Although natively unfolded proteins are being observed increasingly, their physiological role is not well understood. Here, we demonstrate that the Escherichia coli YefM protein is a natively unfolded antitoxin, lacking secondary structure even at low temperature or in the presence of a stabilizing agent. This conformation of the protein is suggested to have a key role in its physiological regulatory activity. Because of the unfolded state of the protein, a linear determinant rather than a conformational one is presumably being recognized by its toxin partner, YoeB. A peptide array technology allowed the identification and validation of such a determinant. This recognition element may provide a novel antibacterial target. Indeed, a pair-constrained bioinformatic analysis facilitated the definite determination of novel YefM-YoeB toxin-antitoxin systems in a large number of bacteria including major pathogens such as Staphylococcus aureus, Streptococcus pneumoniae, and Mycobacterium tuberculosis. Taken together, the YefM protein defines a new family of natively unfolded proteins. The existence of a large and conserved group of proteins with a clear physiologically relevant unfolded state serves as a paradigm to understand the structural basis of this state.

The thermodynamic hypothesis of protein folding, as was introduced more than 40 years ago, suggests that the folded state of a given protein represents a global minimum of free energy (1). Although this theory is widely valid, there is a considerable group of natively unfolded proteins (as were first denoted by Mandelkow and coauthors (2)) which rather favors the thermodynamically unfolded state (3–6; for a recent review on natively unfolded proteins see Ref. 5). The unfolded state of this group of proteins does not signify a requirement for the activity of molecular chaperones to overcome a large energetic barrier to attain a global minimum energy, but a truly energetically favorable unfolded state. The naturally unfolded state is also distinct from the misfolded state in which proteins getically favorable unfolded state. The natively unfolded state serves as a paradigm to understand the structural basis of this state.

Received for publication, July 29, 2003, and in revised form, November 5, 2003
Published, JBC Papers in Press, December 14, 2003, DOI 10.1074/jbc.M308263200

Izhack Cherny and Ehud Gazit‡
From the Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

The YefM Antitoxin Defines a Family of Natively Unfolded Proteins
IMPLICATIONS AS A NOVEL ANTIBACTERIAL TARGET*

The absolute lack of TA systems in eukaryotes, as opposed to their ubiquitous presence in bacteria and archaea, makes the systems a very attractive antibacterial target. Unlike conventional antibiotics, there is no need for the external introduction of toxic material that may affect the host as well. The blockage of the toxin-antitoxin physical interaction may therefore serve as a critical element in the function of the TA1 module. Many “damaged” or misfolded proteins are identified and eliminated by the ClpXP system. These unfolded target proteins may be recognized by ectopic exposure of hydrophobic amino acids, which are normally buried within the hydrophobic core of the protein. Therefore, we assumed that ClpXP recognizes the unfolded Phd protein based on its structural property because it may appear as damaged protein.

TA systems were also identified on chromosomes in both bacteria and archaea, but not in eukaryotes (14–19). These systems share the same paradigm of a stable toxin and an unstable antidote, organization as a polycistronic operon, and the small size of the protein components (70–100 amino acids). Although TA systems are widely present, their physiological role is not fully understood. It is assumed that the systems play a significant role in survival under stringent conditions (14–19).

The work was supported in part by European Union Contract QLK3-CT-2001-00277. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Molecular Microbiology and Biotechnology, Tel-Aviv University, Green Bldg., Rm. 221, Ramat-Aviv, Tel-Aviv 69978, Israel. Tel.: 972-3-640-9030; Fax: 972-3-640-9407; E-mail: ehudg@post.tau.ac.il.

* This work was supported in part by European Union Contract QLK3-CT-2001-00277. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Molecular Microbiology and Biotechnology, Tel-Aviv University, Green Bldg., Rm. 221, Ramat-Aviv, Tel-Aviv 69978, Israel. Tel.: 972-3-640-9030; Fax: 972-3-640-9407; E-mail: ehudg@post.tau.ac.il.

1 The abbreviations used are: TA, toxin-antitoxin; FTIR, Fourier transform infrared; GST, glutathione S-transferase; PBS, phosphate-buffered saline; SPR, surface plasmon resonance; TBS, Tris-buffered saline; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.
some major pathogens. The unfolded YeFM-like proteins are an attractive target for the development of antibacterial agents because the toxin partner of the TA module recognizes a linear determinant with the antitoxin, which could be mimicked by a therapeutic agent.

