

Published in final edited form as:

J Mol Biol. 2008 January 18; 375(3): 720–734.

Restriction endonuclease inhibitor IPI* of bacteriophage T4:

A novel structure for a dedicated target

Dalin Rifat, Nathan T. Wright, Kristen M. Varney, David J. Weber, and Lindsay W. Black
Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108
N. Greene St., Baltimore, MD 21201-1503, USA

SUMMARY

Phage T4 protects its DNA from the two gene encoded *gmrS/gmrD* (glucose modified hydroxymethylcytosine (gHMC) restriction endonuclease) (CT), of pathogenic *E. coli* CT596, by injecting several hundred copies of the 76 amino acid residue nuclease inhibitor, IPI*, into the infected host. Here, the three-dimensional solution structure of mature IPI* is reported as determined by nuclear magnetic resonance (NMR) techniques using 1290 experimental NOE and dipolar coupling constraints (~17 constraints/residue). Close examination of this oblate-shaped protein structure reveals a novel fold consisting of two small β -sheets (β 1: B1, B2; β 2: B3-B5), flanked at the N- and C-termini by alpha helices (H1 & H2). Such a fold is very compact in shape, and allows ejection of IPI* through the narrow 30Å portal and tail tube apertures of the virion without unfolding. Structural and dynamic measurements identify an exposed hydrophobic knob that is a putative *gmrS/gmrD* binding site. A single gene from the uropathogenic *E. coli* UT189, which codes for a *gmrS/gmrD*-like fusion protein (~90% identity to the heterodimeric CT enzyme) has evolved IPI* inhibitor immunity. Analysis of the *gmrS/gmrD* restriction endonuclease enzyme family and its IPI* family phage antagonists reveals an evolutionary pathway that has elaborated a surprisingly diverse and specifically fitted set of co-evolving attack and defense structures.

Keywords

Bacteriophage T4; NMR; restriction endonuclease inhibitor

INTRODUCTION

Host-viral exclusion and anti-exclusion mechanisms co-evolve competitively. Among numerous strategies to prevent phage infection, bacteria synthesize restriction enzymes that digest entering phage DNAs while bacteriophages deploy counterstrategies to circumvent such enzymes. Remarkable specificity is displayed by the dispensable internal protein I (*ip1*) gene that is required for T4 phage to infect successfully the pathogenic *Escherichia coli* isolate CT596¹. In this host, the injected T4 DNA degrades into acid-soluble fragments only in the absence of the injected mature gene product IPI*. An *E. coli* CT596 genomic library identified adjacent *glucose modified restriction* genes *gmrS* (942bp) and *gmrD* (708bp) that are necessary

Corresponding Authors: Lindsay W. Black, Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108 N. Greene St., Baltimore, MD 21201, tel: (410) 706-3510, fax: (410) 706-8297, e-mail: lblack@umaryland.edu David J. Weber, Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108 N. Greene St., Baltimore, MD 21201, Telephone: (410) 706-4354, Fax: (410) 706-0458, Email: dweber@umaryland.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and sufficient to confer upon a host the ability to exclude infection by T4 *ip1*⁻ and other glucosyl-hydroxymethylcytosine (g-HMC) T-evens lacking the *ip1* gene². However, the clones allow infection by phages with non-glucosylated cytosine DNA that lack the *ip1* gene (e.g. lambda, T7, RB49). Thus, g-HMC phages C16 and RB69 that contain the *ip1* gene grow on the clone, whereas g-HMC phages such as T6, T2, RB15, and RB70 (and T4 *ip1*⁻ mutants) that lack the *ip1* gene do not. Moreover a plasmid expressing the mature, encapsidated *ip1* gene product, IPI*, allows T4 *ip1*⁻ and the other phages lacking *ip1* to grow on *E. coli* strains carrying the cloned *gmrS/gmrD* genes; whereas, an isogenic plasmid expressing the mutant *ip1*⁻ product, IPI* (A40T, E53K) does not².

Two proteins, GmrS (36kDa) and GmrD (27kDa), encoded by the *gmrS/gmrD* genes were purified and found to be inactive separately, but together degraded several different g-HMC modified DNAs (T4, T2 and T6) to low molecular mass products (<500bp); whereas, no activity was observed against non-modified DNA including unmodified T4 cytosine (C) or non-glucosylated T4 HMC DNA³. The enzyme activity could be inhibited by IPI*, which binds to the GmrSD proteins. Therefore both *in vivo* and *in vitro* studies show that IPI* specifically inhibits the activity of GmrSD to prevent the g-HMC T4 DNA from being digested.

The *ip1* gene locus has been investigated in the genomes of a number of g-HMC T-even phages related to T4⁴. Five new ORFs, *ip5* to *ip9*, were sequenced within the same flanking sequences 57.B and ORF2 as *ip1*. Despite the overall sequence divergence of these putative *ip1* homolog genes, they all encode for 11-19 amino acid residues in the N-terminal region, which has a remarkably conserved signature sequence. It is known that this consensus capsid targeting sequence (CTS) allows incorporation of an attached protein into the phage head⁵. The CTS is then cleaved from the precursor protein distal to an IXE or LXE amino acid sequence by the gene 21 prohead maturation protease⁵. This allows the processed mature protein to be injected together with the DNA into an infected host. A striking characteristic of the *ip1* gene locus among these phages is its unusual degree of gene expansion and genetic polymorphism in comparison to the highly conserved structural genes of these phages. For example, the *ip1* locus of phages RB70, T6, T2, K3, and RB15 contains *ip5*, *ip5+ip6*, *ip5+ip7*, *ip7+ip8+ip9* and *ip5+ip7+ip8+ip9*, respectively, all genes bearing the signature CTS N-terminal motif⁴. These results suggest that many new *ip1* genes and combinations of these remain to be identified.

The diversity of the *ip1* locus genes may be evolutionarily linked to the diverse sugar-modifications of the 100% HMC DNA containing T-even phages^{6, 7}. Thus, phage T4 completely glucosylates its HMC DNA in 30% α -linkage and 70% β -linkage; T2 in 70% α -linkage and 5% β -linkage in gentiobiose form (a second β -glucose bound to a pre-existing α -linkage); and T6 has 3% α -linkage and 72% gentiobiose DNA. T-even phages modify their DNAs with glucosylated HMC (glc-HMC) to avoid degradation by type I, II and III restriction endonucleases that generally are incapable of digesting such DNAs (summarized in Figure 7). GmrSD is the first discovered modification-dependent restriction (MDR) endonuclease (type IV) that specifically digests sugar-modified HMC DNA³. Different sugar modifications and distributions on HMC residues could account for the selection of other GmrSD family members in the process of the co-evolution between T-even phages and host bacteria. Homology searches show that the *gmrS/gmrD* genes display a remarkable degree of homology with numerous ORFs of unknown function in diverse bacteria. Chief among these, uropathogenic *E. coli* (UPEC) UTI89 and avian pathogenic *E. coli* (APEC) (99% identity to UPEC) contain a single ORF that has 89% and 93% identity with *gmrS* and *gmrD* of CT596, respectively, strongly suggesting that they are members of the GmrSD restriction endonuclease enzyme family.

How the *ip1* gene locus and *gmrS/gmrD* gene products interact with each other and with sugar-modified HMC DNAs of T-even phages is unclear. IPI* (and the other *ip1* gene locus products) display no significant homology with any database protein. Therefore in this study we applied

NMR spectroscopy to determine the three dimensional structure of IPI* in order to understand both evolution and the structure/function relationship of these anti-restriction proteins. Here we report (1) the novel protein fold of IPI* that reveals regions likely involved in interaction with the GmrSD enzyme; (2) how on the basis of this structure an amino acid replacement in IPI* leads to loss of function and absence of the protein from the virion; and (3) how, unlike the *E. coli* CT596 heterodimeric subunit CT enzyme, the closely related (~90% sequence identity) GmrSD UT enzyme from *E. coli* UT189 has evolved a single polypeptide structure that renders it immune to IPI*.

RESULTS

NMR resonance assignments and secondary structure of IPI*

NMR experiments were performed on a recombinant form of IPI* as necessary to sequence specifically assign its ^1H , ^{13}C , and ^{15}N resonances. Initial 2D ^1H - ^{15}N - usually we specify this as HSQC heteronuclear single quantum coherence (HSQC) spectra - were collected in a pH range from 7.5 to 4.5. While most of the HSQC correlations were not perturbed or shifted very slightly with decreasing pH, twelve peaks appeared, and no peaks weakened. Since lowering the pH was beneficial and did not cause large changes in chemical shifts, as measured in the HSQC spectrum, it was concluded that no global structural changes occurred, and the remainder of the NMR studies were completed at pH 4.8. The 2D ^1H - ^{15}N HSQC experiment performed under these conditions revealed a well-dispersed spectrum indicative of a highly structured protein (Figure 1A). This was confirmed with ^1H - ^{15}N heteronuclear NOE data, where only six residues in the interior of the protein were found to have low NOE values (NOE<0.68) (Figure 1B). In order to determine the solution structure of IPI*, it was then necessary to unambiguously assign the resonances *a priori* using 2D, 3D, and 4D NMR experiments (Table 1).

The mature form of IPI*, studied here, contains 76 residues, of which two are proline residues. Of the 74 residues expected to be observed in the HSQC spectrum, 72 were present (Figure 1A). Two residues (H62 and G35) are absent due to exchange-broadening. The amide of the side-chain of W38, along with the side chain of N74, is visible in the HSQC (indicated by parentheses); however, the side-chain ^1H - ^{15}N resonances for N11 were not present. Representative long-range resonances indicative of β -sheet structures and dipolar couplings are shown in Figures 1C-1E. Overall, all of the observable ^1H , ^{13}C , and ^{15}N resonances of IPI* were fully assigned unambiguously using a standard combination of multidimensional heteronuclear NMR experiments (Table 1) and deposited in the BioMagResBank (accession number: [15438](#)).

