Journal of Proteomics & Bioinformatics

Research Article — **Open Access**

Design and Selection of IFI16-PAAD Mutants with Improved dsDNA Destabilization Properties

Desmond Lau#, Kush Dalal#, Benjamin Hon and Frederic Pio*

Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr, Burnaby, British Columbia V5A 1S6, Canada
#Equal contribution for both authors

Abstract

The PAAD Death Domain is a 6-helix bundle domain that contains a disordered region in helix-3. This domain can fold by a two-state mechanism, has low stability and can undergo different partially folded conformations. To identify specific amino acids responsible for these structural features we focus our study on the PAAD domain of IFI16, a protein sensor that recognizes viral nucleic acids and triggers the (Interleukin-β) inflammatory response in response to infection. IFI16-PAAD was randomly mutated using GFP directed evolution and the resulting mutants with increased secondary structure and stability were identified by circular dichroism and tested for nucleic acid binding. Several IFI16-PAAD-GFP fusion mutants that contained I46N/S substitutions translated into improved secondary structure and stability and a direct correlation was observed between secondary structure/stability enhancement and their ability to destabilize dsDNA. In conclusion, GFP directed evolution can be used to obtain more structured and stable mutants and to probe conformational changes. These data suggest that changes in conformation of the PAAD domain of IFI16 occurs when the viral nucleic acid bind to IFI16 and must be a contributing factor to transduce the inflammatory response during viral infection.

Keywords: PAAD; IFI16; GFP directed evolution; Thermodynamic; Circular dichroism; Melting curves

Introduction

The Death Domain (DD) superfamily is a recruitment domain linked to apoptosis and inflammation by protein-protein interactions that allow the propagation of their signalling pathways [1,2]. Members of the super family include DD, Death Effector Domain (DED), Caspase Recruitment Domain (CARD) and the Pyrin, AIM, ASC and Death-domain like (PAAD). Despite low sequence identity (5-30%), members of the super family show a structurally conserved anti-parallel six-helix bundle fold arranged as a Greek key motif [3].

The Pyrin/PAAD domain is located at the N-terminus of many apoptosis and inflammation related proteins [4-12]. NMR studies of NALP1-PAAD [13] and ASC-PAAD domains [14], revealed an anti-parallel Greek key helical bundle. However, helix-3 involved in protein-protein interactions for CARD proteins [15,16] is either partially or completely disordered in both structures, with the exception that MNDA-PAAD (PDB=2DBG) has a greater secondary structure and stability than other PAAD domains [1]. In addition, mutation R42W of the Pyrin protein that causes Familial Mediterranean Fever (FMF) was located in the unfolded helix 3 of ASC-PAAD [14]. Thus, the level of disorder in helix-3 is variable and may contribute to different protein-ligands functions or specificities.

The reasons for the disordering of helix-3 in PAAD domains are currently unknown. The stability of a folded structure depends on many different physical-chemical forces, and proteins can sample many different conformations before reaching its stable structure with the lowest Gibbs free energy. The strength and specificity of these forces will depend on environmental conditions to reduce or even eliminate part of the conformational interactions. Under some environmental conditions the native protein structure can be transformed into new conformations with properties that are intermediate between those of the native and the completely unfolded states [17]. Proteins may also adopt residual structure, which can be in the form of fluctuating secondary structure and dynamic side-chain motions, particularly in hydrophobic clusters [18]. This latter state might only occur to allow the protein to interact with its ligands with greater affinity and in a more energetically favorable manner. The ability of PAAD domain proteins to adopt different stable partially folded conformations may also be encoded in specific features of its amino acids sequence.

If specific mutations have occurred in the PAAD domain during evolution that are responsible for this partially folded conformation, it may not be straightforward to identify them since sequence alignments of the PAAD domain show little conservation in the region near the disordered helix. In addition, other DD share even less similarity to PAAD domains, making comparison of helix-3 between CARD/DD/DED and PAAD domains difficult.

Directed protein evolution describes several mutagenesis and selection techniques that mimic natural protein evolution *in vitro*. It has been extremely useful in protein engineering and *de novo* design [19,20].

In this study, we applied directed protein evolution to the PAAD domain of Interferon-γ Inducible protein 16 (IFI16) [11,21] in order to identify mutants that confer increased secondary structure and stability and that may restructure helix-3 as seen in other DD, DED and CARD domains. These mutants were expressed, purified and tested for their ability to bind to nucleic acid in a destabilization assay. The *in vitro* GFP directed protein evolution method [22-24] was employed as a first selection screen for mutants with greater folding and solubility [22-24].

*Corresponding author: Frederic Pio, Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr, Burnaby, British Columbia V5A 1S6, Canada, Tel: 778-782-5660; Fax: 778-782-5583; E-mail: fpio@sfu.ca

We show that some IFI16-PAAD-GFP fusion mutants showing greater than wild type fluorescence translate into improved protein folding and stability. We also show that there are 4 critical exposed amino acid positions on IFI16-PAAD that can modulate the helicity and stability of the protein, and could be involved in the refolding of the partially disordered structure. Mutants with improved secondary structure and stability have enhanced ability to destabilize double-stranded nucleic acid.

Materials and Methods

Flow diagram for *in vitro* GFP directed evolution of IFI16-PAAD

In vitro GFP directed evolution of IFI16-PAAD and selection of soluble mutants was performed by random mutagenesis followed by three rounds of DNA shuffling. The procedure is represented as a flow diagram in Supplementary Figure 1.

Random mutagenesis of IFI16-PAAD

Random mutagenesis of the 372 bp cDNA fragment of the IFI16-PAAD domain [1] encoding the first 102 amino acids of IFI16 protein, was performed using Genemorph II EZ mutagenesis kit (Stratagene) followed by DNA sequencing (Macrogen, Korea). After error prone PCR, mutated PCR products were cloned into GFP expressing vector via BamHI and NdeI restriction sites. Approximately 100,000 colonies were plated and the plasmid DNA was isolated using Midiprep kit (Qiagen). Plasmid DNA was transformed into bacterial strain BL21 (DE3) and plated on nitrocellulose filters that were placed directly on agar plates. After colonies grew to 1 mm in diameter, nitrocellulose filters were transferred to Petri dishes containing 5 mL LB media with 1 mM IPTG and 40 μg/mL kanamycin for protein expression. Fluorescent colonies expressing the PAAD-GFP mutants were visualized using a hand held UV lamp. Although not quantitative, the UV lamp was able to discriminate large relative changes in fluorescence when comparing bacterial colonies, either positive or negative. Accordingly, out of 5000 colonies plated, 60 colonies showing the greatest fluorescence as judged by visual examination were selected and followed by DNA sequencing of at least 10 colonies to determine the mutagenic rate.

DNA shuffling of IFI16-PAAD mutants

DNA shuffling was used to recombine the mutations generated from the error prone PCR. The recombination of the favourable mutations was used to further increase the folding of the IFI16-PAAD domain. Mutated IFI16-PAAD domains isolated from the 60 brightest colonies (isolated by Midiprep) were subjected to 3 rounds of DNA shuffling as described by Arnold and co-workers [25]. The 5 steps of the DNA shuffling procedure are briefly described below:

PCR and DNase I digestion: ~2 μg of both mutant and wild type PAAD PCR product were mixed and digested with 0.35U DNase I (50 mM Tris-HCl pH 7.4, 10 mM MnCl$_2$) for 5 minutes at 15°C. The wild type DNA was included in the digest to remove non-essential mutations responsible for increasing GFP fluorescence. DNA fragments after digestion were run on a 20% polyacrylamide TAE gel. Fragments of less than 50 bp were purified by gel extraction and concentrated by ethanol precipitation.