EXPERIMENTAL PROCEDURES

Gene Sequence Identification and Alignments—Sequences related to the yeFM and yoeB genes of E. coli were identified by a pair-constrained bioinformatic analysis. Sequences were identified using TBLASTN and PSI-BLAST searches (20) of nonredundant microbial genomes data base at NCBI (www.ncbi.nlm.nih.gov/BLAST/). Putative yeFM and yoeB homologs were obtained and examined for a toxin-antitoxin gene pair module in the chromosome. Low homology unpaired sequences were discarded. Alignments were produced by ClustalW (21) with default settings and edited using JALVIEW editor.

Cloning of the System Genes into the pBAD-TOPO Expression Vector—DNA fragments containing the coding sequence of yeFM, yoeB, and both yeFM-yoeB were produced by PCR using the chromosomal DNA of E. coli K-12 MC1061 and the primers ATG-YEFM (5'-ATGAACTGTA-CAAAAAGG-3') and YEFMEND (5'-GACAACTGTATTGCTACTA-ATG-3') to amplify the yeFM gene; GTG-YOE (5'-GTGAATAAATCTGGTGTCG-3') and YOEEND1 (5'-TGAACTGTTAAATAATGA-ACGAC-3') to amplify the yoeB gene; and ATG-YEFM and YOEEND2 (5'-CTCTGTTGTTCAATTGAATAA-3') to amplify the two genes together. Using the pBAD-TOPO TA cloning kit (Invitrogen), were cloned into the pBAD-TOPO vector to generate pBAD-yeFM, pBAD-yoeB, and pBAD-yeMyoeB. The plasmids were transformed into an E. coli TOP10 strain (Invitrogen).

Growth Rate Analysis—E. coli TOP10 bacteria transformed with pBAD-yeFM, pBAD-yoeB, and pBAD-yeMyoeB were cultured overnight in LB broth supplemented with 100 μg/ml ampicillin at 37 °C. The next day, the three cultures were diluted and adjusted to an absorbance of ~0.1 (A490) in LB-ampicillin. Next, each culture was divided into two equal volumes; at time zero, the first half was added with 0.2% L-arabinose to induce expression of the target gene and the second half with 0.2% L-glucose to suppress low transcription from the pBAD promoter. All cultures were grown at 37 °C/200 rpm, and samples were taken sequentially approximately every 40–60 min for 9 h. Cell density was measured by its absorbance at 600 nm. To inspect the growth rate for gene induction during logarithmic growth phase, the same analysis assay as above was conducted, with the exception of the time of induction. Cultures were divided, and expression was induced (or suppressed) at the time they had reached an absorbance of ~0.45 (A490).

Colony Formation Analysis—E. coli TOP10 bacteria transformed with pBAD-yeFM, pBAD-yoeB, and pBAD-yeMyoeB were grown in LB broth at 37 °C containing ampicillin as indicated. After overnight growth, cultures were diluted to an A490 of 0.01 in LB-ampicillin diluted with fresh medium until an A490 value of 0.5 was reached. At that point, cells were diluted 10⁵–10⁶ times in 10-fold dilution steps and applied as 5-μl drops on LB-ampicillin-agar plates containing arabinose in the following decreasing arabinose dilutions: 0.2%, 0.1%, 0.05%, 0.02%, 0.005%, and 0.0005%. In addition, a negative control plate without arabinose and supplemented with 0.2% glucose was plated. All plates were incubated at 37 °C for at least 20 h.

Cloning, Expression, and Purification of YeFM from E. coli—The DNA fragment containing the coding sequence of yeFM, flanked by primer-encoded EcoRI and HindIII sites, was produced by a PCR using E. coli strain MC1061 chromosome as template and oligonucleotide primers YEFSSTART (5'-AAAGAAGAATTCCATGGAAATCATC-3') and YEFSEND (5'-CTCTGTTGTTCAATTGAATAA-3') with default settings and edited using JALVIEW editor. The product was digested with EcoRI and HindIII enzymes (New England Biolabs), cloned into the EcoRI and HindIII restriction sites of the pET24a expression vector in fusion to GST, and transformed into E. coli BL21(DE3) pLysS. Bacteria were grown, expressed, and lysed in the same manner described above for GST-YeFM fusion. The supernatants were applied to a glutathione-Sepharose column (Amersham Biosciences) preequilibrated with PBS, pH 7.3. The bound protein was eluted using 10 μl of 50 mM Tris-HCl, pH 8.0, 10 mM glutathione. Eluted fractions containing the GST-YeFM protein were collected and assessed quantitatively by Coomassie staining of SDS-PAGE. Circular Dichroism (CD)—CD spectra were recorded using an Aviv 202 spectropolarimeter equipped with temperature-controlled sample holder and a 5-mm path length cuvette. Mean residual ellipticity, [θ], was calculated as