Nuclear Overhauser effect (NOE) data, chemical shift data, and hydrogen exchange data indicate that IPI* has many regions of defined secondary structure (Figure 2). Chief among these are two long alpha helical regions extending from residues 3 to 10 (helix 1) and from residues 65 to 76 (helix 2). These two N- and C-terminal helices flank two small β -sheets (β 1 and β 2). The first of these, β 1, is a small two-stranded sheet comprising two strands (B1, residues 19-21; B2, residues 26-28) separated by a Type 1 β -turn (T1; residues 22-25). The second β -sheet, β 2, has three β -strands (B3, residues 38-42; B4, residues 46-50; and B5, residues 60-63) with B3 and B4 separated by a type I β -turn (T2; residues 43-45). Lastly, IPI* has several loops including those connecting helix 1 to β 1 (Loop 1; residues 11-18), β 1 to β 2 (Loop 2; residues 29-37) and a third loop (Loop 3; residues 51-59) that connects strands B4 and B5 of the second β -sheet (β 2). Representative NOE data (Figures 1 & 3) between strands B3 and B4 of IPI* are illustrated and are typical long-range connectivities observed in an antiparallel β -sheet. Likewise, hydrogen exchange rate data indicates that residues in the helical and β -sheet regions of IPI*, which are expected to be involved in hydrogen bonds, do indeed exchange more slowly than residues in loops or extended regions of the protein, further

validating the regions of secondary structure and β -strand alignment determined for IPI* by NMR (Figures 2 and 3).

3D structure calculations for IPI*

A total of 1244 distance constraints were used in the structure calculation of IPI* including 180 intraresidue, 389 sequential, 301 medium-range, 398 long-range, and 46 hydrogen bond constraints. In addition, 85 dihedral constraints (Φ and Ψ) and 55 dipolar coupling constraints were used for a total of 1384 constraints. NOE correlations were assigned on the basis of ^1H , ^{13}C , and ^{15}N backbone and side chain chemical shift assignments, and all NOE-based distance constraints used in the structure calculation were checked for their reciprocal correlation. Some degeneracy in chemical shift assignments made it difficult, in some cases, to assign NOE correlations using 3D NOE spectroscopy (NOESY) data alone. We therefore confirmed all of the NOE assignments using 4D ^{13}C , ^{15}N -edited NOESY-HSQC and 4D ^{13}C , ^{13}C -edited NOESY data sets to resolve ambiguities and provide unambiguous NOE assignments. Additional NOE correlations could then be confidently included in an iterative manner using preliminary structural models in which only unambiguous data was used initially. Likewise, we incorporated long-range $\text{H}^{\text{N}}\text{-N}$ residual dipolar coupling (RDC) constraints into the structure calculations throughout the entire protein to supplement the short-range NOE-derived constraints ($<6 \text{ \AA}$). Typical dipolar coupling data collected for IPI* is illustrated in Figure 1E, and RDC constraints were included for more than two-thirds of the residues, including a large number in loops and extended regions for which NOE constraints were minimally observed.

A family of the best 20 structures of IPI* is illustrated in Figure 4 with none of these structures having a dihedral violation greater than 5° or a distance violation greater than 0.40 \AA . The structures agree well with one another, with a backbone RMSD of 0.47 for regions of secondary structure and well-defined loops, and an RMSD of 0.56 for all heavy atoms. If all residues are included in the RMSD calculation, including residues in poorly defined loops (i.e. residues 12-17 and 33-38), then the backbone RMSD drops to 0.74 . On average, the family of structures is of high quality since 95.3% of residues in the 20 structures are in the most favorable and additionally allowed regions of the Ramachandran plot, with the best structure having 96.9% in the most favorable and additionally allowed regions. Additional structural parameters were assessed by the computer program PROCHECK, and the family of twenty structures is comparable with that of a 2 \AA X-ray crystal structure with on average about 10 bad contacts per 100 residues. Although hydrogen bonds were partially derived experimentally from amide exchange data, a family of structures that did not include hydrogen bond constraints was calculated to ensure the validity of incorporating these constraints. The pairwise RMSD between the average structures calculated with and without hydrogen bonds was 0.47 \AA , indicating these constraints do not perturb the structure. Lastly, the quality of the ensemble of structures was evaluated using the long-range dipolar coupling constraints (i.e. the Q-factor). In structures with dipolar couplings, the Q-factor was calculated to be 0.24 by omitting a randomly selected subset set of the J_{NH} data (10%). This Q-factor is a significant improvement from the $Q=0.35$ value, which was calculated when no dipolar couplings were included in the structure calculation and further validates the accuracy of the structure of IPI*.

IPI* represents a novel 3-dimensional fold

IPI* folds into a compact oblate ellipse that is roughly $30 \text{ \AA} \times 30 \text{ \AA} \times 15 \text{ \AA}$ with about half the residues in the protein involved in defined secondary structure (Figure 5). At the N-terminus, the hydrophobic face (L3, T4, V7, and I8) of the first amphipathic helix (H1) packs into Ala-52 and perhaps Phe-58 of Loop 3; however, much of loop 3 has fast timescale backbone dynamics (i.e. L57 to K59) that may compromise the overall stability of this small hydrophobic core (Fig. 6). Numerous long-range NOE constraints define a second, much larger hydrophobic core.

Specifically, this core comprises residues from the hydrophobic face of helix 2 (L68, I71, and L75), the second β -sheet (W38, A40, I47, I49, and I63), loop 2 (T31, V32, and L34) and the first and last residues of the first β -sheet (M19, I28), respectively (Figure 6). Helix 2 also displays an amphipathic character, with one side of the helix made up almost exclusively of isoleucine and leucine residues, while the other side contains mostly charged or polar residues. The hydrophobic side of helix 2 packs onto the face of the second β -sheet and against one end of the first β -sheet; furthermore, residues from the loop connecting the two β -sheets (loop 2) provide a cap for this relatively extensive hydrophobic core. Because it is generally recognized that hydrophobic interactions represent the major factor stabilizing folded structures of proteins ^{8, 9}, this relatively extensive hydrophobic core, for such a small protein, could provide an explanation for why IPI* is so stable in solution for such an extended period of time.

The surface of IPI* is relatively hydrophilic on one face of the oblate ellipse with the other face having a small hydrophobic patch (Figs. 5B, 5C). In fact, IPI* has a total of 13 positively charged residues with 11 of these 13 residues (K9, K12, R14, K17, K29, K59, K61, H62, H67, K69, K73) located on the hydrophilic basic side of IPI* (Fig. 5B); whereas, only two remaining positively charged residues (K41, K42) and a few negatively charged residues are on the other face of IPI* (Figure 5C). Along with the charged residues on the first face of IPI*, there exists an exposed hydrophobic protrusion (Fig 5B, circled in red) comprising of residues only from loop 3 (T55, A56, L57, and F58). That these exposed hydrophobic residues also are the only sequence of residues that show any real dynamic character makes them very likely candidates for being part of a protein-protein interaction involving IPI*. Lastly, a BLAST search on IPI* revealed that it shares very low sequence homology to any other protein. Likewise, when the structure of IPI* is compared to others in the protein data base using the online program DALI, the results indicate that the IPI* structure in fact comprises a unique protein fold. The novel IPI* compact fold does not support the view that all single-domain compact folds have now been identified ¹⁰; possibly other three dimensional search methods might find structurally more distantly related protein folds ¹¹.

IPI* inhibitor specificity

IPI* displays remarkable specificity as an inhibitor. Not only does IPI* function only as a specific inhibitor of *gmrS/gmrD* enzymes that target sugar modified and cytosine modified (HMC) T-even phage DNAs ^{1, 2}, but it is effective as an inhibitor against only some of these closely related enzymes (Table 3). IPI* inhibition allows growth of any sugar-HMC DNA containing phage on the *gmrS/gmrD* CT restriction endonuclease (from *E. coli* CT596) expressing clones, either when IPI* is injected by the infecting phage (column 3, see column 7 for the phage IP content) or when it is synthesized by an IPI* expression plasmid (column 2 IPI* versus column 1 mutant IPI*). Thus, IPI* is necessary and sufficient for protection against the CT restriction enzyme. However, clones expressing the ~90% sequence identical UT enzyme (from *E. coli* UT189) are immune to IPI* injected by the phage (column 4) or synthesized by the IPI* expression plasmid (column 5). Thus, the UT enzyme is immune to IPI* inhibition. Evidently some phages contain sugar HMC, or *ip* or other phage gene modifications that are sufficient to protect against the UT enzyme (e.g. T4, T4 *ip1*⁻, DDY1) and that are insufficient to protect against the CT enzyme (T4 *ip1*⁻). In fact, it appears more likely that these differences relate to CT and UT enzymatic differences in recognition and degradation of specific sugar HMC DNA modifications, since none of the *ip5-9* gene encoded proteins are effective inhibitors of the UT enzyme (rows 3-7).

There is apparently a significant gain of fitness associated with acquisition of *ip1* resistance by the UT enzyme, since it can protect against significantly more of the phages that can grow on a CT producing host (e.g. RB69, RB27 and D8 versus AC3 and DDY1). Although UT resistance found among phages such as AC3 and DDY1 is unrelated to the *ip1* inhibitor carried

by such phages and likely is due to unknown differences in the sugar HMC DNA modifications, greater knowledge of the CT and UT enzyme sugar HMC DNA directed nuclease specificities, and of the nature of the DNA sugar modifications among the phages is necessary to account fully for these resistance differences. In fact, it is not excluded that sequence and/or structural differences among the *ip1* genes themselves (e.g. 4 amino acid differences from T4 in IPI* of RB69), or the presence of other non *ip1* locus *ip* encoded inhibitor proteins among some of these phages may account for differences in resistance to the *gmrS/gmrD* restriction endonucleases. In any case, it appears that the UT enzyme has evolved from the CT enzyme specifically to respond to the phage IPI* inhibitor, a fourth generation response in this co-evolutionary pathway of increasingly elaborate modifications (Figure 7).