Fragment reassembly and amplification: The DNA fragments (10 μL of re-concentrated stock from the previous ethanol precipitation) were reassembled with 40 cycles of PCR (1 min 94°C, 1 min 55°C, 1 min 72°C) without primers using 0.06 U/μL high fidelity pfu DNA polymerase and 0.4 mM dNTP. The reassembled fragments (shuffled library) were used as a template for 30 cycles of PCR (30 s 94°C, 45 min 55°C, 1 min 72°C) with a 1:1 mixture (2.5 U) of pfu: taq polymerases and 0.4 mM dNTP.

TA cloning of shuffled mutants: For TA cloning, a final 30 cycles PCR (using the PCR product amplified from the fragment reassembly as the template) were performed with 0.04 U/μL taq polymerase (30 cycles-30 s 94°C, 45 s 55°C, 1 min 72°C) to add T and A overhangs to the amplification products. Shuffled PAAD domains were cloned into the TOPO TA cloning vector (Invitrogen) and transformed into electro-competent DH5α cells. Plasmid DNA from the 10,000 colonies pooled in LB (40 μg/ml Kanamycin) was isolated by Midiprep (Stratagene).

Cloning into GFP folding vector: The PAAD domain mutants (shuffled) were excised from the TOPO vector by NdeI/BamH1 digestion and cloned into the GFP expressing vector. The ligation product was transformed into DH5α competent cells and grown overnight on LB plates containing 40 μg/mL kanamycin.

Expression in BL21 (DE3): DH5α colonies were pooled and plasmid DNA corresponding to the first round shuffled PAAD mutants were isolated by Miniprep (Stratagene). After transformation into BL21 (DE3), cells were plated on nitrocellulose filters and induced as described in the random mutagenesis of IFI16-PAAD section. From the 5,000 colonies screened, the 40 brightest colonies after IPTG induction were selected, pooled and used in 2 additional round of mutagenesis.

DNA sequencing and protein multiple sequence alignments

After random mutagenesis and each round of DNA shuffling, plasmid DNA (~100 ng/μL) from colonies showing greater fluorescence than wild type were sent for 5' sequencing. After the third round of shuffling, plasmid DNA corresponding to the ~45 brightest colonies were sent for 5' sequencing, of which 39 returned suitable high fidelity results for mutagenesis analysis. The mutant DNA sequencing was compared to the wild type using Vector NTI (InforMax) to generate multiple sequencing alignments. From these DNA alignments, the mutagenic rate was calculated at each stage of the protein evolution (i.e., Error prone PCR, 1X, 2X and 3X shuffling) by summing all nucleotide substitutions and dividing by the total # of mutant sequences.

Multiple protein sequence alignments were created by first translating the DNA sequences using Vector NTI, creating a FASTA file with the protein sequences from each mutant, and then performing a clustalW alignment (www.ebi.ac.uk/clustalw/). The ALN file from the clustalW alignments was loaded in JalView (www.jalview.org) and coloured by percentage identity.

Protein expression and purification of wild type and MNDA-PAAD or mutant IFI16-PAAD, RAIDD-CARD and NAC-CARD

The PAAD domain from the mutants corresponding to the brightest colonies after the 3rd round of shuffling were digested out of the GFP folding vector using NdeI and BamH1 and subsequently, cloned into PET28b using the same restriction sites. The wild type IFI16-PAAD or mutant recombinant proteins were expressed in BL21(DE3), purified by metal affinity chromatography and dialyzed into 50 mM sodium acetate, pH 4.0, as previously described [1]. MNDA-PAAD (3–88, CAH73797), RAIDD-CARD (1–92, GI:4388927) and NAC-CARD (1373–1473, AAG30288) domains were expressed and purified the same way [1], except they were dialyzed into 50 mM Tris-HCl, 100 mM NaCl, pH 8.0 after purification. All proteins were soluble and concentrated (Amicon Ultra centrifugation filter, 5000 dalton M.W. cut-off) to 10 mg/mL stock concentration before use.

Secondary structure and protein stability measurements

Secondary structure contents were determined by analysis of far UV Circular Dichroism (CD) and stability parameters by analyzing the thermal denaturation curves of the protein mutants as previously described [1].

Briefly, CD spectra of IFI16-PAAD and mutants domains were acquired on a JASCO-J-810 spectropolarimeter (equipped with a peltier type PFD-425S constant temperature cell holder). UV spectra were recorded (190-260 nm, 25°C, 0.05 cm path length cell) then converted into mean residue ellipticity (deg cm dmol^{-1}). Data were recorded using 200 nm/min scan rate, 100 mdeg sensitivity and 0.1 s response with approximately 40 μM protein concentration in each experiment. Secondary structure content of each PAAD or CARD proteins were determined by the cdPRO program (http://lamar.colostate.edu/~sreeram/CDPro/main.html) using the CDSSTR algorithm to accurately predicted alpha helical content based on a database of CD spectra from known protein structures.

IFI16-PAAD and mutants (0.04 mg/mL) were denatured thermally and monitored by CD at 222 nm to detect the loss of helical content. The data were recorded using a 10 mm quartz cell, 100 mdeg sensitivity and 0.1 s response at 1°C intervals from 20-100°C. Thermodynamic parameters were obtained by using the linear extrapolation model assuming a two-state unfolding transition model [1].

Nucleic acid binding and melting experiments

Guanine rich oligonucleotides were used in all binding and melting experiments: Guanine rich 5' 35mer (GC-5): 5'-GGAAGAAGGAA-GTGGGATCAGGATCCGCTGGCTCC-3' (complementary to GC-3); Cytosine rich 3' 35mer (GC-3)–5'-GGAGCCA-GCGGATCCTGATCCCACTTCCTTCTTCC-3' (complementary to GC-5). Oligonucleotides (500 pmol) were labelled with 50 μCi of ^{32}P gamma ATP using T4 polynucleotide kinase (1 hour, 37°C), and extracted from a 10% acrylamide native TBE gel. Recombinant proteins were incubated with ^{32}P-labeled GC-5 single stranded nucleic acid and loaded on an native 7% acrylamide gel (no SDS) to observe shifted bands (100 volts, 40 min). Visualization of free and shifted GC-5 probe was performed on a Typhoon 9410 variable mode imager. The DNA destabilization assay was performed with annealed unlabelled double stranded DNA (dsDNA) composed of GC-5 and GC-3 complementary oligonucleotides by heating to 95°C (10 min) and cooling to 20°C in 20 mM HEPES, 133 mM NaCl, pH 7.5. Destabilization of the annealed GC-5/GC-3 dsDNA was measured by calculating a delta Tm corresponding to the difference in melting point (Tm) observed by melting dsDNA in absence or presence of the protein at a molar excess of protein relative to DNA. Melting experiments were carried out on a cary UV spectrophotometry equipped with a temperature regulator, with measurement of hyperchromicity of 1 μM dsDNA and 2.5 μM IFI16-PAAD or MNDA-PAAD at UV 260 nm over the temperature range and in the corresponding dialysis buffer for each PAAD protein. After normalization of the data (0-100% UV260), the melting temperature, Tm, was calculated by recording the temperature at 50% of the hyperchromic shift.

Results and Discussion

IFI16-PAAD binds and destabilizes double stranded nucleic acid

Different recombinant proteins of the DD superfamily (Figure 1) were tested for binding to DNA: the CARD domain of NAC, the CARD domain of RAIDD (negative control) and the PAAD domain of IFI16. Only IFI16 showed binding to single-stranded DNA (ssDNA) oligonucleotide containing guanine rich sequences. To determine if IFI16 has the ability to bind to double-stranded DNA (dsDNA), GC-5/GC-3 hybridized nucleic acid were melted in the presence or absence of IFI16-PAAD (Figure 2A) or MNDA PADD (negative control, Figure 2B) in order to determine the Tm of the hyperchromic transition from double-to single-stranded DNA. The results show that only IFI16-PAAD could cause a reduction in Tm, indicating that the domain can bind to dsDNA and facilitate its destabilization, possibly through a preference for ssDNA.