\[
[\theta] = \left[ 100 \times \theta \times m \times l \right] \times L
\]

where θ is the observed ellipticity, m is the mean residual weight, c is the concentration in mg/ml, and L is the path length in cm. All experiments were performed in PBS, pH 7.3, at a protein concentration of 10 μM. For thermal denaturation experiments, samples were equilibrated at each temperature for 0.5 min, and CD ellipticity at 222 nm and 217 nm was averaged for 1 min.

Fourier Transform Infrared Spectroscopy (FTIR)—Infrared spectra were recorded using a Nicolet Nexus 470 FTIR spectrometer with a DTGS detector. The sample, 1 μl of lyophilized YeFM suspended in 30 μl of PBS in D₂O, pD 7.3, was suspended on a CaF₂ plate. The measurements were taken using a 4 cm⁻¹ resolution and 2,000 scans averaging. The transmittance minima values were determined by the OMNIC analysis program (Nicolet Imaging systems). Analysis of YeFM Stability—Overnight culture of E. coli carrying the pBAD-yeFM plasmid was grown at 37 °C/200 rpm in LB broth to stationary phase (A490 = 1.4). YeFM expression was then induced for 10 min with 0.2% arabinose and subsequently treated with 200 μg/ml rifampicin and 0.2% glucose to repress further expression from pBAD promoters. Aliquots of 2 ml were filtered and the absorption at 260 nm was monitored for 20 min. Protein samples were centrifuged and the protein fractions were collected. After centrifugation and analyzed by Western blot (see below) to assess YeFM quantity in bacteria. Densitometer assessment of YeFM was achieved using an ImageScanner (Amersham Biosciences) and the Image Master one-dimensional prime (version 3.01) program (Amersham Biosciences).

Western Blot Analysis— Aliquots (2 ml) were centrifuged at 14,000 rpm for 5 min at 4 °C and resuspended in 80 μl of double-distilled water. Samples of 60 μl were added to 20 μl of 4× sample buffer, and the remaining 20 μl was used to quantify the total protein using the Coomassie Plus protein assay reagent (Pierce). Aliquots containing

yeFM proteins were separated from the GST using 16 units of factor Xa protease (Novagen) at 1 mg of YeFM fusion. After a 14-h incubation at 37 °C, the reaction was terminated by the addition of 1 ml phenylmethylsulfonyl fluoride. Two different methods were applied for YeFM purification. In the first method, gel filtration was conducted to remove the GST and linker protein (~40 kDa) from YeFM (~11 kDa) using a Sepharose HR 10/30 (fast protein liquid chromatography) gel filtration column (Amersham Biosciences) and a fast protein liquid chromatography instrument (Amersham Biosciences). Proteins were eluted with PBS, pH 7.3, 0.8 ml/min, and a peak that included the ~11-kDa YeFM proteins was collected after 13 min. Fractions containing the YeFM protein were completely purified using 1 μmol of immobilized glutathione-agarose (Sigma) agitated for 16 h at room temperature. At this point, YeFM was greater than 95% pure as estimated by Coomassie staining of SDS-PAGE. In the second purification method, the YeFM and GST protein mixture was divided into 0.5-ml fractions, boiled for 10 min, and then centrifuged at 14,000 rpm for 10 min. The supernatants, containing the purified YeFM, were collected and united.

To determine YeFM concentration, tyrosine absorbance measurement in 0.1 M KOH was used. Protein concentrations were calculated using the extinction coefficients of 2391 M⁻¹ cm⁻¹ (293.2 nm in 0.1 M KOH) for single tyrosine.

The molecular mass of YeFM was verified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry using a Voyager-DE STR Biospectrometry work station (Applied Biosystems).