DISCUSSION

Absence of significant homology between IPI* and any other protein has previously ruled out structure-function predictions. Determination of its structure should facilitate future work elucidating the molecular mechanism and specificity of IPI* nuclease inhibition as well as of the *ip1* family of gene products. The IPI* restriction endonuclease inhibitor forms a unique and compact protein fold whose dimensions ($30 \times 30 \times 15$ Å) may allow the protein to be ejected through the 30 Å phage portal and tail apertures without unfolding. Although larger proteins that require refolding for biological function are ejected into the host by the T-even phages⁵, DNA protection may require immediate nuclease inhibition that would be impaired by a requirement for refolding. The most remarkable feature of the IPI* structure is the novel fold of the protein. The fold utilizes hydrophobic packing of one face of a beta sheet along the hydrophobic side of an amphipathic helix. Additional hydrophobic residues from a loop region are also incorporated into the core, creating a U-shaped domain that can apparently withstand quite harsh conditions; in the course of this work we observed that this protein survives multiple freeze-thaw cycles without visible breakdown products, remains folded at room temperature for upwards of 12 months, and stays folded from pH 3 to 10. The importance of the integrity of the structure to the function of the protein is demonstrated by the loss of inhibitor function of the single amino acid replacement *ip1*⁻ mutant KAI⁻ (A40T) protein². A closely related IPI* protein, containing the mutations A40T and E53K, was found to be completely unfolded in solution (data not shown), and as discussed in previous work, the E53K mutation is strongly suspected to be silent². The current work offers an explanation for why the A40T replacement disrupts the IPI* structure. A40, labeled in red in Figure 6, is located in the middle of a well-packed hydrophobic core and is important for retaining folded IPI*. The introduction of a slightly larger residue, coupled with the introduction of a polar group in the middle of an otherwise nonpolar environment (A40T), is apparently structurally disastrous to the beta-sheet core of the IPI* fold.

IPI* evidently inhibits the activity of the *E. coli* CT596 GmrS/GmrD restriction endonuclease through direct protein-protein interaction². Recent theories on the energetics dictating protein binding suggest that highly dynamic regions of otherwise well-ordered proteins often are involved in protein-protein interaction; thus binding interfaces are often flexible (reviewed in¹²). The heteronuclear NOE experiment (Figure 1B and 6B) identifies ¹H-¹⁵N pairs involved in fast time scale dynamics (picosecond regime) of the protein backbone. As shown in figure 6B, 9 residues (A1, T2, V23, D43, L57, E54, F58, K59, and L76) display rapid motion (noe/ref <0.68). The loop between B4 and B5 has a particularly large number of highly dynamic residues. Interestingly, this loop is also particularly solvent exposed, yet has the distinction of containing multiple hydrophobic residues (T55, A56, L57, F58). A combination of exposed hydrophobic residues and fast internal dynamics therefore lead us to hypothesize that IPI* binds *gmrS* or *gmrD* primarily through this loop region. Studies are ongoing to test this idea.

IPI* is observed to bind to DNA without apparent sequence specificity at low ionic strength (unpublished observations). It is likely that the highly electropositive side of IPI* binds electrostatically to the DNA and thus promotes co-ejection of IPI* with the DNA into the host. Such an association of the ~360 IPI* molecules over the length of the DNA entering the cell might promote nuclease inhibition, although DNA entry together with associated IPI* is evidently not required for inhibition, since IPI* protein synthesized within the host is also effective in protecting the entering DNA (Table 3). Whether the multiple other *ipI* locus homolog encoded internal proteins found among other T-even phage family members have different specificities and structures is an interesting question.

It is remarkable that the IPI* inhibitor has evolved apparently with the sole function of blocking a single specific CT heterodimeric type IV DNA modification dependent restriction endonuclease. It does not inhibit the 90% sequence identical monomeric type IV UT enzyme that targets comparable sugar DNA modifications and thus the UT enzyme apparently has evolved to block the IPI* inhibition. It will be of interest to determine whether or not IPI* binds to the UT enzyme, and, if so, whether the single polypeptide structure of this closely related enzyme is responsible for its IPI* immunity. The UT enzyme is a protein of ~70 kDa (unpublished observations) that may be more amenable to characterization than the heterodimeric CT enzyme, which has a tendency to unfold and bind strongly to the GroEL protein³. It appears on the basis of genetic and enzymatic criteria that the CT (and likely UT) enzymes are related to the McrBC enzyme (Figure 7) that attacks non-sugar modified hydroxymethylcytosine containing DNAs (discussed in³). McrBC is a unique Type IV modification dependent restriction endonuclease that translocates DNA by hydrolyzing GTP^{13, 14}. It appears that various NTPs provide for CT enzyme activity³, and likely the UT enzyme would have a comparable NTP dependency. NTP driven DNA translocation coupled to DNA hydrolysis is a fascinating feature of this type of restriction enzyme.

What is the connection between the diverse sugar modifications and the diverse *ipI* gene proteins of the T-even phages? One argument for a connection is the observation that the unmodified cytosine DNA containing PseudoT-even RB49 phage has no internal protein genes¹⁵. Except in the capsid targeting sequences the *ipI* locus genes display few sequence similarities. It is known that there are recombinogenic sequences that promote horizontal gene transfers among the *ipI* locus genes in the T-even phage family⁴. It is likely that only a small fraction of the gene diversity at the *ipI* locus in the T-even phage family has been sampled to date, thus likely there are a very large number of small proteins found associated with this single locus that could display novel protein folds. The *ipI* gene family relationship to the CT and UT family of *gmr* encoded enzymes that act only on glc-HMC DNA shows an interesting evolutionary progression of ever increasing complexity and specificity. What accounts for the extent of the gene diversity generated by this phage versus host arms race (Figure 7)? As in some other attack and defense wars (or cold wars) it appears that the phage and host have elaborated a wasteful set of increasingly elaborate armaments. Since the *gmr* enzymes do not attack cytosine phage DNAs (Table 3)^{2, 3}, why have the T-evens set out to add increasingly complex sugars, and sugars plus protein inhibitors to block enzymes directed against these modified DNAs when they could reverse course to non-modified DNA and focus on defenses against the enzymes that attack those DNAs? What maintains a forward direction along the evolutionary pathway shown, and what are the gains versus costs of this forward elaboration strategy? A detailed understanding of the components at play and their diversity is necessary to begin to answer this evolutionary question. What is the approximate magnitude of the unique and highly specific IPI*-like proteins that exist in this phage family, and where did they originate and undergo testing^{15, 16}? These are specific questions about the evolution of an enormous set of protein structures with a very circumscribed set of functions arising within a population estimated at over 10^{31} members¹⁷.

MATERIALS AND METHODS

Plasmids and phages

The IPI* precursor (95 amino acid residues) is assembled into the phage prohead scaffold where it is processed by the morphogenetic protease (gp21) that removes 19 N-terminal amino acids at a consensus ITE cleavage site, which leaves the mature IPI* (76 residues) within the capsid for injection with the DNA (5). Therefore the *ipI** and *ipI*⁻ genes cloned into pET15b for functional protein expression studies started from the cleavage site of the IPI* precursor of wild-type T4 and T4 *ipI*⁻ HA35 (A40T, E53K), with mutant residues numbered from the N-terminus of the IPI* polypeptide chain (2). Since the single amino acid replacement *ipI*⁻ mutant KAI⁻ (A40T) has the same null phenotype and lacks IPI* from the virion as does HA35 (2) it is likely that it is A40T that converts the IPI* chain to a random coil. The two *ip** gene containing clones maintained in *E. coli* BL21(DE3) for high level protein expression called DL13 and DL14, respectively, were constructed as described previously².

The *E. coli* CT596 *gmrS/gmrD* homolog from *E. coli* UT189 was cloned using primers 5'-CCCAAGCTTGATGAAGAGATAAGAAAGAATAACCAG (*Hind*III) and 5'-CGCGGATCCCTTTTGTTCAGGGAGATCTACCATGC (*Bam*HI) to insert the ~2.3 kb PCR product into pBELOBac11. Sequencing of the desired clone that restricted the growth of numerous T-even phages without restricting the growth of the cytosine DNA containing phages tested (see Table 3) confirmed the identity of the *gmrS/gmrD* homologs gene encoding a single 628 amino acid polypeptide chain comprising the *S* plus intervening sequence plus *D* residues of CT596. The sequence was identical to that reported from the genomic sequence of the uropathogenic *E. coli* UT189¹⁸. T-even phages, the cytosine DNA containing PseudoT-even RB49¹⁵, and other phages (λ , T3) tested (table 3) were either from this laboratory or were generous gifts of Henry Krisch (Toulouse), or Jim Karam (Tulane). Phage growth on plasmid containing bacteria was tested as described previously².

Protein expression and purification

DL13 and DL14 were used to over express and purify a His-tagged version of both the IPI* and mutant IPI* (HA35) proteins. Overnight cultures of these strains were diluted into 1 L of LB broth and grown exponentially at 37°C until at O.D. 600nm \approx 0.5 IPTG was added to cultures to a final concentration of 1 mM. Growth was continued at 30°C for 2.5 hours after which bacterial pellets were harvested by centrifugation at 4000rpm for 20 min at 4°C in a Sorvall GSA rotor. The remaining protein purification was conducted by following the protocol suggested in the His-Bind Kit (Novagen; San Diego, CA), and carried out at room temperature.

The IPI* and IPI*⁻ samples uniformly labeled with either ¹⁵N or ¹⁵N, ¹³C were prepared by growing strains DL13 and DL14 in minimal medium that contained either ¹⁵NH₄Cl (99%, ¹⁵N), or ¹⁵NH₄Cl and ¹³C-glucose (>98%, ¹³C) (Cambridge Isotope Laboratories, Inc; Woburn, MA) as the sole nitrogen and carbon sources, respectively. The expression and purification protocols were the same as described above. The eluted IPI* and IPI*⁻ were dialyzed against 20 mM sodium acetate buffer, pH 4.8 and then concentrated to reach a final concentration of 500 μ M. The samples were stored at -20°C for NMR experiments.