Mutagenesis and DNA shuffling of IFI16-PAAD

To create the cDNA expression mutant library we performed error prone PCR and 3 rounds of DNA shuffling of the IFI16-PAAD domain cDNA (Figures 3 and 4). The complexity of the mutant library pool consisted of approximately 10^5 colonies. The mutagenesis rate of 3.76 ± 1.46 nt/kB showed an equal distribution of mutations across the entire 372 bp of the IFI16-PAAD coding region. To determine if the IFI16-PAAD mutants could increase solubility and folding of the downstream fused GFP, the library containing the mutants was sub-cloned into GFP expressing vector for bacterial expression. After plating the library and expressing the recombinant protein, UV exposure revealed that some colonies could express mutant IFI16-PAAD-GFP proteins at different intensities of GFP fluorescence, suggesting that some IFI16-PAAD mutants may improve folding and solubility of the downstream GFP protein. Based on visual examination under UV light, approximately 1.5% of the screened colonies showed increases in fluorescence, 8.5% were non-fluorescent and 90% fluoresced with the same intensity as the wild type PAAD-GFP fusion. The 60 visually brightest colonies were selected as candidates for improved folding and stability of IFI16-PAAD. To assess if the number of mutations per clone were sufficient by obtaining one to two amino acid substitutions, 10 colonies were selected for DNA sequencing. Mutations leading to "brighter"
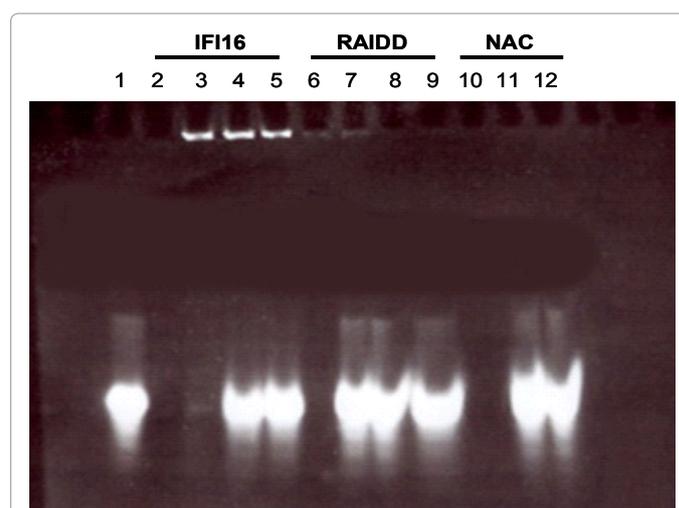
Figure 1: IFI16-PAAD but not the CARD domains from RAIDD and NAC interact with ssDNA.
Each lane has a constant amount of 200 pmol (2 μg) GC-5 radio-labelled ssDNA. Lane 1: Free oligonucleotide; Lane 2: IFI16-PAAD protein only; Lane 3: 5 fold excess IFI16-PAAD; Lane 4: 2 fold excess IFI16-PAAD; Lane 5: Equal IFI16-PAAD and ssDNA; Lane 6: RAIDD-CARD protein only; Lane 7: 10 fold excess RAIDD-CARD; Lane 8: 3;3 fold excess RAIDD-CARD; Lane 9: 1;7 fold excess RAIDD-CARD; Lane 10: NAC-CARD protein only; Lane 11: 2 fold excess NAC-CARD; Lane 12: Equal NAC-CARD and ssDNA.

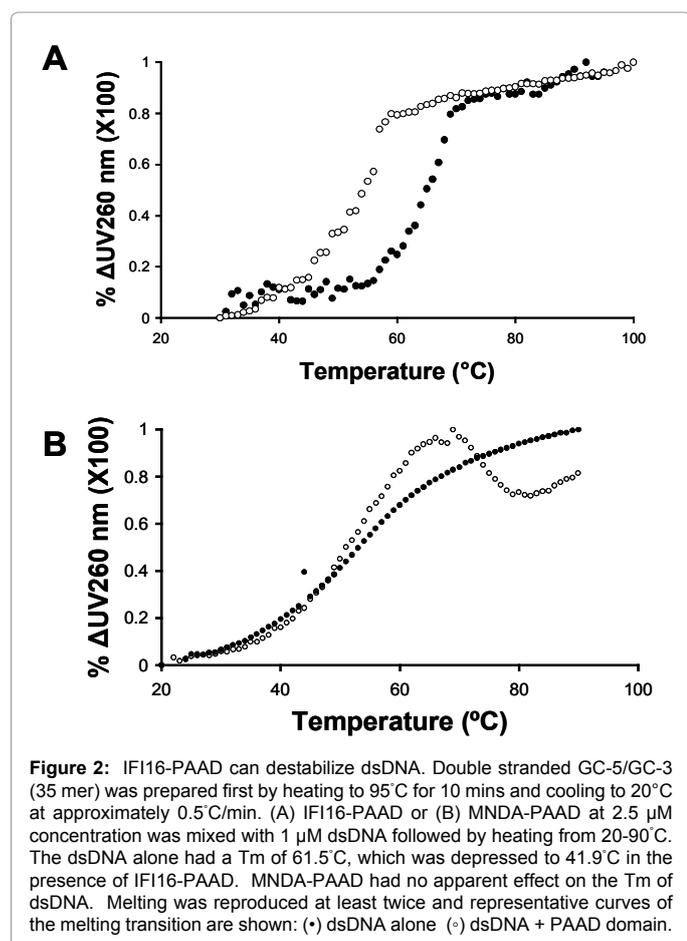**Figure 2:** IFI16-PAAD can destabilize dsDNA. Double stranded GC-5/GC-3 (35 mer) was prepared first by heating to 95°C for 10 mins and cooling to 20°C at approximately 0.5°C/min. (A) IFI16-PAAD or (B) MNDA-PAAD at 2.5 μM concentration was mixed with 1 μM dsDNA followed by heating from 20-90°C. The dsDNA alone had a Tm of 61.5°C, which was depressed to 41.9°C in the presence of IFI16-PAAD. MNDA-PAAD had no apparent effect on the Tm of dsDNA. Melting was reproduced at least twice and representative curves of the melting transition are shown: (•) dsDNA alone (○) dsDNA + PAAD domain.