Cloning, Expression, and Purification of GST-YeFM from E. coli—The DNA fragment containing the coding sequence of yoeB, flanked by primer-encoded EcoRI and HindIII sites, was produced by a PCR using E. coli strain MC1061 chromosome as template and oligonucleotide primers YOEBSSTART (5'-AAAGAAGAATTCCATGGAAATCATC-3') and YOEBSEND2 (5'-CTCTGTTGTTCAATTGAATAA-3'). The product was digested with EcoRI and HindIII enzymes (New England Biolabs), cloned into the EcoRI and HindIII restriction sites of the pET24a expression vector in fusion to GST, and transformed into E. coli BL21(DE3) pLysS. Bacteria were grown, expressed, and lysed in the same manner described above for GST-YeFM fusion. The supernatant was applied to a glutathione-Sepharose column (Amersham Biosciences) preequilibrated with PBS, pH 7.3. The bound protein was eluted using 10 μl of 50 mM Tris-HCl, pH 8.0, 10 mM glutathione. Eluted fractions containing the GST-YeFM protein were collected and analyzed quantitatively by Coomassie staining of SDS-PAGE.
equal total protein amounts were loaded on a Tris-Tricine SDS 15% polyacrylamide slab gel. After electrophoresis, the proteins were electrotransferred to polyvinyldene difluoride membrane filters (Bio-Rad). The detection of YefM was performed using anti-YefM serum raised in rabbit. The membrane was then incubated with peroxidase-conjugated anti-rabbit antibodies, and YefM proteins were detected through the enhanced chemiluminescence reaction after an exposure to a sensitive film.

**Amino Acid Composition and Charge-Hydrophobicity Values Analysis**—The rate of occurrence of each amino acid in the YefM family proteins (P_M) was determined by averaging its 30 frequencies in each of the 30 YefM homolog sequences. The general amino acid occurrence statistics (P_M) were compiled by the Rockefeller authors using the NCBI data base (prowl.rockefeller.edu/aainfo/masses.htm). The comparison ordnates between the amino acid occurrences are given by their fractional difference: \( \text{Var}(P_M) = P_M \times P_{M} \). The variances of these ratios were calculated as \( \text{Var}(P_M) \). This was accepted (24). Surprisingly, overexpression of YefM was inhibited the bacterial growth to maximum \( A_{600} \) of \(-0.15 \) (Fig. 3B). Overexpression of both YefM and YoeB as an operon abolished this toxic effect, indicating a TA relationship between YoeB and YefM (Fig. 3C), as accepted (24). Surprisingly, overexpression of YefM alone had displayed an effect on cell growth similar to that by YoeB (Fig. 3A). The same results had been witnessed when cells expressing the system genes were induced during the logarithmic growth stages (Fig. 3D-F): 0.2% arabinose was added to the different cultures at the time they reached \( A_{600} \) of \(-0.45 \). In the cases of YefM or YoeB expression, absolute growth inhibition had been observed after less than 1 h (Fig. 3D and E) as cells reached \(-0.7 A_{600} \) whereas the expression of both genes together enabled normal growth (Fig. 3F).

To confirm that the YefM is an actual antitoxin, we tested the colony formation capability of each of the clones at decreasing expression levels (Fig. 3G). On the whole, yefM clones have consistently demonstrated a certain degree of growth in all arabinose concentrations, whereas yoeB clones did not form colonies at most concentrations. Moreover, in the presence of 0.005% arabinose, growth of the yoeB clone was disabled, whereas the yefM clone still demonstrated clear growth, indicating that YoeB is a real toxin whereas YefM displays toxicity upon high expression levels.

**Biophysical Characterization of YefM**—YefM was purified as described under “Experimental
Procedures,” either by performing gel filtration (obtaining ~0.1 mg/ml) or by boiling GST and YefM proteins subsequent to factor Xa cleavage (~0.35 mg/ml).

The far-UV CD spectra of the purified YefM protein (in both purification methods) at increasing temperatures (25, 37, 42 °C) show a typical random coil pattern with a minimum in the vicinity of 200 nm (25), with only slight changes in spectra caused by an increase in temperature (Fig. 4A). FTIR spectroscopy also indicates that YefM protein is random coil-structured (Fig. 4B). The FTIR spectrum of the purified YefM (room temperature) showed a transmittance minimum at 1,643 cm⁻¹ relating to random coil structure (26).

A thermal denaturation experiment (Fig. 4C) proves that YefM keeps a consistent predominant random coil structure at the entire temperature range, as a continuous temperature increase of the YefM sample from 2 to 80 °C did not significantly shift the CD ellipticity at 222 nm or at 217 nm (wavelengths specifying for maximum CD ellipticity of α-helix and β-sheet structures, respectively), implying that the structure remained unchanged. Another support for the natively unfolded state of YefM comes from its extraordinary solubility during boiling (Fig. 4D).