NMR Spectroscopy

All NMR samples contained ¹³C, ¹⁵N-labeled or ¹⁵N-labeled IPI* (0.25-0.75 mM), sodium acetate buffer (20 mM), pH 4.8, and 10% D₂O. Samples were passed through a Sephadex G25 column to remove trace impurities, re-concentrated in *d*₃ sodium acetate (>99 atom% deuterium; Cambridge Isotope Laboratories), degassed for at least one hour, flushed with argon gas and sealed. Samples prepared in this manner were remarkably stable (i.e. for at least a year at room temperature) as judged by the lack of change in all of its ¹H-¹⁵N correlations in 2D

HSQC experiments. Sequential backbone and side-chain assignments of IPI* were obtained using standard heteronuclear multidimensional NMR spectroscopy as described previously¹⁹. Specifically, heteronuclear NMR spectra were collected at 37°C on either a Bruker DMX600 NMR spectrometer (600.13 MHz for protons) or a Bruker AVANCE 800 NMR spectrometer (800.27 MHz for protons) equipped with pulse-field gradients, four frequency channels and a triple-resonance z-axis gradient 5mm cryogenic probehead. Unless otherwise stated, multidimensional NMR data were collected in the indirect dimensions using States-TPPI phase-cycling using a 1 s relaxation delay²⁰. For most experiments, initial delays in the indirect dimensions were set to give zero- and first-order phase corrections of 90° and -180°, respectively²¹. In general, pulse field gradients were included to purge undesired transverse magnetization, and the WATERGATE technique was used in most cases to suppress the solvent signal in samples dissolved in H₂O. Data were processed on Linux Fedora 2.6 using the software package NMRPIPE²². All proton chemical shifts are reported with respect to the H₂O or HDO signal, which is taken as 4.658 ppm downfield from external TSP (0.00 ppm) at 37 °C. The ¹³C and ¹⁵N chemical shifts were indirectly referenced using the following ratios of the zero-point frequencies at 37 °C: 0.10132905 for ¹⁵N-¹H and 0.25144953 for ¹³C-¹H^{21, 23, 24}.

Uniformly ¹⁵N-labeled IPI* was used to collect the 2D ¹H, ¹⁵N fast HSQC²⁵, 3D ¹⁵N, ¹⁵N-edited HMQC-NOESY-HSQC²⁶, and 3D ¹⁵N-edited NOESY-HSQC spectra. The ¹³C, ¹⁵N-labeled IPI* was used to collect 3D CBCA(CO)NH²⁷, 3D HNCACB²⁸, 3D HNCO²⁹, 3D C(CO)NH²⁷, 3D H(CCO)NH, 4D ¹³C, ¹⁵N-edited NOESY-HSQC³⁰, and 4D ¹³C, ¹³C-edited NOESY HSQC spectra (Table 1). Hydrogen bonds were determined through a hydrogen-deuterium exchange experiment, where IPI* in H₂O was lyophilized, redissolved in D₂O, and a series of fast HSQC spectra were collected every 10 min for 2 hours, then every 30 min. for 10 hours. A 2D IPAP ¹H-¹⁵N HSQC was collected on isotopic and partially aligned ¹⁵N-labeled IPI* samples to obtain H-N dipolar coupling data, as previously described³¹. These data were separated into complementary in-phase and anti-phase matrices, each with 512* × 2048* points in *t*₁ and *t*₂, and acquisition times of 80.9 ms (¹⁵N) and 122.1 ms (¹H). The processed matrices contained 1024 × 4096 real points F1 and F2. In the final processed data, the peak positions were determined using the contour averaging method with the data analysis program PIPP³². The 2D ¹⁵N-¹H heteronuclear NOE experiment was run on a ¹⁵N-labeled sample, and ¹⁵N-¹H NOE ratios were measured by fully interleaving the NOE and reference experiments as previously described³³. The heteronuclear NOE experiment used a 3-s presaturation period and a 2-s relaxation delay; the control experiment had an equivalent 5-s delay. In these experiments, water was returned to the z-axis prior to each scan to avoid saturation transfer. Both NOE and control experiments were acquired with 1024* (*t*₂) × 246* (*t*₁) points, with 128 scans. Spectra from the heteronuclear NOE experiment were extended by linear prediction and zero-filling prior to apodization with a 5%-shifted mixed Gaussian/exponential function. NOE peak heights were taken from Gaussian fits of the frequency-domain data. The NOE was calculated as $NOE = \eta + 1 = I_{\text{presat}}/I_{\text{no-presat}}$, where *I*_{presat} and *I*_{no-presat} are the peak heights with and without ¹H presaturation, respectively. The heteronuclear NOE experiment was used to identify residues of the protein (NOE < 0.68) that are dynamic on a fast timescale; such residues in loops (loops 2-5), and at the N- and C-termini, including residues in the His-tag pre-sequence; these residues were not included in structure calculations.

Structure calculations

Interproton distance restraints derived from NOESY experiments were classified into five levels including strong, medium, medium weak, weak and very weak representing distance ranges of a lower limit of 1.8Å and an upper limit of 6Å, as described previously^{34, 35}. Pseudoatom corrections were applied to the upper limit for degenerate methyl, methylene and

aromatic ring protons, and restraints for methyl protons were shifted down one category. Only structurally useful constraints were used, and thus NOE correlations between geminal protons and vicinal methylene protons were excluded. Dihedral constraints $\Phi \pm 20^\circ$ $\psi \pm 15^\circ$ for α -helix and $\Phi \pm 40^\circ$ $\psi \pm 40^\circ$ for β -sheet were included on the basis of hydrogen exchange rate, predictions from the program TALOS³⁶, and chemical shift index of $^1\text{H}^\alpha$ and $^{13}\text{C}^\alpha$ ³⁷. In regions of known secondary structure, the values fell within the normal bond angles for α helices ($\varphi = -60 \pm 20^\circ$, $\psi = -40 \pm 15^\circ$) and β strands ($\varphi = -120 \pm 30^\circ$, $\psi = -140 \pm 30^\circ$). Omission of dihedral constraints produced no significant structural changes to the structure.

A total of 55 internuclear dipolar coupling constraints were determined by recording the difference in the $^1J_{\text{HN}}$ values (in hertz) between standard (isotropic) and axially stretched polyacrylamide (aligned) samples. The axial (\mathbf{A}_a) and rhombic (\mathbf{A}_r) components of the alignment tensor \mathbf{A} were obtained from a histogram of the dipolar coupling that were normalized to $^1D_{\text{NH}}$ values as previously described³⁸. The force constant for the dipolar coupling energy was slowly increased in concert with the force constants of the NOE energy during the cooling stage of the structure calculations. The initial force constants were originally small ($0.001 \text{ kcal Hz}^{-2}$), and the final force constants were determined for the $^1D_{\text{NH}}$ dipolar couplings such that the calculated structures represented the estimated uncertainty in the actual dipolar coupling data set³⁹.

Hydrogen bond constraints of $r_{\text{NH-O}} = 1.5\text{-}2.8 \text{ \AA}$ and $r_{\text{N-O}} = 2.4\text{-}3.5 \text{ \AA}$ were included in the final stage of structure calculations. Also, pseudopotentials for secondary $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts and a conformational database potential were included in the final SA refinement⁴⁰. Two hundred structures were calculated using the computer program XPLOR version 2.9.3 using a $(\sum r^{-6})^{-1/6}$ sum for the distance dependent constraints⁴¹. Standard protocols for sub-structure embedding, regularization, hybrid distance geometry-simulated annealing (DGSA) regularization/refinement, and simulated annealing (SA) were used in these calculations. Overall, this refinement process yielded >60 structures with no distance constraint violations greater than 0.4 \AA and no dihedral violations $>5^\circ$. From this group, the 20 best structures were chosen based on their overall energy level (Table 2). Coordinates and chemical shift values of IPI* have been deposited in the Protein Data Base under accession number **2JUB**.

ACKNOWLEDGEMENTS

DR and LWB were supported by a grant from the NIH (AI11676). DJW was supported by a grant from the NIH (GM58888), and NTW was supported by an American Heart Association predoctoral training grant (0615343U). We thank Swain Chen and Scott Hultgren (Washington University, St. Louis) for providing uropathogenic (UPEC) *E. coli* UT189.

REFERENCES

1. Abremski K, Black LW. The function of bacteriophage T4 internal protein I in a restrictive strain of *Escherichia coli*. *Virology* 1979;97(2):439–47. [PubMed: 382610]
2. Bair CL, Rifat D, Black LW. Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. *J Mol Biol* 2007;366(3):779–89. [PubMed: 17188711]
3. Bair CL, Black LW. A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* 2007;366(3):768–78. [PubMed: 17188297]
4. Repoila F, Tetart F, Bouet JY, Krisch HM. Genomic polymorphism in the T-even bacteriophages. *The EMBO journal* 1994;13(17):4181–92. [PubMed: 8076614]
5. Mullaney JM, Black LW. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J Mol Biol* 1996;261(3):372–85. [PubMed: 8780780]
6. Carlson, K.; Raleigh, EA.; Hattman, S. Restriction and Modification. In: Karam, JD., editor. In the *Molecular Biology of bacteriophage T4*. ASM Press; Washington, DC: 1994. 1994. p. 369-381.

7. Karam, JD.; Drake, JW. Molecular biology of bacteriophage T4. American Society for Microbiology; Washington, DC: 1994.
8. Dill KA. Dominant forces in protein folding. *Biochemistry* 1990;29(31):7133–55. [PubMed: 2207096]
9. Kim PS, Baldwin RL. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 1990;59:631–60. [PubMed: 2197986]
10. Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. On the origin and highly likely completeness of single-domain protein structures. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(8):2605–10. [PubMed: 16478803]
11. Hou J, Jun SR, Zhang C, Kim SH. Global mapping of the protein structure space and application in structure-based inference of protein function. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(10):3651–6. [PubMed: 15705717]
12. Ishima R, Torchia DA. Protein dynamics from NMR. *Nat Struct Biol* 2000;7(9):740–3. [PubMed: 10966641]
13. Panne D, Raleigh EA, Bickle TA. The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J Mol Biol* 1999;290(1):49–60. [PubMed: 10388557]
14. Bourniquel AA, Bickle TA. Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 2002;84(11):1047–59. [PubMed: 12595133]
15. Comeau AM, Bertrand C, Letarov A, Tetart F, Krisch HM. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* 2007;362(2):384–96. [PubMed: 17289101]
16. Comeau AM, Krisch HM. War is peace--dispatches from the bacterial and phage killing fields. *Current opinion in microbiology* 2005;8(4):488–94. [PubMed: 15979391]
17. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96(5):2192–7. [PubMed: 10051617]
18. Chen SL, Hung CS, Xu J, et al. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(15):5977–82. [PubMed: 16585510]
19. Baldisseri DM, Rustandi RR, Zhang Z, et al. 1H, 13C and 15N NMR sequence-specific resonance assignments for rat apo-S100A1(alpha alpha). *J Biomol NMR* 1999;14(1):91–2. [PubMed: 10382312]
20. Marion D, Driscoll PC, Kay LE, et al. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* 1989;28(15):6150–6. [PubMed: 2675964]
21. Spera S, Ikura M, Bax A. Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment. *J Biomol NMR* 1991;1(2):155–65. [PubMed: 1668721]
22. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6(3):277–93. [PubMed: 8520220]
23. Edison AS, Abildgaard F, Westler WM, Mooberry ES, Markley JL. Practical introduction to theory and implementation of multinuclear, multidimensional nuclear magnetic resonance experiments. *Methods Enzymol* 1994;239:3–79. [PubMed: 7830587]
24. Live DH, Davis DG, Agosta WC, Cowburn D. Long range hydrogen bond mediated effects in peptides: ¹⁵N NMR study of gramicidin S in water and organic solvents. *J Am Chem Soc* 1984;106:1939–41.
25. Mori S, Abeygunawardana C, Johnson MO, van Zijl PC. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J Magn Reson B* 1995;108(1):94–8. [PubMed: 7627436]
26. Ikura M, Bax A, Clore GM, Gronenborn AM. Detection of Nuclear Overhauser Effects between Degenerate Amide Proton Resonances by Heteronuclear Three-dimensional Nuclear Magnetic Resonance Spectroscopy. *Journal of American Chemical Society* 1990;112:9020–2.