fluorescence had a mutagenic rate of 4.27 ± 1.48 nt/kB, which is similar to the rate before GFP selection. A better example of the difference between wild type and increased fluorescence is demonstrated when wild type and mutant colonies are streaked on the same plate and induced for protein expression (Supplemental Figure 2).
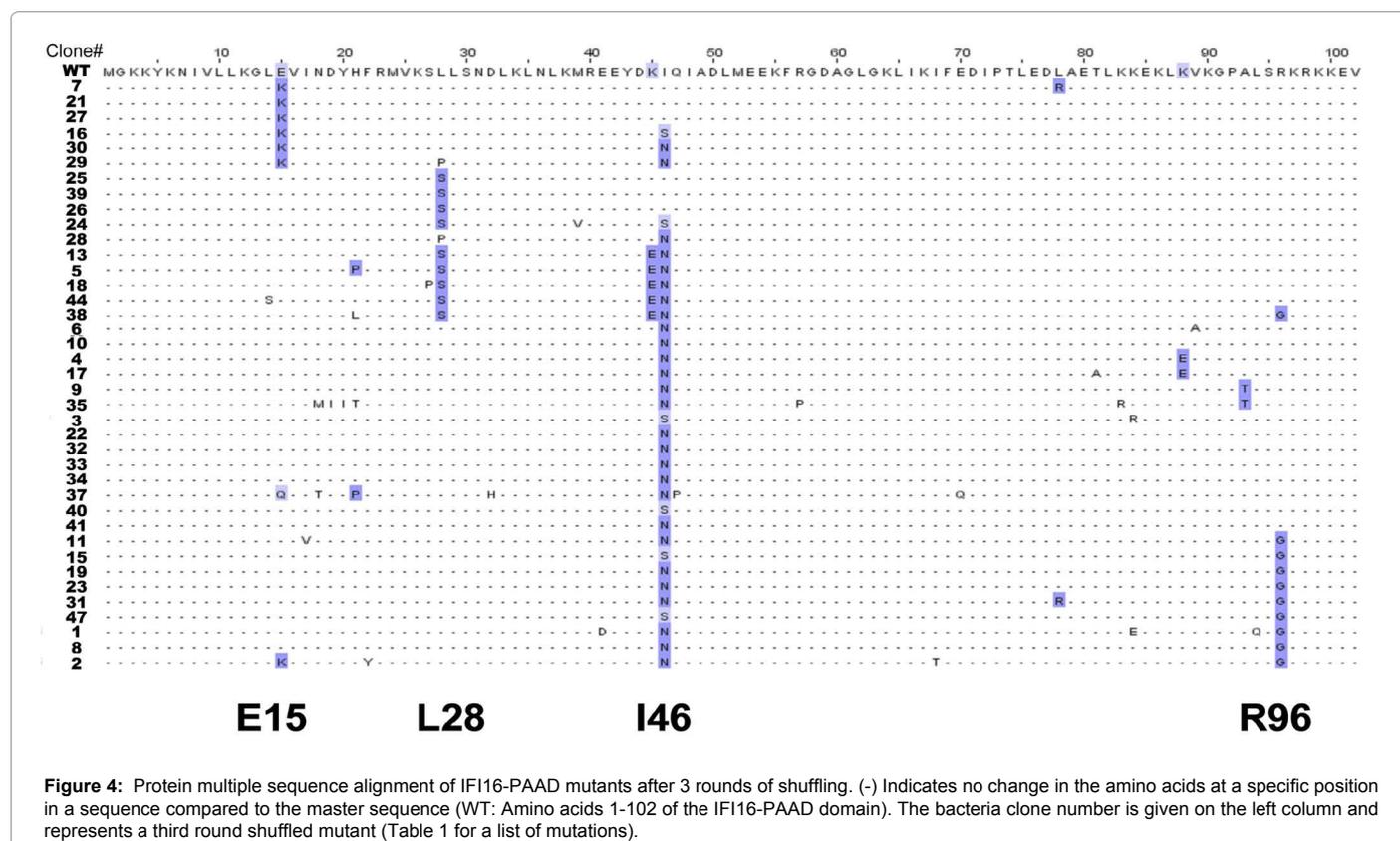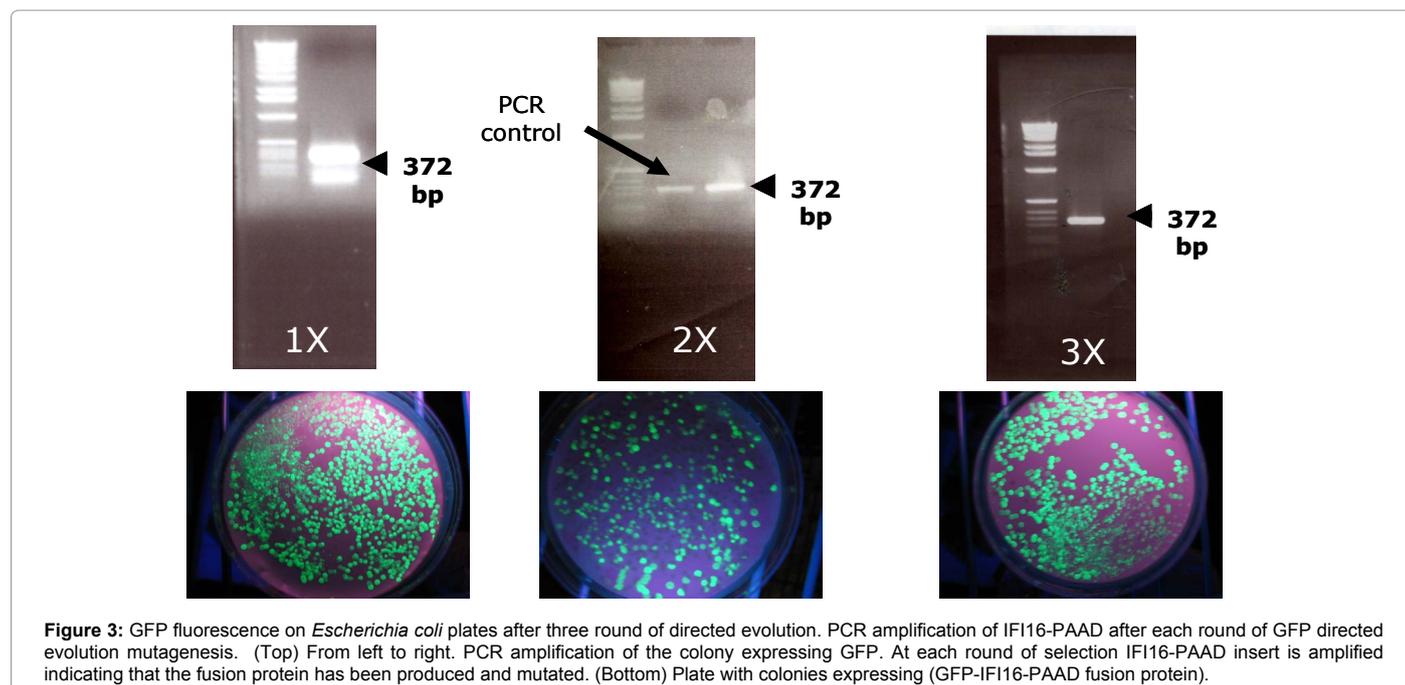
The pooled plasmid DNA (unshuffled) from the 60 brightest colonies were isolated and used in three iterative rounds of DNA shuffling to recombine favourable mutations (40 colonies picked for 2nd and 3rd round shuffling). Each round of shuffling was intended to incrementally improve the solubility and possible folding/stability of IFI16-PAAD. At each round of the DNA shuffling, the correct fragment reassembly was assessed by performing PCR amplification of the reassembled templates. Figure 3 shows the confirmation of the 372 bp PCR amplification of IFI16-PAAD after each round of shuffling, in addition to the differential fluorescence of IF16-PAAD-GFP mutants at each evolutionary stage. After the first round of shuffling, the rate of IFI16-PAAD mutagenesis was increased to 5.38 ± 2.69 nt/kB, which is higher than that of the random library and consistent with the recombination of mutations by the shuffling procedure. These mutations were separately observed and not combined in any PAAD-mutant sequences. By the third round of shuffling, the mutagenic rate of brightly fluorescent mutants was 12.0 ± 6.7 nt/kB, which represents a greater recombination frequency of mutations. We excised the PAAD domain from the pool of 3rd round PAAD-GFP fusion mutants to remove the C-terminal GFP and sub-cloned into PET28b vector containing N-terminal His-tag for sequencing and subsequent protein expression. After translation of 39 readable DNA sequences into

protein (see Materials and Methods), every 3rd round shuffled mutant had at least one amino acid change at four separate regions of IFI16-PAAD: E15, L28, I46 and R96 (multiple sequence alignment shown in Figure 4). These four amino acid positions are important for increased fluorescence since every PAAD mutant sequence contains a mutation in one or more of these regions. Unlike the sequencing results after the first round of shuffling, many PAAD mutants contained combinations of the four mutants due to the shuffling procedure. Additional mutations beyond E15K, L28S, I46N and R96G were found in some PAAD sequences and could be non-essential mutations that were not eliminated by back-crossing with the wild type. This is further suggested by the fact that the non-essential mutations are never found alone on any PAAD mutant sequence. Lastly, there were eight mutants that had the same protein sequence, suggesting that additional rounds of DNA shuffling past the third will not reveal many more mutations that lead to greater fluorescence. The mutation that occurred most frequently was I46N or I46S, suggesting that the isoleucine at position 46 of IFI16-PAAD may be critical for the partially disordered structure of the protein. This hypothesis was tested and is described in next section. I46 is found at the beginning of helix four of IFI16-PAAD while E15 and L28 are found in helix-1 and -2, respectively. Overall, the convergence of mutations onto four distinct regions of IFI16-PAAD in more than 30 sequences indicates that L28, I46 and R96 of IFI16-PAAD, when mutated, are potentially important for increased secondary structure and stability of PAAD-GFP fusion mutants.