**Determination of YefM Stability in Vivo**—To get insight into the structural stability of the YefM antitoxin in its native state within cells, we have examined its proteolytic stability in vivo. For that end, we performed a short expression of YefM followed by its full repression under stationary growth. Analysis of YefM levels in *E. coli*, before and after repression at different intervals, reveals that the YefM antitoxin is proteolytically unstable (Fig. 5). YefM degraded in vivo with a half-life of approximately 1 h. This result correlates with expected features of TA systems, where the antitoxin proteins are preferred substrates for a protease, and is consistent with the half-life reported for the unfolded Phd antitoxin (13).

**Amino Acid Composition of YefM Family Proteins**—To visualize differences between amino acid composition of the YefM proteins and the general amino acid composition and to gain further insight into the role of the sequence in providing disorder characteristics, we have compared the general occurrence of each amino acid in relation to its mean occurrence in YefM proteins. As shown in Fig. 6A, YefM family proteins are considerably enriched in Met and Glu (30–50%) and substantially depleted in Trp, Cys, Pro, Phe, and Gly (~50%). The obtained results for these amino acids are significant, with a p value < 0.001, as determined by a one-sample t test. Other amino acids do not display significant enrichment or depletion from the general occurrence of amino acids.

**Charge-Hydrophobicity Relationships in the YefM Family Proteins**—A comparative study that was published by Uversky et al. (3) demonstrated well that it is possible to predict whether a given sequence encodes a folded or natively unfolded protein by a two-dimensional plot of the overall hydrophobicity and the net charge of the studied proteins. To assess whether the charge-hydrophobicity properties of the YefM family proteins correlate with those previous findings, we have examined these relationships for YefM, Phd, and their homolog sequences as described previously (3) (Fig. 6B). Unexpectedly, the YefM-Phd family proteins were found to be localized mostly within the defined “folded region” of the plot. The localization of Phd protein and its homologs is indistinguishable from the YefM homologs.

**Identification of YefM Recognition Determinant**—On the basis of the YefM natively unfolded structure, we assumed a linear determinant rather than a conformational one to be recognized by its toxin partner. To identify this determinant in...
the YefM sequence, we have designed an array consisting of 41 overlapping tridecamer peptides corresponding to amino acids residues 1–12 up to 80–92 of the whole YefM sequence in successive order with 2-amino acid shifts (Fig. 7A) synthesized on a cellulose membrane matrix. The YefM fragments capable of binding GST-YoeB fusion were identified by immunoblotting. Using a low stringency procedure to obtain maximum putative interaction sites, we have identified three such regions. As seen in Fig. 7A, first region included three tridecamer peptides (YefM11–23–YefM15–27) in decreasing binding capacity, including the sequence RTISYSEARQNLSATMM (underlined sequence represents major bound site); the second region included the single YefM33–45 peptide sequence, APILITRQNGEAC; the third region comprised the two YefM75–87 and YefM77–89 peptides, which cover the MDSIDSLKSGK-TEKD sequence.