27. Grzesiek S, Anglister J, Bax A. Correlation of backbone amide and aliphatic side-chain resonances in $^{13}\text{C}/^{15}\text{N}$ -enriched proteins by isotropic mixing of ^{13}C magnetization. *Journal of Magnetic Resonance* 1993;B101:114–9.
28. Wittekind M, Mueller L. HNCACB a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta proton resonances in proteins. *J Magn Reson* 1993;B101:205–10.
29. Kay LE, Clore GM, Bax A, Gronenborn AM. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. *Science* 1990;249(4967):411–4. [PubMed: 2377896]
30. Muhandiram DR, Guang YX, Kay LE. An enhanced-sensitivity pure absorption gradient 4D ^{15}N , ^{13}C -edited NOESY experiment. *J Biomol NMR* 1993;3:463–70.
31. Wright NT, Varney KM, Ellis KC, Markowitz J, Gitti RK, Zimmer DB, Weber DJ. The three-dimensional solution structure of $\text{Ca}(2+)$ -bound S100A1 as determined by NMR spectroscopy. *J Mol Biol* 2005;353(2):410–26. [PubMed: 16169012]
32. Garrett DS, Powers R, Gronenborn AM, Clore GM. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* 1991;95:214–20.
33. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* 1994;4(5):727–34. [PubMed: 7919956]
34. Clore GM, Nilges M, Sukumaran DK, Brunger AT, Karplus M, Gronenborn AM. The Three-dimensional structure of a 1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *EMBO J* 1986;5:2729–35. [PubMed: 16453716]
35. Drohat AC, Baldisseri DM, Rustandi RR, Weber DJ. Solution structure of calcium-bound rat S100B (betabeta) as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* 1998;37(9): 2729–40. [PubMed: 9485423]
36. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 1999;13(3):289–302. [PubMed: 10212987]
37. Wishart DS, Sykes BD. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR* 1994;4(2):171–80. [PubMed: 8019132]
38. Clore GM, Gronenborn AM, Tjandra N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* 1998;131(1):159–62. [PubMed: 9533920]
39. Drohat AC, Tjandra N, Baldisseri DM, Weber DJ. The use of dipolar couplings for determining the solution structure of rat apo-S100B(betabeta). *Protein Sci* 1999;8(4):800–9. [PubMed: 10211826]
40. Kuszewski J, Gronenborn AM, Clore GM. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J Magn Reson* 1997;125(1):171–7. [PubMed: 9245376]
41. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 2003;160(1):65–73. [PubMed: 12565051]

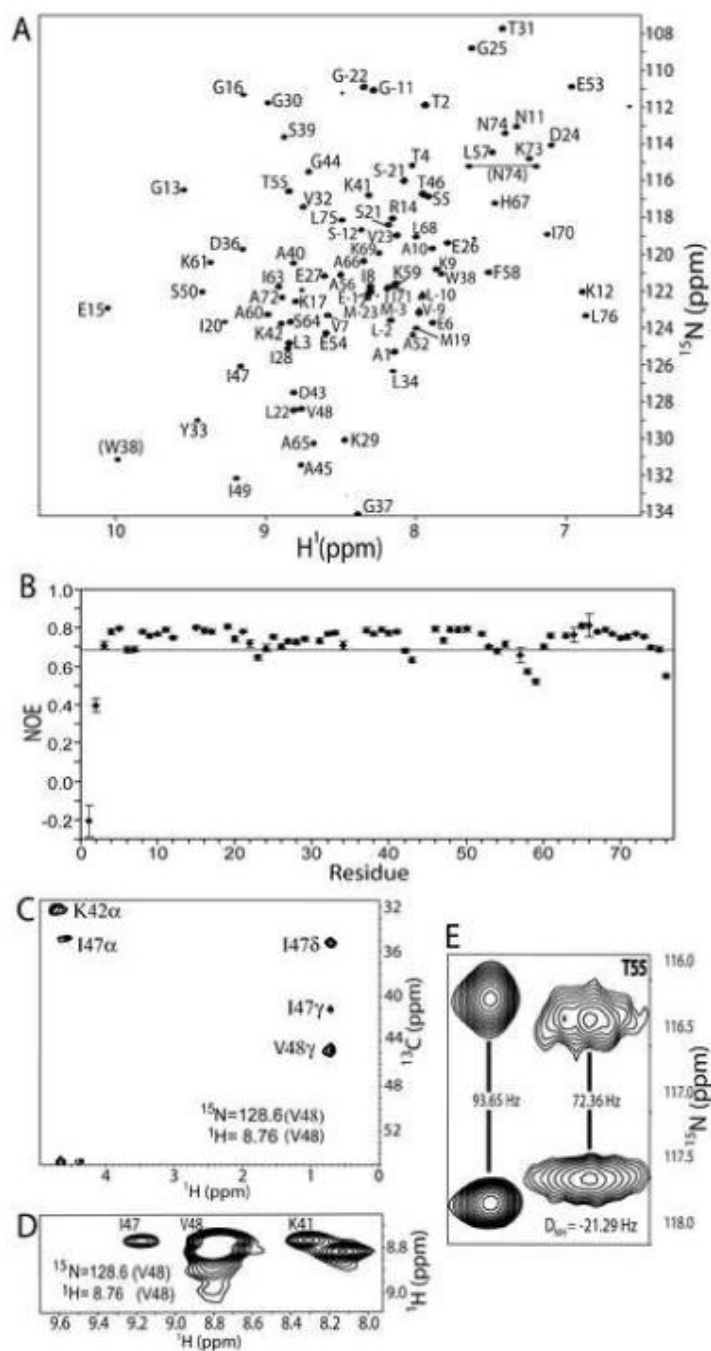


Figure 1.

(A) 2D ^1H - ^{15}N HSQC of IPI* at 600 MHz (^1H) with residue assignments included. Backbone ^1H - ^{15}N correlations are labeled sequence-specifically, and correlations connected by the horizontal line correspond to asparagine side-chain NH_2 groups. (B) A graph representing data obtained from a 2D ^{15}N - $\{^1\text{H}\}$ heteronuclear NOESY experiment shows that A1, T2, V23, D43, L57, F58, K59, and L76 all have significant fast-motion contributions ($\text{NOE} < 0.68$; residues below the horizontal line in graph). A plane from the 4D ^{13}C , ^{15}N -edited NOESY-HSQC (C) and a 3D ^{15}N -edited NOESY (D) show NOE correlations of the amide hydrogen of V48 to sequential and long-range resonances. In particular, NOE correlations of $\text{H}^\alpha\text{H}^\text{N}$ ($i, i+1$), $\text{H}^\alpha\text{H}^\text{N}$ (K42, V48), $\text{H}^\text{N}\text{H}^\text{N}$ (K41, V48) are visible. The presence of a strong

$H^{\alpha}H^N(i, i+1)$ resonance and strong long-range resonances is indicative of a residue in the middle of a beta sheet. (E) A representative region of the 2D 1H - ^{15}N HSQC IPAP NMR spectra used to measure dipolar coupling data (RDCs). This region illustrates splitting of T55. Dipolar couplings were determined by comparing the splitting in the presence (right) or absence (left) of the aligning acrylamide media.

