., 1997. Raster3D: photorealistic molecular graphics. Methods Enzymol. 277, 505–524.

Murakami, K., Stewart, M., Nozawa, K., Tomii, K., Kudou, N., Igarashi, N., Shirakihara, Y., Wakatsuki, S., Yasunaga, T., Wakabayashi, T., 2008. Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-T. Proc. Natl. Acad. Sci. USA 105, 7200–7205.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., Dodson, E.J., 1999. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. 55, 247–255.

O'Shea, E.K., Klemm, J.D., Kim, P.S., Alber, T., 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254, 539–544.

Pauling, L., Corey, R.B., 1953. Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. Nature 171, 59–61.

Rapali, P., Radnai, L., Suveges, D., Harmat, V., Tolgyesi, F., Wahlgren, W.Y., Katona, G., Nyitray, L., Pal, G., 2011. Directed evolution reveals the binding motif preference of the LC8/DYNLL hub protein and predicts large numbers of novel binders in the human proteome. PLoS One 6, e18818.

Reiter, D.M., Frierson, J.M., Halvorson, E.E., Kobayashi, T., Dermody, T.S., Stehle, T., 2011. Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. PLoS Pathog. 7, e1002166.

Strelkov, S.V., Burkhard, P., 2002. Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J. Struct. Biol. 137, 54–64.

Strelkov, S.V., Herrmann, H., Geisler, N., Wedig, T., Zimbelmann, R., Aebi, U., Burkhard, P., 2002. Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. EMBO J. 21, 1255–1266.

Testa, O.D., Moutevelis, E., Woolfson, D.N., 2009. CC+: a relational database of coiled-coil structures. Nucleic Acids Res. 37, D315–322.

Vagin, A., Teplyakov, A., 2000. An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. Crystallogr. 56, 1622–1624.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature 387, 426–430.

Weissenhorn, W., Carfi, A., Lee, K.H., Skehel, J.J., Wiley, D.C., 1998. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol. Cell 2, 605–616.

Welch, B.D., Francis, J.N., Redman, J.S., Paul, S., Weinstock, M.T., Reeves, J.D., Lie, Y.S., Whitby, F.G., Eckert, D.M., Hill, C.P., Root, M.J., Kay, M.S., 2010. Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. J. Virol. 84, 11235–11244.

Welch, B.D., Liu, Y., Kors, C.A., Leser, G.P., Jardetzky, T.S., Lamb, R.A., 2012. Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. Proc. Natl. Acad. Sci. USA 109, 16672–16677.

Wolfe, S.A., Grant, R.A., Pabo, C.O., 2003. Structure of a designed dimeric zinc finger protein bound to DNA. Biochemistry 42, 13401–13409.

Wu, Y., Yang, Y., Ye, S., Jiang, Y., 2010. Structure of the gating ring from the human large-conductance Ca(2+)-gated K(+) channel. Nature 466, 393–397.

Yin, H.S., Paterson, R.G., Wen, X., Lamb, R.A., Jardetzky, T.S., 2005. Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. Proc. Natl. Acad. Sci. USA 102, 9288–9293.

Yin, H.S., Wen, X., Paterson, R.G., Lamb, R.A., Jardetzky, T.S., 2006. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. Nature 439, 38–44.

Zhou, G., Ferrer, M., Chopra, R., Kapoor, T.M., Strassmaier, T., Weissenhorn, W., Skehel, J.J., Oprian, D., Schreiber, S.L., Harrison, S.C., Wiley, D.C., 2000. The structure of an HIV-1 specific cell entry inhibitor in complex with the HIV-1 gp41 trimeric core. Bioorg. Med. Chem. 8, 2219–2227.