To verify our results, we used a second peptide array membrane comprising those regions with the intention of performing a high resolution analysis of the putative binding sites (Fig. 7B). We used a high stringency procedure (see “Experimental Procedures”) to minimize unspecific binding of the GST-YoeB fusion protein or antibodies. The examined sites were extended to include YefM8–31 as the first region, YefM29–48 as the second region, and YefM72–92 as the third region. The shift between each arrayed tridecamer peptide was reduced to a single amino acid. Of all examined regions, the YefM11–23 peptide (RTIS-YSEARQNLS) was detected as the best YoeB binding sequence. SPR analysis was used to quantitative determine the affinity between the YoeB toxin and the YefM11–23 peptide fragment. The recognition determinant sequence peptides were immobilized onto the sensor chip, and the kinetics of GST-YoeB binding and dissociation was estimated at 12.5, 25, and 50 nM concentrations (Fig. 8). According to data analysis, a $k_a$ of 3.06 × 10^3 (M$^{-1}$ s$^{-1}$) and a $k_d$ of 1.22 × 10^3 (M$^{-1}$ s$^{-1}$) were calculated (arithmetic mean). Accordingly, an equilibrium con-
E. coli strain TOP10 carrying one of the pBAD-TOPO vectors expressing YefM (A and D), YoeB (B and E), or YefM-YoeB together as an operon (C and F) were grown in LB-ampicillin medium at 37 °C. Transcription of the respective genes was induced by the addition of 0.2% arabinose (closed circles) at two different growth phases: stationary (at time zero (A–C)) and logarithmic (when cultures reached $A_{600}$ of 0.45 (D–F)). In parallel, equal culture volumes were added with 0.2% glucose as a negative control (open circles). G, effect of overexpressing YefM, YoeB, or YefM-YoeB together in a TOP10 strain. Dropouts of the different clones (as indicated) were plated on arabinose gradient plates in 10-fold dilutions and incubated for 20 h at 37 °C. The arabinose gradient plates are in the following order (top to bottom): 0%, 0.0005%, 0.005%, 0.02%, 0.05%, 0.1%, and 0.2%. Plates missing L-arabinose were added with 0.2% glucose.
stant \( K_D \) of 0.4 \( \mu M \) was determined for the YoeB-YefM\(_{11-23} \) complex. This dissociation constant is consistent with a specific binding between the toxin and the peptide fragment.

The Arginine in Position 19 Is Essential for YefM-YoeB Interaction—Alongside the verification of the major binding sequence, we tried to detect a single amino acid that would be crucial for YefM-YoeB interaction. The identified binding sequence is rather conserved through the YefM-Phd protein families. However, two amino acids are notably conserved within: arginine (position 19) and leucine (position 22), as seen in Fig. 2A. We have examined the binding capability of a GST-YoeB fusion to a cellulose membrane array using tridecamer peptides corresponding to the YefM\(_{11-23} \) sequence, containing Arg-19 or Leu-22 replacements to alanine or glycine (Fig. 7C).

Although L22A and L22G replacements only attenuated the binding of YoeB, R19A or R19G totally interrupted the binding, suggesting that the arginine in position 19 is essential for the binding of the YoeB toxin.

**DISCUSSION**

Non-native protein structures attract an increasing degree of intention because of their abundance on the one hand and the lack of understanding of their physiological significance on the other. Identification of distinct families of natively unfolded proteins, understanding their conservation on the structural level, and understanding their physiological role are therefore of high importance. Here, using a combination of bioinformatics and biophysical and physiological analysis, we define a new
family of natively unfolded proteins, the YefM-Phd family. Using a pair-constrained bioinformatic approach, we were clearly able to demonstrate that members of the family are present in a large number of bacteria. Although the level of homology within the antitoxins family is relatively low (Fig. 2A), we were surprised to find Phd homologs that share higher percentage of homology to YefM than Phd does (Y. enterocolitica, K. pneumoniae, and S. typhimurium). Although YefM and Phd proteins share very low sequence homology, the key feature that the proteins share is the natively unfolded state at physiological temperatures (Fig. 4 and Refs. (11 and 27)). Because both Phd-Doc (12) and YefM-YoeB (Fig. 3) are proven to be functional TA systems, these findings may suggest that Phd and YefM antitoxins have evolved from a common ancestor system and that at a certain point in the past the antitoxin may have branched out to establish new TA systems consisting of different toxins.

Interestingly, the level of homology within the YoeB family (Fig. 2B) appears to be significantly higher compared with the YefM family of proteins (Fig. 2A). The level of conservation observed with the YoeB proteins is highly consistent with a toxic activity that explicitly targets specific cellular determinants and that requires a well defined fold such as a keylock or induced fit recognition. On the other hand, the low degree of conservation of the extended YefM-Phd family is consistent with a protein missing a clear structural recognition and/or catalytic activity that otherwise requires a defined configuration. It is important to note that YefM and Phd proteins could be irregularly conjugated to a Doc-like or YoeB-like toxins, two families of toxins that could not be aligned and do not share any substantial homology. It is more consistent with a family of protein that is essentially designed to be recognized as a damaged protein and does not represent an interactive or catalytic scaffold. Moreover, the relatively small area of YefM which shows the highest level of conservation was identified to include the target of linear recognition by the YoeB protein (Fig. 7).

Physiological assays have verified that the yoeB gene encodes a toxin that is lethal or inhibitory to host cells and that yefM encodes an antitoxin that prevents the lethal action of the toxin (Fig. 3 and Ref. 24). Unexpectedly, upon overexpression, YefM inhibited the bacterial growth. However, the dose-dependent behavior of toxicity may suggest that it is an artifact of overexpression rather than a true physiological phenomenon (Fig. 3G).