Soluble IFI16-PAAD mutants have greater than wild type secondary structure and stability

To confirm that the increase in fluorescence of IFI16-PAAD-GFP mutants after our shuffling procedure could be translated into secondary structure and stability enhancement, we attempted to produce and purify 31 of the mutant proteins in *Escherichia coli* without the C-terminal GFP fusion partner. These mutants were chosen to include as many diverse amino acid substitutions/combinations as possible. Secondary structure and stability of the mutants were determined by scanning the far UV spectrum using circular dichroism and monitoring the 222 nm signal upon thermal denaturation. Among the 31 mutants purified and tested, 11 could not be expressed and 5 redundant sequences involving I46 mutations were removed, suggesting that the mutagenesis had converged by the third round of shuffling. Of the 15 mutants tested for both secondary structure and stability, 3 did not give proper spectra that could be analyzed by cdPRO and linear extrapolation to obtain thermodynamic parameters. Table 1 shows a summary of the folding and stability of the 12 PAAD mutants, of which robust structural and thermodynamic data could be obtained, in descending order of the best-folded mutants.

Secondary structure contents of each IFI16-PAAD mutant were assessed by far UV and cdPRO analysis in order to identify which mutations could improve folding. The far UV spectra of mutants were recorded and compared to the wild type IFI16-PAAD (Figure 5). The mutations which lead to the greatest increases in IFI16-PAAD secondary structure were I46N when coupled with other mutations (Figure 5A). Every mutant that displayed a α-helical content greater than the wild type (40.4%) (Clones 4, 6, 8, 9, 15 and 17) had one of these I46 amino acid changes. This is consistent with the high frequency of mutations at I46. As a proof of principle, these results confirm the correctness of the previously untested hypothesis that increases in GFP fluorescence can be translated into improved secondary structure of an upstream N-terminal protein. The best-folded triple mutant L28F-K88E-I46N (44.7% α-helix) had a 4.3% increase in helical content over the wild-

Figure 3: GFP fluorescence on *Escherichia coli* plates after three round of directed evolution. PCR amplification of IFI16-PAAD after each round of GFP directed evolution mutagenesis. (Top) From left to right. PCR amplification of the colony expressing GFP. At each round of selection IFI16-PAAD insert is amplified indicating that the fusion protein has been produced and mutated. (Bottom) Plate with colonies expressing (GFP-IFI16-PAAD fusion protein).

Figure 4: Protein multiple sequence alignment of IFI16-PAAD mutants after 3 rounds of shuffling. (-) Indicates no change in the amino acids at a specific position in a sequence compared to the master sequence (WT: Amino acids 1-102 of the IFI16-PAAD domain). The bacteria clone number is given on the left column and represents a third round shuffled mutant (Table 1 for a list of mutations).

type. The I46N mutant alone was slightly less folded (38.9%) but had an increase in stability according to $\Delta G_{folding}$ 25°C. Several mutations lead to decreased secondary structure compared to the wild type (clones 11, 16, 22, 27, 29 and 30, Figure 5B). Many of these mutants had the E15K or L28F/P mutation, which lead to a collapse of IFI16-PAAD secondary

structure when combined with others. The least folded mutants were clones 11 and 29 which both had less than 23% α-helix.

The stability of IFI16-PAAD mutants were obtained by recording their thermal denaturation using CD at 222 nm (Figure 6), and fitting the curves to a two state unfolding model by linear extrapolation. The

Figure 5: Secondary structure analysis of IFI16-PAAD mutants. (A) Representative far UV spectra of IFI16-PAAD mutants with increased secondary structure: (solid line) Wild type IFI16-PAAD; (short dashed line - -) Clone 4; Clone 6; Clone 8; Clone 9; Clone 15; Clone 17; Clone 22; Clone 24. (B) Representative far UV spectra of IFI16-PAAD mutants with decrease secondary structure: (solid line) Wild type IFI16-PAAD; Clone 5; Clone 11; Clone 16; Clone 28; Clone 29; Clone 30.

Figure 6: Thermodynamic stability of IFI16-PAAD mutants. Thermal denaturation was monitored from 20°C to 98°C at 222 nm Far UV CD and presented as fraction unfolded protein. (A) Representative thermal denaturation of more stable IFI16-PAAD mutants: (solid line) Wild type IFI16-PAAD; (short dashed line - -) Clone 4; Clone 6; Clone 8; Clone 9; Clone 17; Clone 22. (B) Representative thermal denaturation spectra of IFI16-PAAD mutants with decreased stability than wild-type: (solid line) Wild type IFI16-PAAD; Clone 11; Clone 15; Clone 16; Clone 27; Clone 29; Clone 30.

most stable mutant was clone 4 (I46N, L28F, K88E) with a $\Delta G_{folding}$ 25°C=-2.44 kcal/mol, which represents a 38% increase in stability over the wild type protein. The improvement in stability of this mutant is consistent with its improved secondary structure content. Other clones with increased stability were clones 6, 8, 9, 17 and 22(6:I46N, V89E; 8:I46N, R96G; 9: I46N, A93T 17:I46N, T81A, K88E, 22:I46N). As expected, these clones have the I46N mutation, which as seen above, is responsible for increased secondary structure except clone 22 which had I46N alone. In contrast, every mutant with the E15K mutation had decreased thermal stability, whether it was coupled to other mutations or not. This was consistent with the finding that the E15K mutation leads to dramatically reduced secondary structure than wild type, as

assessed by far UV CD measurements described above. Notably, clones 16, 29 and 30 have both the E15K mutation as well as one of the I46 substitutive mutation N46 or S46 but have dramatically decreased stability (between -0.38 and -0.75 kcal/mol), suggesting that the E15K mutation is dominant for decreased stability.