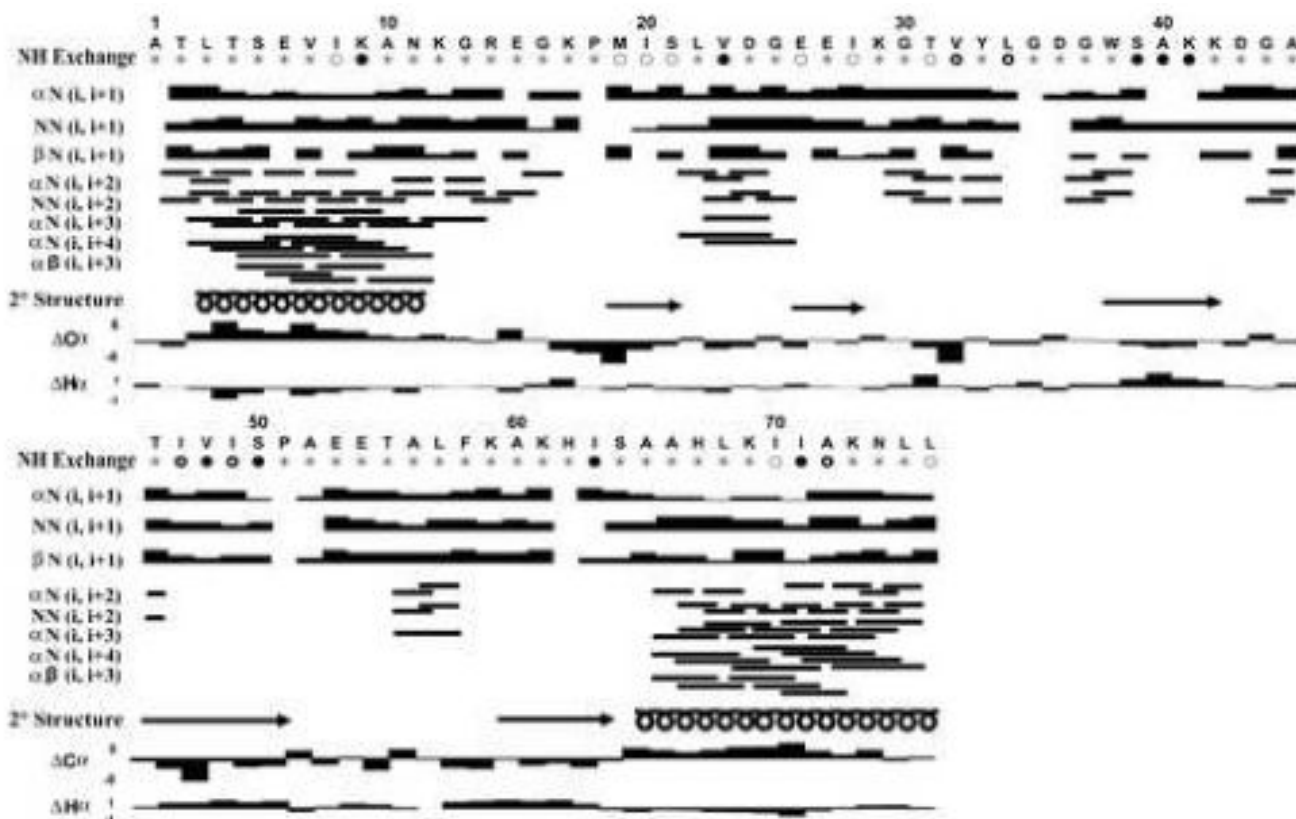


Figure 2.

Secondary structure of IPI. Circles indicate relative amide hydrogen exchange rates at 37°C as determined using 2D ^1H - ^{15}N HSQC spectra with stars ($T < 0.125\text{h}$) and open circles ($0.125 < T < 0.5\text{h}$) arbitrarily referred to as fast exchanging amide protons. Residues with partially filled in circles are arbitrarily referred to as medium exchanging amide protons ($0.5 < T < 1\text{h}$). Residues with solid circles ($T > 4\text{h}$) are arbitrarily referred to as slowly exchanging amide protons. The NOE correlations were determined from 3D ^{15}N -edited NOESY-HSQC, 4D ^{13}C , ^{15}N -edited NOESY, and 4D ^{13}C , ^{13}C -edited NOESY at 37°C. The height of the bar indicated the strength of the NOE (strong, medium, medium-weak, weak, or very weak). Deviations in the $^{13}\text{C}^\alpha$ and $^1\text{H}^\alpha$ chemical shift from those of a random coil are illustrated such that regions of contiguous upfield-shifted $^{13}\text{C}^\alpha$ chemical shifts (positive values) and downfield-shifted $^1\text{H}^\alpha$ chemical shifts (negative values) are indicative of helical regions. Likewise, regions of contiguous downfield-shifted $^{13}\text{C}^\alpha$ chemical shifts (negative values) and upfield-shifted $^1\text{H}^\alpha$ chemical shifts (positive values) are indicative of regions of β -strand. The secondary structure is represented by α -helices (spirals), β -strands (arrows), turns (crosses), loops and flexible regions (no symbols), as indicated under the appropriate residues in the sequence of IPI*.

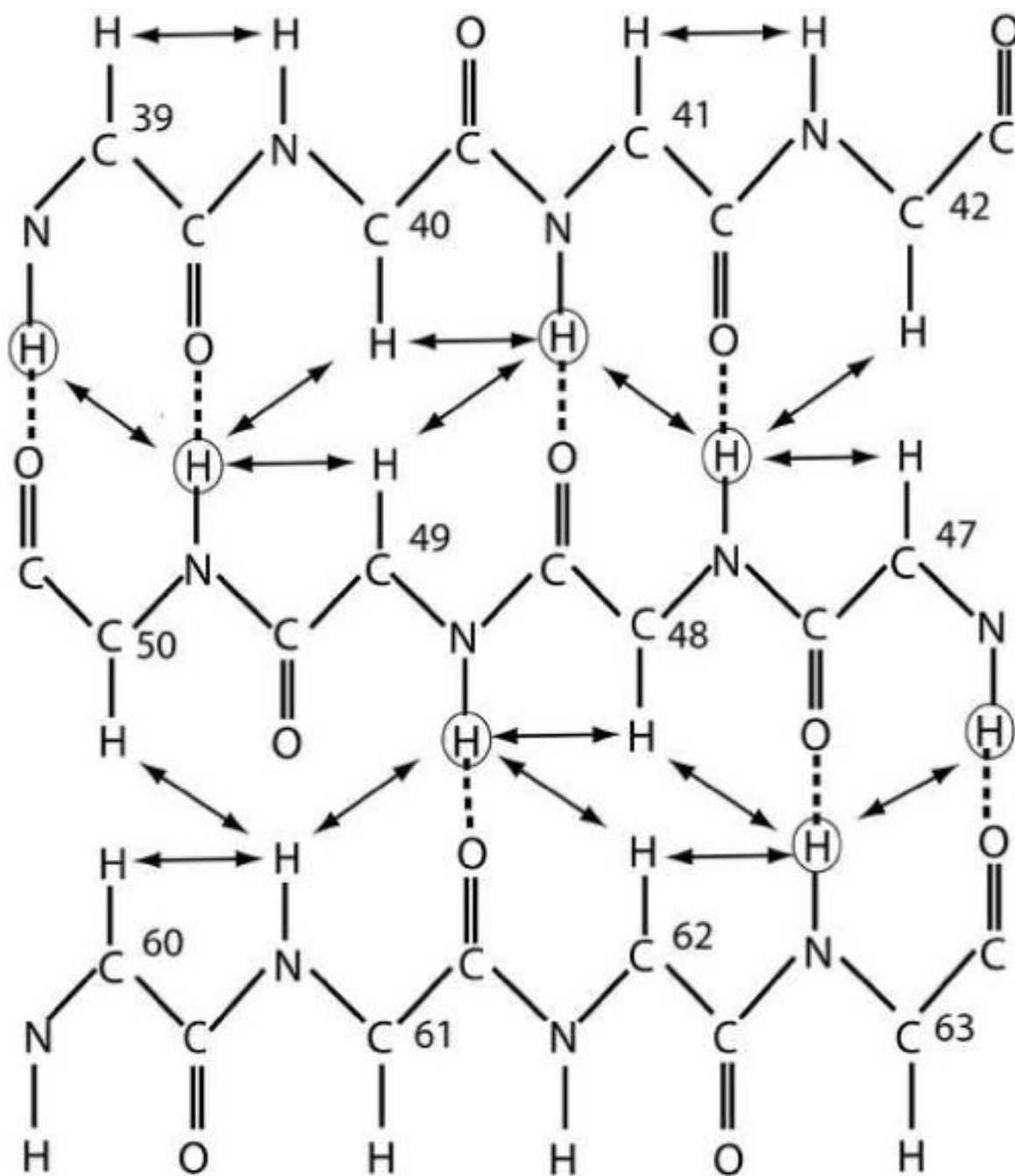


Figure 3. Schematic of the 3-strand antiparallel beta sheet. Unambiguous NOEs are marked by solid arrows. Circled hydrogen atoms are those involved in slow exchange with the solution, and dotted lines represent hydrogen bonds.

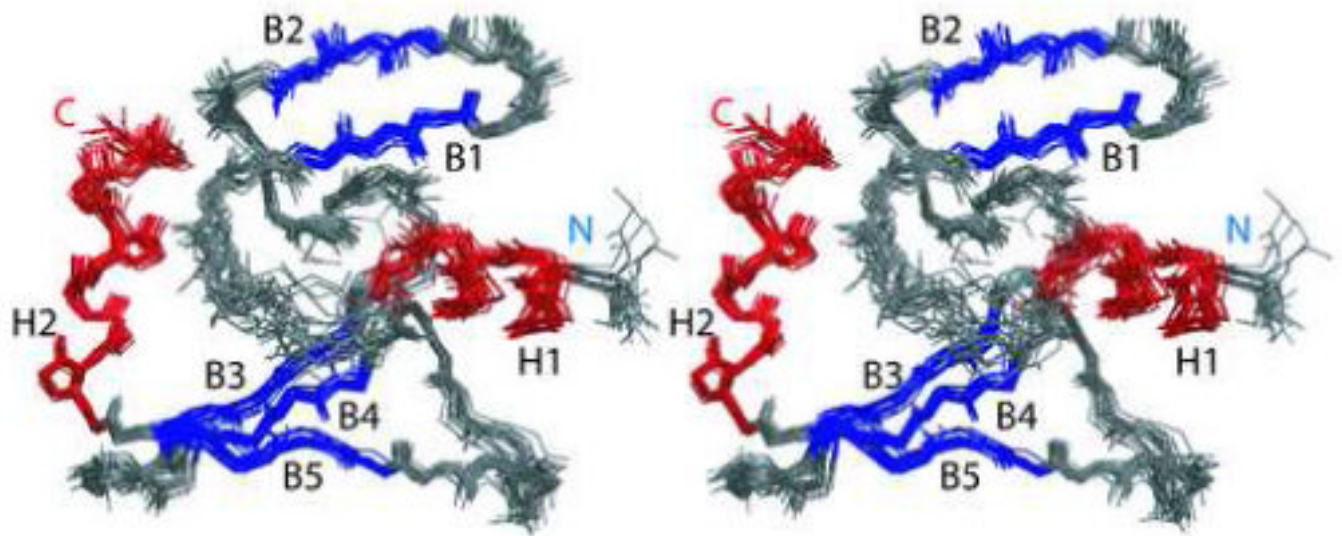


Figure 4. Stereoview of the 20 lowest energy structures of IPI*. Helices are colored red, beta strands are colored blue, turns and loops and flexible regions are colored gray.