It is hypothesized that the proteolytic stability difference of the TA system components arises from their thermodynamic stability difference. YefM strongly supports this hypothesis as it was demonstrated to be a natively unfolded protein. Furthermore, among all structurally described antitoxins (Phd of P1 (11, 27), ParD of RK2/RP4 (28), CcdA of F (29), and ϵ of pSM19035 (30)), YefM is the most unstable protein. One of the general structural characteristics of a natively unfolded protein is the lack of secondary structures. At 37 °C, the Phd antitoxin seems to be in a largely unfolded, random coil conformation as well (11). However, at 4 °C or at 37 °C in the presence of the trimethylamine N-oxide chemical chaperone, Phd folds into an ordered protein containing ~45% α-helix. Analysis of the YefM far-UV CD spectra yields a low content of ordered secondary structure (α-helices and β-sheets) and does not change even at low temperature of 2 °C (Fig. 4, A and C) or upon the addition of trimethylamine N-oxide chemical chaperone (data not shown). YefM was also confirmed to be random coil by FTIR analysis (Fig. 4B). Additional substantiation for YefM being a

![Image](https://example.com/image.png)
most unstructured protein comes from its unusual resistance to aggregation upon boiling (Fig. 4D), which is consistent with a lack of secondary structure elements that mediate aggregate formation through intermolecular association (see Fig. 4D).

Indeed, YefM is proteolytically unstable in vivo (Fig. 5), suggesting that it maintains an unfolded conformation within cells. This feature further correlates with the observed proteolytic instability of other antidotes, as Phd and MazE (13, 19).

It was suggested recently that the relations between sequence and disorder proteins include amino acid compositional bias and high predicted flexibility (6, 31). According to this study, it was demonstrated that natively unfolded proteins are substantially depleted in Trp, Cys, Phe, Ile, Tyr, Val, Leu, and Asn and substantially enriched in Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys. Indeed, we found that the same amino acid compositional bias is valid when comparing the occurrence of the above disordered sequences (using the ALL-disorder sequences data base (31)) with the general occurrence of amino acids (prowl.rockefeller.edu/aainfo/masses.htm) (data not shown). In addition, the depleted amino acids were shown to correspond to low flexibility residues, whereas the enriched amino acids corresponded to high flexibility ones (6). The flexibility ranking is based on a scale developed by Vihinen et al. (32) and reflects the propensity of a given residue to be buried or exposed (i.e. low or high flexibility, respectively) in the crystal structure of globular proteins. However, the amino acid composition of the natively unfolded YefM family proteins is rather different (Fig. 6A). Although both the studied disordered proteins and the YefM family proteins are depleted significant in Trp, Cys, and Phe, the YefM proteins are depleted further in Gly and Pro, amino acids considered as disorder-promoting (6, 22). Moreover, Glu is the sole amino acid that seems to be significantly enriched in both. Noteworthy, the most rigid residues (Trp, Cys, and Phe) remained depleted in both surveys, insinuating an essential importance in the absence of core-forming side chains in the coding of intrinsically disordered sequences.

Recent comparative studies suggested that it is possible to predict whether a given sequence encodes a folded or natively unfolded protein (3–5). This suggests that a natively unfolded protein must possess the combination of low mean hydrophobicity and relatively high net charge under physiological conditions. However, the majority of the YefM family proteins do not correlate with this determination, including YefM and Phd proteins (Fig. 6B). Obviously, this result is coupled with the
unique amino acid compositional bias of the YefM family proteins mentioned above, which does not fit the established characteristics of disordered sequences. The relative lack in high flexibility side chains (e.g., Lys, Pro, Gly, Ser, and Gln) together with an insufficient depletion in hydrophobic rigid side chains (e.g., Ile, Tyr, Val, and Leu), account for the relatively low net charge and rather high overall hydrophobicity that characterize the YefM family. Furthermore, in the case of the YefM family proteins, we propose that the lack of aromatic residues, rather than hydrophobic, maintains the disordered state of YefM. As seen in Fig. 6A, the depletion in the aromatic residues Phe and Trp, unlike other hydrophobic residues, is conserved through the YefM family. The lack of aromatic moieties is consistent with the lack of organized and packed hydrophobic core.