dsDNA destabilization by IFI16-PAAD mutants

Melting curve depression assays of dsDNA in the presence of different IFI16-PAAD mutants with either increased (clones 4, 6, 22) or decreased (clone 29) secondary structure content indicate a correlation between greater protein folding and destabilization (ΔTm) of dsDNA (Figure 7A). The correlation was quantified by plotting the
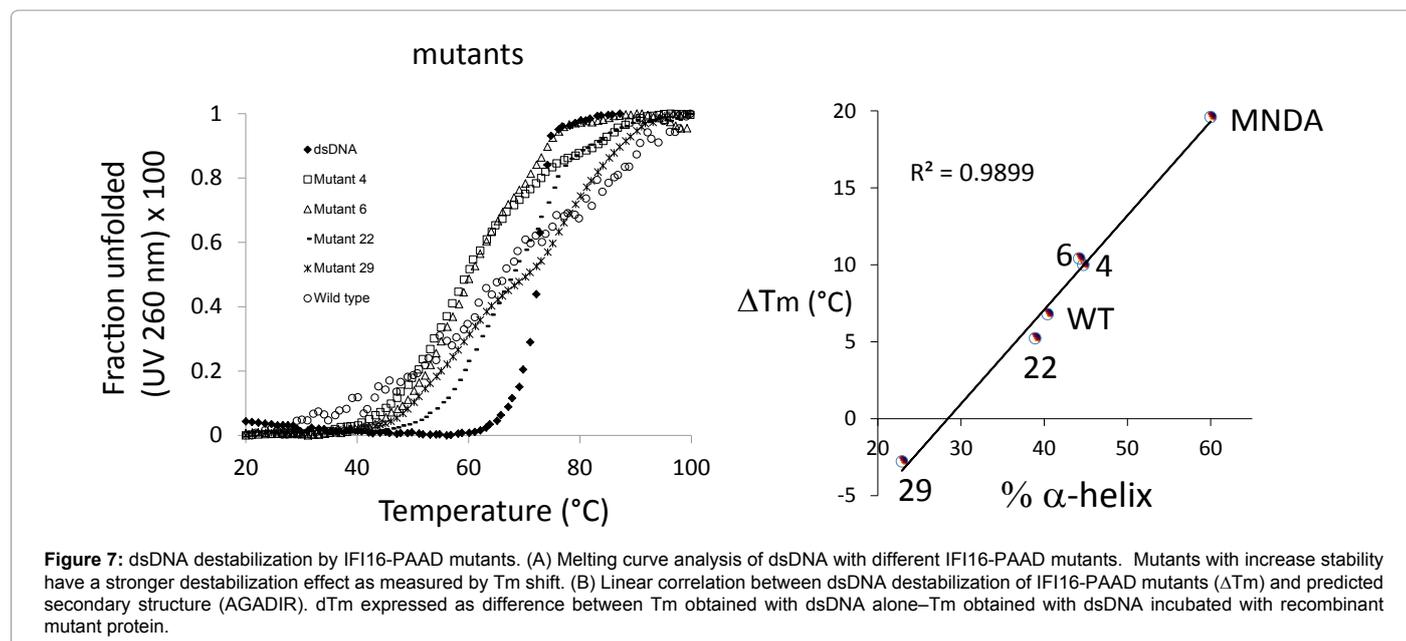
Figure 7: dsDNA destabilization by IFI16-PAAD mutants. (A) Melting curve analysis of dsDNA with different IFI16-PAAD mutants. Mutants with increase stability have a stronger destabilization effect as measured by Tm shift. (B) Linear correlation between dsDNA destabilization of IFI16-PAAD mutants (ΔTm) and predicted secondary structure (AGADIR). dTm expressed as difference between Tm obtained with dsDNA alone–Tm obtained with dsDNA incubated with recombinant mutant protein.

change in Tm versus the % alpha helical content, establishing the linear relationship between these two parameters (Figure 7B). These results suggest that the native, unfolded state of IFI16-PAAD is critical to maintain the structure and integrity of the duplex dsDNA.

Discussion

In this study, a combination of error prone PCR, DNA shuffling and GFP selection was used to find mutations in IFI16-PAAD that could re-order the disordered regions of the protein domain. Since increased GFP fluorescence is usually attributed to solubility, we have shown that *in vitro* GFP directed evolution can be used to identify mutants with increased secondary structure and stability, but not all the mutants obtained in our selection were more stable and better folded than the wild type IFI16-PAAD.

It should be pointed out that the brighter than wild type GFP fluorescence seen in colonies expressing PAAD-GFP fusion mutants could also be due to greater over-expression of the fusion protein. Thus, the selection of false positive mutants is a possibility. To address this problem, we tested many 3rd round shuffled mutants for greater stability and secondary structure after protein expression without the C-terminal GFP fusion partner, as shown in Tables 1 and 2.

The four amino acid positions of IFI16-PAAD (E15K, L28P/F, I46N, R96G) selected by the directed evolution procedure do not lie in helix-3 (Figure 8). All these mutants seem to affect helix propensity using the secondary structure prediction software AGADIR (http://www.embl-heidelberg.de/ExternalInfo/serrano/softwarePage.html), but the helix-2 and helix-3 regions are strongly predicted to be α-helical by the program for the wild type and mutant. This prediction contradicts the current NMR model of PAAD domains from NAC and MNDA that show disorder in helix-2 and -3. These data may imply that helix-2 and -3 have the ability to fold into α-helix on their own but are restrained in the context of the full PAAD domain. The E15K and L28F/P mutations are found in helix-1 and -2, respectively, while the I46N mutation lies at the beginning of helix-4. The latter mutation is predicted to be involved in helix capping by the software AGADIR. Effectively, the asparagine residue in the I46N mutation might promote helix formation in the

disordered region by forming favourable polar interactions [26,27] with either the helix-1 or -3. Since we have already established that IFI16-PAAD has a partially folded intermediate structure [1], it is possible that the refolding of helix-3 can explain the dramatic increases in secondary structure seen in clones 4, 6, 8 and 9. The I46N/S mutation was present in every mutant that had improved folding; I46 must be a critical position for the native structure of IFI16-PAAD. In contrast, many of the mutations lead to a collapse of IFI16-PAAD secondary structure, namely E15K, illustrates that improved solubility does not always translate into improved folding and stability. The I46N/S mutations do not occur within the PAAD family members and the DD superfamily in a multiple sequence alignment of the PAAD domain sequence of IFI16 with protein sequences from DD, CARD, DED or PAAD families (results not shown). This alignment further suggest that our mutagenesis has not selected naturally occurring mutations during evolution but *de novo* mutations that increase stability and secondary structure of the protein.

IFI16-PAAD mutants that had increased folding correlated well with those having increased stability (clones 4, 6, 8, 9, 15 and 17), suggesting that the directed evolution procedure can select for both traits simultaneously.

Protein appears to simplify their folding pathway by breaking it down into sequential steps. The search for a simple and unique mechanism of protein folding has lead to 3 different models that may not be exclusive: (i) The nucleation-growth model, where tertiary structure propagates rapidly from an initial nucleus of local secondary structures, but this model fails to predict folding intermediates. (ii) The framework model, where secondary structure folds first, followed by docking of the preformed secondary structural units to yield the native, and folded protein. (iii) The hydrophobic collapse model, where the hydrophobic collapse drives the compaction of the protein so that the folding can take place in confined volume to limit the conformational search space leading to the native state [17,18,28-42].

The rate and mechanism of protein folding of all α-helix proteins that fold by a two state mechanism is dependent upon their content of secondary structure and stability [43,44]. Effectively, it was observed