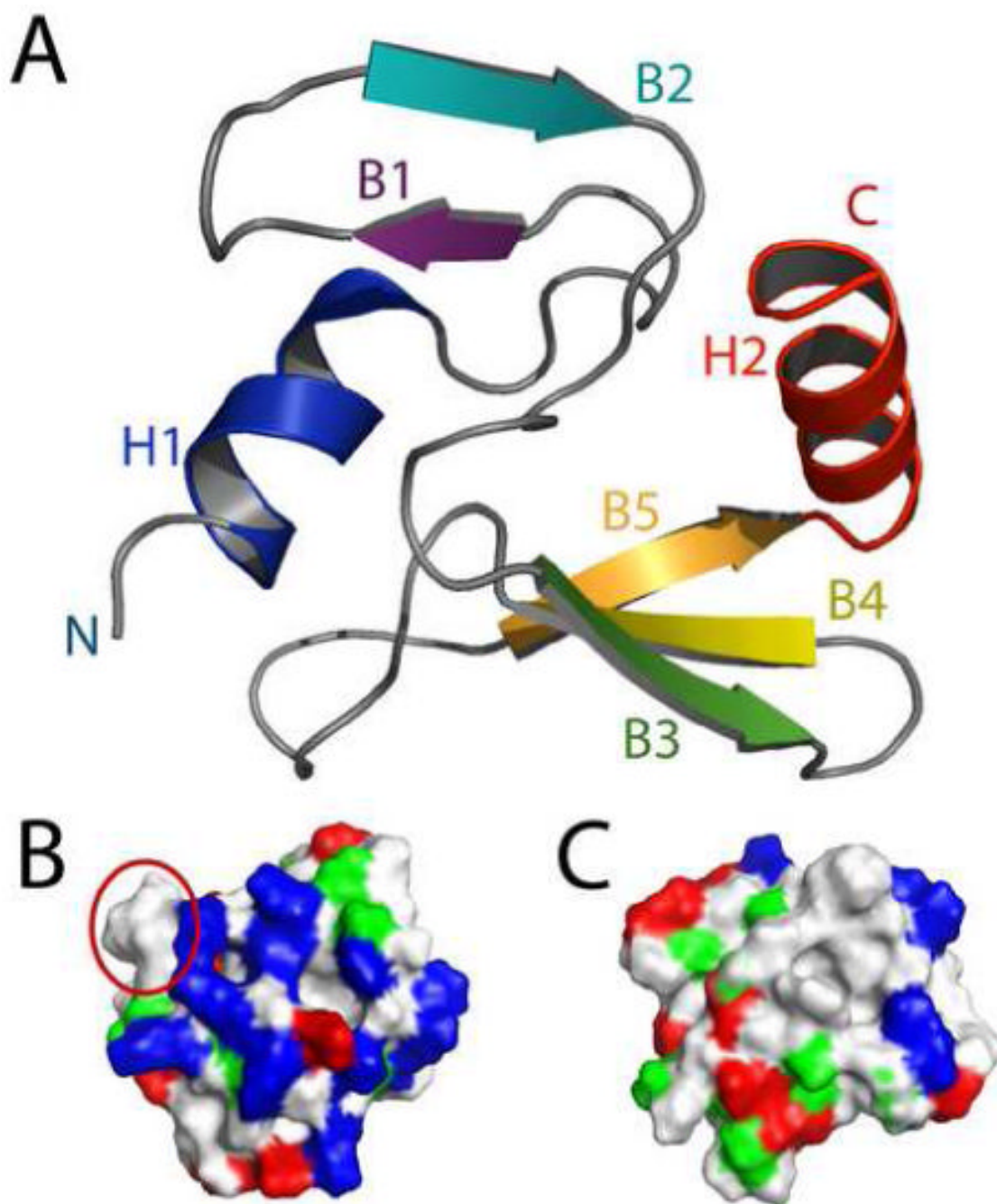


Figure 5. NMR solution structure of IPI*. (A) Ribbon diagram of the lowest-energy structure of IPI*. (B) and (C) The electrostatic surface potential of IPI. Electronegative (acidic) regions are colored *red*, electropositive (basic) regions are colored *blue*, uncharged regions are colored *white* and polar group residues are colored *green*. In (B), exposed hydrophobic patch residues 55-58 are circled in red.

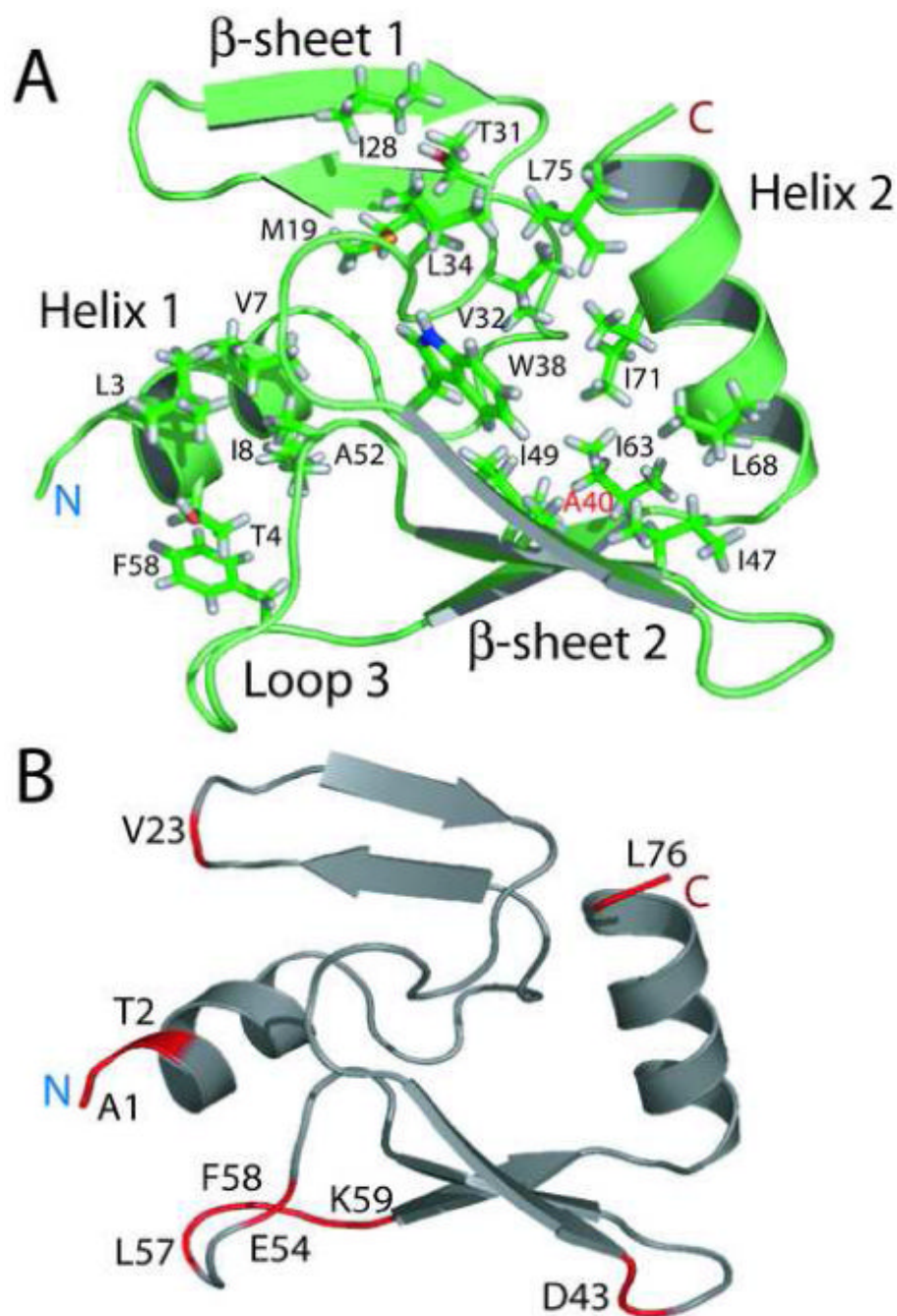


Figure 6. Hydrophobic residues involved in the unique fold of IPI*. (A) IPI* maintains its fold through interactions within a large hydrophobic fold. Residues included in this hydrophobic pocket are L3, T4, V7, and I8 on helix 1, M19 and I28 on the first β -sheet, T31, V32, and L34 in loop 2, W38, A40, I47, I49, and I63 in the second β -sheet, A52 and F58 in the loop 3, and L68, I71, and L75 in helix 2. Labeled in red is residue A40, where an A40T substitution eliminates IPI* structure and inhibitor function. (B) Ribbon diagram illustrating residues (in red) that exhibit fast time-scale motions. Several of the residues with significant fast-motion contributions are located in loop 3.

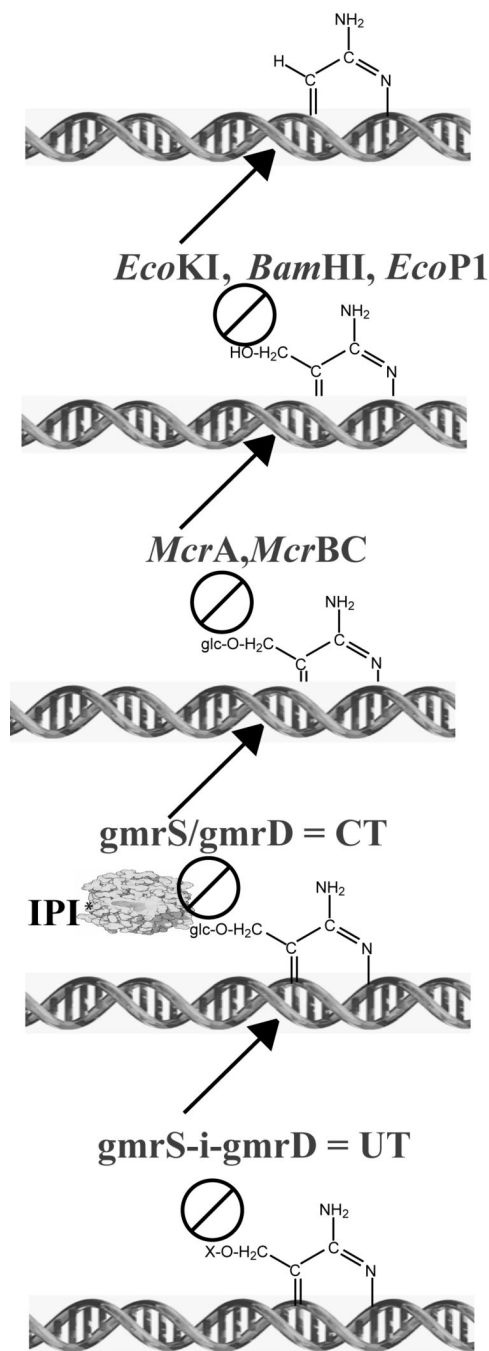


Figure 7.