As discussed in the introduction section, the TA system may serve as an excellent target for antibacterial agent. One approach is to prevent the toxin and antitoxin components from interacting in vivo, which would trigger their inhibitory (or lethal) effect on cell growth. Here, we have identified the molecular recognition sequence within the YefM protein using peptide array (Fig. 7) and SPR analysis (Fig. 8). In the future we intend to use this information for the design of agents that will affect the YefM-YoeB interaction.

Acknowledgments—We thank members of the Gazit laboratory for helpful discussions, Kfir Madjar for assisting in protein purifications and membrane procedures, Dr. Susana Shochat (Hebrew University, Jerusalem) for assistance with BIAcore analysis, and Prof. Eliora Ron and her laboratory for supplying purified chromosomal DNA.

REFERENCES
1. Anfinsen, C. B. (1973) Science 181, 223–230
2. Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) J. Biol. Chem. 269, 24290–24297
3. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Proteins Struct. Funct. Genet. 41, 415–427
4. Uversky, V. N. (2002) Protein Sci. 11, 739–756
5. Uversky, V. N. (2002) Eur. J. Biochem. 269, 2–12
6. Dunker, A. K., Lawson, J. D., Brown, C. J. Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hips, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C.-H., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovia, Z. (2001) J. Mol. Biol. 315, 713–739
7. Dobson, C. M. (1999) Trends Biochem. Sci. 24, 329–332
8. Rochet, J. C., and Lansbury, P. T., Jr. (2000) Curr. Opin. Struct. Biol. 10, 60–68
9. Gazit, E. (2002) Curr. Med. Chem. 9, 1725–1735
10. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovia, Z. (2002) Biochemistry 41, 6573–6582
11. Gazit, E., and Sauer, E. T. (1999) J. Biol. Chem. 274, 2652–2657
12. Lehnherr, H., Maguin, E., Jafari, S., and Yarmolinsky, M. B. (1993) J. Mol. Biol. 233, 414–428
13. Lehnherr, H., and Yarmolinsky, M. B. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3274–3277
14. Christensen, S. K., and Gerdes, K. (2003) Mol. Microbiol. 48, 1389–1400
15. Pedersen, K., Zavialov, A. V., Pavlov, M. Y., Elf, J., Gerdes, K., and Ehrenberg, M. (2003) Cell 112, 131–140
16. Pedersen, K., Christensen, S. K., and Gerdes, K. (2002) Mol. Microbiol. 45, 501–510
17. Christensen, S. K., Mikkelsen, M., Pedersen, K., and Gerdes, K. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 14328–14333
18. Gerdes, K. (2000) J. Bacteriol. 182, 561–572
19. Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 6059–6063
20. Aartschot, S. F., Madden, T. L., Schiffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
21. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
22. Radivojac, P., Obradovia, Z., Brown, C. J., and Dunker, A. K. (2003) Pac. Symp. Biocomput. 8, 216–227
23. Pomerantsev, A. P., Golovlyov, I. R., Ohara, Y., Mokrievich, A. N., Obuchi, M., Norvigst, A., Kuoppa, K., and Pavlov, V. M. (2001) Nucleic Acids Res., 24, 210–222
24. Grady, B., and Hayes, F. (2003) Mol. Microbiol. 47, 1439–1452
25. Jenness, D. D., Sprecher, C. A., and Johnson, W. C. (1976) Biopolymers 15, 513–521
26. Harris, P. I., and Chapman, D. (1995) Biopolymers 37, 251–263
27. Gazit, E., and Sauer, E. T. (1999) J. Biol. Chem. 274, 16813–16818
28. Oberer, M., Lindner, H., Glatter, O., Kretzky, C., and Keller, W. (1999) Biol. Chem. 380, 1413–1420
29. Thi, M. H. D., Besenze, J., Wynne, L., and Backmann, J. (2000) J. Mol. Biol. 299, 1373–1386
30. Camaioni, A. G., Marseletz, R., Belhka, J., Ayora, S., Welle, K., Meihart, A., Lenz, E., Staiger, W., Welle, H., and Alonso, J. C. (2002) Mol. Microbiol. 46, 1701–1713
31. Romero, P., Obradovia, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, K. (2001) Proteins Struct. Funct. Genet. 42, 38–48
32. Vihinen, M., Torkkla, E., and Rikhonen, P. (1994) Proteins 19, 141–149