Clone	% helix	Tm (°C)	ΔH (kcal/mol)	ΔS (kcal/mol/K)	(kcal/mol)	n	Mutation 1	Mutation 2	Mutation 3
4	44.7 ± 4.6	47.80 ± 1.18	-34.80 ± 4.87	-0.109 ± 0.015	-2.44 ± 0.51	3	I46N	L28F	K88E
6	44.2 ± 1.3	44.78 ± 0.75	-29.41 ± 2.64	-0.091 ± 0.008	-1.84 ± 0.18	3	I46N	V89E	-
9	44.0 ± 4.3	45.03 ± 2.86	-28.14 ± 2.32	-0.089 ± 0.007	-1.77 ± 0.34	3	I46N	A93T	-
8	42.9 ± 0.8	45.18 ± 0.75	-25.60 ± 3.92	-0.080 ± 0.012	-1.63 ± 0.27	3	I46N	R96G	-
17	41.5 ± 2.2	46.70 ± 0.90	-26.94 ± 1.65	-0.084 ± 0.005	-1.83 ± 0.18	3	I46N	T81A	K88E
15	41.2 ± 0.9	39.47 ± 1.21	-23.44 ± 2.47	-0.076 ± 0.007	-1.36 ± 0.40	3	I46S	R96G	-
Wild Type	40.4 ± 0.2	42.43 ± 0.50	-27.17 ± 2.50	-0.086 ± 0.008	-1.52 ± 0.12	3	none	-	-
16	39.8 ± 0.5	33.63 ± 0.88	-27.34 ± 8.95	-0.089 ± 0.029	-0.89 ± 0.20	3	I46S	E15K	-
30	39.8 ± 1.0	35.72 ± 0.58	-25.93 ± 5.27	-0.084 ± 0.017	-1.14 ± 0.07	3	I46N	E15K	-
22	38.9 ± 0.9	46.74 ± 2.41	-28.32 ± 3.13	-0.089 ± 0.009	-1.94 ± 0.41	3	I46N	-	-
27	38.6 ± 0.5	43.41 ± 0.99	-24.94 ± 8.10	-0.079 ± 0.026	-1.45 ± 0.44	3	-	E15K	-
29	22.9 ± 2.0	31.38 ± 0.61	-20.51 ± 1.04	-0.067 ± 0.003	-0.77 ± 0.07	3	I46N	E15K	L28P
11	20.1 ± 4.2	31.48 ± 2.11	-58.78 ± 56.52	-0.193 ± 0.185	-1.36 ± 1.41	3	I46N	R96G	I17V

Table 1: Secondary structure prediction and thermodynamic stability of 3rd round shuffled IFI16-PAAD mutants. Text shaded in red refers to predicted stabilizing mutations while blue text is for destabilizing mutations. The percent helix was calculated by cdPRO analysis of far UV spectra of each clone. The thermal parameters ΔH, ΔS, Tm and ΔG$_{folding}$ 25°C were calculated by Van't Hoff analysis of thermal denaturation monitored by CD at 222 nm. The Van't Hoff plot was generated as previously described [1].

Clone	ΔG$_{folding}$ 25ºC (kcal/mol)	ΔG$_{H2O}$ 25ºC (kcal/mol)	Tm (ºC)	ΔH (kcal/mol)	ΔS (kcal/mol/K)	m (kcal/mol/M)	Cm (M)	ΔASA (Å2)	Mutation 1	Mutation 2	Mutation 3
4	-2.44 ± 0.51	-2.22	47.80 ± 1.18	-34.80 ± 4.87	-0.109 ± 0.015	-1.12	1.99	1186	I46N	L28F	K88E
6	-1.84 ± 0.18	-2.26	44.78 ± 0.75	-29.41 ± 2.64	-0.091 ± 0.008	-1.12	2.01	1186	I46N	V89E	-
Wild Type	-1.52 ± 0.12	-1.75	42.43 ± 0.50	-27.17 ± 2.50	-0.086 ± 0.008	-1.7	1.03	3822	none	-	-
22	-1.94 ± 0.41	-1.56	46.74 ± 2.41	-28.32 ± 3.13	-0.089 ± 0.009	-0.96	1.64	3550	I46N	-	-
29	-0.77 ± 0.07	-1.09	31.38 ± 0.61	-20.51 ± 1.04	-0.067 ± 0.003	-0.76	1.44	2640	I46N	E15K	L28P

Table 2: Thermodynamic of folding and surface accessible area parameters of the different IFI16-PAAD mutants. Thermodynamic parameters for IF16-PAAD wild type and clones 4, 6, 22, and 29 are presented as in Table 1 but with chemical denaturation parameters (ΔG^{H2O}-energy of unfolding, m-unfolding constant, C_m-[Gdn-HCl] of 50% unfolding, ΔASA-Calculated accessible surface area of folded protein) determined by monitoring helical content (CD 222 nm) following chemical denaturation with Guanidine Hydrochloride (Gdn-HCl) as described in Dalal et al. [1].
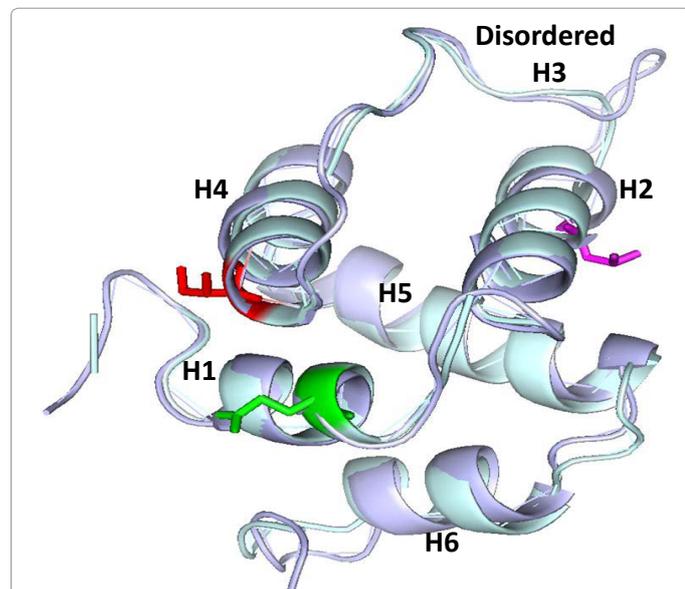

Figure 8: Comparative modelling of IFI16-PAAD on NALP1-PAAD. Superimposition of NALP1-PAAD domain (light green) (pdb code: 1PN5) with the IFI16-PAAD model generated with Modeler software (light purple) (RMSD=0.88 Å) (1). The stick representation of amino acids E15, L28 and I46 are shown in green, pink and red respectively. Helix numbers are shown (H1 to H6) with the disordered helix-3 identified.

that an increase of the intrinsic stability of helices by substitution of solvent-exposed residues may stabilize the native protein [39,40]. These results are consistent with our study since the mutations that increase stability and secondary structure of IFI16-PAAD involve amino acid side chains that are exposed to solvent. However, the effect of these mutations on the rate of folding of IFI16-PAAD remains to be established by kinetics studies.

The modification of hydrophobic amino acids side chains exposed to solvent or located in the core of the protein have also been shown to affect folding and stability in other studies. Some of the mutations obtained in this work may change hydrophobicity by which secondary structure and stability may be modulated. The effect of these mutations on secondary structure and protein stability are difficult to predict and has been shown to be dependent of the system studied [39-44].

The EMSA (Figure 1) and dsDNA destabilization (Figures 2 and 7) experiments suggest that IFI16-PAAD interacts with nucleic acids, a protein-ligand interaction that has not been reported for this domain. However, it is well known that the full-length IFI16 protein and related transcriptional regulators that belong to the Hematopoietic Interferon-inducible nuclear proteins with 200 amino-acids repeats (HIN-200) can bind to DNA and include proteins such as MNDA, AIM-2, ASC and NALP1 [45]. The HIN-200 protein domains in these family members are found C-terminal to their PAAD domains and facilitate diverse functions as DNA binding [46,47] and activation of the inflammasome assembly [48,49]. Conversely, more work is required to fully understand the role and mechanism of the DNA-binding function of IFI16-PAAD, and is under active investigation in our laboratory.

In conclusion, we have identified mutations that can increase stability and secondary structure of IFI16-PAAD suggesting that the partially folded state of IFI16-PAAD may be determined at least in part by its primary sequence. A detailed analysis of the folding pathway of IFI6-PAAD using synthetic peptide, mutagenesis, thermal/urea unfolding transition and folding kinetics approaches [18,19] is needed to determine which folding model describe best the partially folded structure of IFI16-PAAD observed in solution. These studies may

also allow us to determine if this conformation constitutes the native state of the protein which may be essential to accommodate diverse interacting partners in physiological conditions.