Evolution of the T-even phage IPI* inhibitors and DNA modifications in response to type IV DNA modification dependent restriction endonucleases. Bacterial encoded type I, II, and III (e.g. *E. coli* K12, *BamHI*, P1) restriction endonucleases protect against infecting phage DNAs containing cytosine (first line). Many of these enzymes are blocked by 5-methyl or 5-hydroxymethylcytosine (HMC) DNA modification (line 2). The *McrA* and *McrBC* type IV modification dependent restriction endonucleases of *E. coli* specifically attack HMC modified DNAs, but are inhibited by the glucosylation of HMC (glc-HMC) (line 3). The heterodimeric CT (*gmr* (glucose modified restriction) enzyme of *E. coli* CT596 hydrolyzes only the sugar modified (glc)-HMC containing DNAs of a number of T-even phages, but its activity is

inhibited by the encapsidated phage IPI* protein injected with the DNA (line 4). The UT enzyme comprising the S and D subunits in a single polypeptide chain is immune to IPI*, but is blocked by further unknown modifications to the DNAs or internal proteins of some of the T-even phages (line 5).

Table 1

Parameters for NMR Experiments ¹

experiment	Dim	Nuc	Parameters ²		Acq time (ms)
			time pts	freq pts	
2D ¹ H, ¹⁵ N fast HSQC	t ₁	¹⁵ N	256	512	20.5
	t ₂	¹ H	1024	2048	79.9
3D ¹⁵ N -edited HNCA	t ₁	¹⁵ N	60	128	20.5
	t ₂	¹³ C	128	256	12.4
	t ₃	¹ H	1024	2048	46.0
3D ¹⁵ N -edited NOESY-HSQC ³	t ₁	¹⁵ N	180	512	4.5
	t ₂	¹⁵ N	80	160	19.3
	t ₃	¹ H	1024	2048	46.0
3D ¹⁵ N, ¹⁵ N -edited HMQC-NOESY-HSQC	t ₁	¹⁵ N	42	128	19.2
	t ₂	¹⁵ N	42	128	19.3
	t ₃	¹ H	1024	2048	46.0
	t ₄	¹³ C ⁴	128	256	3.3
3D CBCA(CO)NH	t ₁	¹⁵ N	40	128	8.6
	t ₂	¹ H	1024	2048	91.8
	t ₃	¹³ C	126	256	5.0
3D HNCACB	t ₁	¹⁵ N	60	128	20.5
	t ₂	¹⁵ N	1024	2048	32.0
	t ₃	¹ H	128	256	11.3
3D HNCO	t ₁	¹⁵ N	40	128	8.6
	t ₂	¹ H	1024	2048	91.8
	t ₃	¹³ C ⁵	110	512	5.5
3D H(CCO)NH	t ₁	¹⁵ N	44	128	22.8
	t ₂	¹ H	1024	2048	106.5
	t ₃	¹³ C ⁴	64	128	25.7
3D C(CO)NH	t ₁	¹⁵ N	36	128	13.8
	t ₂	¹ H	128	256	6.4
	t ₃	¹³ C ⁷	34	64	5.7
4D ¹³ C, ¹⁵ N-edited ⁶ NOESY-HSQC	t ₁	¹⁵ N	512	1024	61.1
	t ₂	¹³ C ⁴	36	64	13.8
	t ₃	¹ H	1024	2048	13.8
	t ₄	¹³ C ⁴	36	64	7.3
4D ¹³ C, ¹³ C-edited ⁶ NOESY-HSQC	t ₁	¹³ C ⁴	128	256	7.3
	t ₂	¹³ C ⁴	128	256	71.3
	t ₃	¹ H	1024	2048	
	t ₄	¹ H	1024	2048	

¹Data were collected in H₂O at 37 °C at 800.27 MHz for ¹H except for the H(CCO)NH, C(CO)NH, 4D NOESY experiments, and dynamic experiments, which were collected at 600.13 MHz for ¹H.

²The number of points in the time domain is complex. The number of points in the frequency domain is real. The carrier frequency is 4.658 and 118.0 ppm for ¹H and ¹⁵N, respectively, unless otherwise stated.

³The NOE mixing time was 150 ms.

⁴The ¹³C carrier position was set at 43.78 ppm.

⁵The ¹³C carrier position was set at 173.78 ppm.

⁶The NOE mixing time was 130 ms.

⁷The ¹³C carrier position was set at 40.78 ppm.

Table 2
NMR-derived restraints and statistics of 20 NMR structures ⁸

	<20>	Best
rmsd from distance constraints (Å) ⁹		
total (1244)	0.027 ± .002	0.025
intraresidue (180)	0.004 ± .001	0.004
sequential (i - j = 1) (389)	0.017 ± .003	0.018
medium range (1 < i - j ≤ 1) (301)	0.030 ± .004	0.024
long range (i - j > 5) (328)	0.033 ± .004	0.032
hydrogen bonds (46)	0.060 ± .009	0.050
rmsd from dihedral constraints (°)		
Φ, Ψ (85)	0.269 ± .172	0.119
rmsd from dipolar coupling restraints (Hz)		
<i>D</i> _{NH} (46)	0.49 ± .05	0.43
rmsd from exptl ¹³C chemical shifts		
¹³ Cα (ppm) (76)	1.31 ± .09	1.27
¹³ Cβ (ppm) (69)	1.31 ± .10	1.24
rmsd from idealized geometry		
bonds (Å)	0.004 ± .001	0.004
angles (°)	0.609 ± .035	0.590
impropers (°)	0.601 ± .027	0.580
Lennard-Jones potential energy (kcal/mol) ¹⁰	-235 ± 12	-238
Region of the Ramachandran plot ¹¹		
Most favorable	73.2 ± 4.1	72.3
Additionally allowed	22.1 ± 4.3	24.6
Generously allowed	4.6 ± 2.0	3.1
Dissallowed	0.0 ± 0.0	0
Rmsd to the mean structure (Å) ¹²		
Backbone	0.743 ± .108	0.590
all heavy atoms	1.303 ± .182	1.05
ordered backbone	0.484 ± .119	0.406
ordered heavy atoms	.562 ± .116	0.492

⁸The 20 ensemble structures, <20>, are the results of simulated annealing calculations. The best structure is chosen from the lowest overall RMSD value. The values shown for the <20> are the mean ± standard deviation.

⁹None of the 20 structures has a distance violation > 0.4 Å or a dihedral angle violation of > 5°. The force constants used in the SA calculations are as follows: 1000 kcal mol⁻¹ Å² for bond length, 500 kcal mol⁻¹ rad⁻² for angles and improper torsions, 4 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals (vdw) repulsion term (hard-sphere effective vdw set to 0.8 times their values in CHARMM parameters), 50 kcal mole⁻¹ Å⁻² for experimental distance constraints, 100 kcal mol⁻¹ Å⁻² for non-crystallographic symmetry, 1kcal mol⁻¹ Å⁻² for distance symmetry constraints, 0.5 kcal mol⁻¹ ppm⁻² for the ¹³C chemical shift constraints, and 1.0 for the conformational database potential. The force constant (in kcal Hz⁻²) used for dipolar coupling restraints was 0.70 for ¹⁵N-¹H^N.

¹⁰Lennard-Jones van der Waals energies were calculated using CHARMM parameters and were not used in any stage of the structure determination

¹¹PROCHECK was utilized to generate the Ramachandran plot.

¹²Backbone calculations include C^α, N, and C^γ atoms. Ordered backbone residues consist of 3-10, 18-28, 39-49, and 59-76 as these are the residues are in secondary structural regions or well-defined loops (65% of the molecule). ¹²The 20 ensemble structures, <20>, are the results of simulated annealing calculations. The best structure is chosen from the lowest overall RMSD value. The values shown for the <20> are the mean ± standard deviation.

Table 3
IPI* inhibition of the *E. coli* CT and UT *gmrS/D* restriction endonucleases ^{1/3}

pIPI* <i>pgmrS/D</i>	1 mIPI* CT	2 IPI* CT	3 CT	4 UT	5 IPI* UT	6 -	7 <i>ipI</i> locus	8 glt-HMC DNA
Phage								
T4	+	+	+	+	+	+	<i>ipI</i>	+
T4 <i>ipI</i>	-	+	-	+	+	+	<i>ipI</i>	+
RB15	-	+	-	-	-	+	<i>ip5-7-8-9</i>	+
K3	+	+	+	-	-	+	<i>ip7-8-9</i>	+
T2	-	+	-	-	-	+	<i>ip5-7</i>	+
T6	-	+	-	-	-	+	<i>ip5-6</i>	+
RB70	-	+	-	-	-	+	<i>ip5</i>	+
C16	+	+	+	-	-	+	<i>ipI</i>	+
RB69	+	+	+	+	-	+	<i>ipI</i>	+
AC3	+	+	+	+	-	+	<i>ipI</i>	+
RB27	+	+	+	+	-	+	<i>ipI</i>	+
D8	+	+	+	+	-	+	<i>ipI</i>	+
RB30	+	+	+	+	-	+	<i>ipI</i>	+
DDYI	+	+	+	+	-	+	<i>ipI</i>	+
N209	+	+	+	+	+	+	<i>ipI</i>	+
RB49, λ, T3	+	+	+	+	+	+	<i>ipI</i>	+

^{1/3} Phage growth is given as + (efficiency of plating = 1 relative to growth on *E. coli* DH10B (column 6) or - (efficiency of plating <10⁻⁶). Phages are shown together with the *ipI* locus genes they contain (column 7) and the presence or absence of sugar-hydroxymethylcytosine modification in their DNA (column 8). Clones: all plasmids are tested in *E. coli* DH10B that lacks restriction-modification systems. Compatible plasmids are CT=*pgmrS/gmrD* (pBexB from pBel6BAC11 (columns 1-3), or UT=*pgmrS-i-gmrD* from *E. coli* UT189 in pBel6BAC11 (columns 4 & 5); and mIPI* or IPI* (pdI4 and pdI5 encoding mIPI* (A40T, E53K) (column 1) and IPI* (columns 2 & 5) in pET15B).