Acknowledgment

References

1. Dalal K, Pio F (2006) Thermodynamics and stability of the PAAD/DAPIN/PYRIN domain of IFI-16. FEBS Lett 580: 3083-3090.

2. Liu T, Rojas A, YeY, Godzik A (2003) Homology modeling provides insights into the binding mode of the PAAD/DAPIN/pyrin domain, a fourth member of the CARD/DD/DED domain family. Protein Sci 12: 1872-1881.

3. Liang H, Fesik SW (1997) Three-dimensional structures of proteins involved in programmed cell death. J Mol Biol 274: 291-302.

4. Pawlowski K, Pio F, Chu Z, Reed JC, Godzik A (2001) PAAD-a new protein domain associated with apoptosis, cancer and autoimmune diseases. Trend Biochem Sci 26: 85-87.

5. Bertin J, DiStefano PS (2000) The PYRIN domain: a novel motif found in apoptosis and inflammation proteins. Cell Death Differ 7: 1273-1274.

6. Fairbrother WJ, Gordon NC, Humke EW, O'Rourke KM, Starovasnik MA, et al. (2001) The PYRIN domain: a member of the death domain-fold superfamily. Protein Sci 10: 1911-1918.

7. Anonymous (1997) A candidate gene for familial Mediterranean fever. The French FMF Consortium. Nat Genet 17: 25-31.

8. Stehlik C, Fiorentino L, Dorfleutner A, Bruey JM, Ariza EM, et al. (2002) The PAAD/PYRIN-family protein ASC is a dual regulator of a conserved step in nuclear factor κB activation pathways. J Exp Med 196: 1605-1615.

9. Karin M, Lin A (2002) NF-kappaB at the crossroads of life and death. Nat Immunol 3: 221-227.

10. Reed JC, Doctor K, Rojas A, Zapata JM, Stehlik C, et al. (2003) Comparative analysis of apoptosis and inflammation genes of mice and humans. Genome Res 13: 1376-1388.

11. Aglipay JA, Lee SW, Okada S, Fujiuchi N, Ohtsuka T, et al. (2003) A member of the Pyrin family, IFI16, is a novel BRCA1-associated protein involved in the p53-mediated apoptosis pathway. Oncogene 22: 8931-8938.

12. Johnstone RW, Kerry JA, Trapani JA (1998) The human interferon-inducible protein, IFI 16, is a repressor of transcription. J Biol Chem 273: 17172-17177.

13. Hiller S, Kohl A, Fiorito F, Herrmann T, Wider G, et al. (2003) NMR structure of the apoptosis- and inflammation-related NALP1 pyrin domain. Structure 11: 1199-1205.

14. Liepinsh E, Barbals R, Dahl E, Sharipo A, Staub E, et al. (2003) The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition. J Mol Biol 332: 1155-1163.

15. Qin H, Srinivasula SM, Wu G, Fernandes-Alnemri T, Alnemri E, et al. (1999) Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature 399: 549-557.

16. Chu ZL, Pio F, Xie Z, Welsh K, Krajewska M, et al. (2001) A novel enhancer of the Apaf1 apoptosome involved in cytochrome c-dependent caspase activation and apoptosis. J Biol Chem 276: 39-9245.

17. Uversky V (2002) Craking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't. FEBS Lett 514: 181-183.

18. Daggett V, Fersht A (2003) Is there a unifying mechanism for protein folding. Trends Biochem Sci 28: 18-25.

19. Otten LG, Quax WJ (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng 22: 1-9.

20. DeGrado WF (1997) Proteins from scratch. Science 278: 80-81.

21. Dawson MJ, Trapani JA (1995) The interferon-inducible autoantigen, IFI 16: localization to the nucleolus and identification of a DNA-binding domain. Biochem Biophys Res Commun 214: 152-162.

22. Waldo GS (2003) Improving protein folding efficiency by directed evolution using the GFP folding reporter. Methods Mol Biol 230: 343-359.

23. Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17: 691-695.

24. Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, et al. (2005) Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis. J Struct Funct Genomics 6: 113-119.

25. Zhao H, Arnold FH (1997) Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Res 25: 1307-1308.

26. Shoemaker KR, Kim PS, Brems DN, Marqusee S, York EJ, et al. (1985) Nature of the charged-group effect on the stability of the C-peptide helix. Proc Natl Acad Sci USA 82: 2349-2353.

27. Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL (1987) Tests of the helix dipole model for stabilization of alpha-helices. Nature 326: 563-567.

28. Levinthal C (1969) How to fold graciously. In: Monticello IL (ed.) Mossbauer Spectroscopy in Biological Systems. University of Illinois Press, USA pp: 22-24.

29. Wetlaufer DB (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci USA 70: 697-701.

30. Kim PS, Baldwin RL (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51: 459-489.

31. Kim PS, Baldwin RL (1990) Intermediates in the folding reactions of small proteins. Annu Rev Biochem 59: 631-660.

32. Ptitsyn OB, Rashin AA (1975) A model of myoglobin selforganization. Biophys Chem 3: 1-20.

33. Ptitsyn OB (1987) Protein folding: hypotheses and experiments. J Protein Chem 6: 273-293.

34. Karplus M, Weaver DL (1976) Protein folding dynamics. Nature 260: 404-406.

35. Schellman JA (1955) Stability of hydrogen bonded peptide structures in aqueous solution. Compt Rend Lab Carlsberg Ser Chim 29: 230-259.

36. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14: 1-63.

37. Tanford C (1962) Contribution of hydrophobic interactions to the stability of globular confirmation of proteins. J Am Chem Soc 84: 4240.

38. Baldwin RL (1989) How does protein folding get started? Trends Biochem Sci 14: 291-294.

39. Brown JE, Klee (1971) WA Helix–coil transition of the isolated amino terminus of ribonuclease. Biochemistry 10: 470-476.

40. Bierzynski A, Kim PS, Baldwin RL (1982) A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A. Proc Natl Acad Sci USA 79: 2470-2474.

41. Epand RM, Scheraga HA (1968) The influence of long range interactions on the structure of myoglobin. Biochemistry 7: 2864-2872.

42. Taddei N, Chiti F, Fiaschi T, Bucciantini M, Capanni C, et al. (2000) Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. J Mol Biol 300: 633-647.

43. Taddei N, Capanni C, Chiti F, Stefani M, Dobson CM, et al. (2001) Folding and aggregation are selectively influenced by the conformational preferences of the alpha-helices of muscle acylphosphatase. J Biol Chem 276: 37149-37154.

44. Cranz-Mileva S, Friel CT, Radford SE (2005) Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9. Protein Eng Des Sel 18: 41-50.

45. Johnstone RW, Wei W, Greenway A, Trapani JA (2000) Functional interaction between p53 and the interferon-inducible nucleoprotein IFI 16. Oncogene 19: 6033-6042.

46. Yan H, Dalal K, Hon BK, Youkharibache P, Lau D, et al. (2008) RPA nucleic acid-binding properties of IFI16-HIN200. Biochim Biophys Acta 1784: 1087-1097.

47. Ni X, Ru H, Ma F, Zhao L, Shaw N, et al. (2016) New insights into the structural basis of DNA recognition by HINa and HINb domains of IFI16. J Mol Cell Biol 8: 51-61.

48. Jin T, Perry A, Smith P, Jiang J, Xiao TS (2013) Structure of the absent in melanoma 2 (AIM2) pyrin domain provides insights into the mechanisms of AIM2 autoinhibition and inflammasome assembly. J Biol Chem 288: 13225-13235.

49. Jin T, Perry A, Jiang J, Smith P, Curry JA, et al. (2012) Structures of the HIN domain: DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor. Immunity 36: 561-571.