The complete nucleotide sequences of over 37 microbial and three eukaryote genomes are already publicly available, and more sequencing is in progress. Despite this accumulation of data, newly sequenced microbial genomes continue to reveal up to 50% of functionally uncharacterized "anonymous" genes. A majority of these anonymous proteins have homologues in other organisms, whereas the rest exhibit no clear similarity to any other sequence in the data bases. This set of unique, apparently species-specific, sequences are referred to as ORFans. The biochemical and structural analysis of ORFan gene products is of both evolutionary and functional interest. Here we report the cloning and expression of Escherichia coli ORFan ykfE gene and the functional characterization of the encoded protein. Under physiological conditions, the protein is a homodimer with a strong affinity for C-type lysozyme, as revealed by co-purification and co-crystallization. Activity measurements and fluorescence studies demonstrated that the YkfE gene product is a potent C-type lysozyme inhibitor (K_i ~ 1 mM). To denote this newly assigned function, ykfE has now been registered under the new gene name IvY (inhibitor of vertebrate lysozyme) at the E. coli genetic stock center.

Despite the accumulation of sequence information from a large diversity of species and phyla, newly sequenced bacterial genomes continue to reveal a high proportion of genes of unknown function (1), including a significant subset of "ORFans" (2), i.e. putative open reading frames (ORFs) without significant similarity to any previously encountered protein (or conceptual translation) sequences. Most genes found in data bases have only been predicted by computer methods and never experimentally validated. It is thus expected that some annotated ORFs, in particular among the ORFans, might not correspond to real genes. In a previous study, we verified the existence of a cognate transcript for 25 Escherichia coli ORFans with a surprising rate of success (92%) (3). Given that most ORFans appear to be transcribed, we have now initiated a systematic expression and structure determination program for the proteins encoded by these (apparently) unique genes. Because three-dimensional structures are more resilient to evolution and change than amino acid sequences, it is expected that some ORFans should exhibit structural similarity to previously described protein families, hence providing some functional hints. Alternatively, targeting ORFans for structure determination is also a suitable strategy to optimize the discovery of original protein folds, one of the goals of structural genomics.

In a pilot study involving five ORFan genes, we succeeded in producing four of them in E. coli as soluble proteins, and we report here the most advanced project, ykfE. YkFE (Swiss-Prot accession number P45552; b0220 in the Blattner database) is a 474-nucleotide-long uncharacterized ORF. It is part of a single gene operon and was found to exhibit a high level of expression during the exponential and stationary phases of E. coli growth (3). The ykfE ORF exhibits an N-terminal signal peptide cleaved to produce the mature protein (5, 6). Initial purification steps and biochemical analyses suggested a strong interaction between this protein and hen egg white lysozyme (HEWL). The existence of a stable complex was confirmed by biophysical analyses, and enzymatic studies revealed the capacity of ykfE to inhibit hen and human C-type lysozymes through a specific interaction. The X-ray structure determination of ykfE, both in isolation (7) and in a complex with HEWL, is currently in progress and should allow us to understand the molecular basis of the ykfE-lysozyme interaction at atomic resolution. To denote its newly assigned function, ykfE has now been registered under the new gene name IvY (inhibitor of vertebrate lysozyme) at the E. coli genetic stock center.

**EXPERIMENTAL PROCEDURES**

**Cloning of ykfE**—The 474-base pair ykfE ORF including its own signal peptide was polymerase chain reaction amplified from E. coli K-12 MG1655 genomic DNA using two DNA polymerase (Roche Molecular Biochemical) primers. Sequences, 5'-TTATACCATGGAACAGGAT-AAGCTC-3' (sense) and 5'-GCTTTAGATCTGATGATGATCATGGCGAGGAT-AAGCTC-3' (antisense) were used. After digestion with NcoI and BglII, the polymerase chain reaction product was cloned into a pQE-60 vector (Qiagen) to express ykfE in phase with a C-terminal His_6 tag (plasmid pQE-0220).

**Expression and Purification of the YkFE Gene Product**—The ykfE gene product (IvY) was expressed by culturing E. coli XLI-Blue carrying the plasmid pQE-0220 in LB + Amp medium. After initial growth at 37 °C, temperature was set at 30 °C when A_600 reached 0.4. Medium was replaced by 1 mM isopropyl-1-thio-β-D-galactopyranoside. Cells were harvested at A_600 around 2–2.5 and resuspended in Buffer A (20 mM sodium phosphate, pH 8.0, 300 mM NaCl containing 1.5% Triton X-100, 1.5% glycerol, and 1 mM MgCl_2); Buffer C (20 mM sodium phosphate, pH 8.0, 300 mM NaCl containing 1.5% Triton X-100, 1.5% glycerol, and 1 mM MgCl_2) was washed with 10 column volumes of Buffer A, followed by 10 column volumes of Buffer B containing 25 mM imidazole, and 5 column volumes of Buffer C containing 70 mM imidazole at a flow rate of 1 ml/min. Elution was performed with a linear gradient over 8
column volumes from 70 to 500 mM imidazole. The recombinant protein was eluted with 150–200 mM imidazole, and fractions were pooled and desalted against 20 mM Tris, pH 8.0, on a fast desalting column HR 10/10 (Amersham Pharmacia Biotech) at a flow rate of 5 ml/min. Protein concentration was determined by UV absorption at 280 nm using molar absorption coefficients calculated on the basis of tyrosine and tryptophan contents (8). Protein purity was assessed by SDS polyacrylamide gel electrophoresis and isoelectrofocusing (IEF) using 3 to 10 pH gradient pre-cast gels (Novex). Puriﬁed molecular weights for the puriﬁed proteins were estimated by gel ﬁltration using a calibrated Superdex 75 HR10/30 column (Amersham Pharmacia Biotech) equilibrated with a 20 mM sodium citrate buffer, pH 6.5, at a ﬂow rate of 0.5 ml/min. The puriﬁed proteins were characterized by mass spectrometry (matrix-assisted laser desorption ionization/time of ﬂight, Voyager DE-RP, PerSeptive Biosystems) and by N-terminal Edman sequencing (473A; Applied Biosystems).

Interaction Measurements—To assay the interaction between Ivy and HEWL, 50 mg of HEWL were loaded on a nickel column at a flow rate of 1 ml/min of 20 mM sodium phosphate, pH 8.0, 300 mM NaCl in the presence or absence of 3 mg of pure Ivy protein. After extensive washing with 20 ml sodium phosphate, pH 8.0, 1 mM NaCl, elution was performed with a linear gradient over 5 column volumes to 1 M imidazole.

Intrinsic protein ﬂuorescence was measured with a Spex Fluorolog3 photon-counting spectrophotometer (Jobin Yvon-Spex, Longjumeau, France) equipped with a 450-watt Xenon source and a cooled photomultiplier. Intrinsic protein ﬂuorescence emission spectra were recorded between 290 and 450 nm from solutions containing the individual proteins and from solutions containing a mixture of proteins excited with 280 nm of light. The degree of protein-protein interaction was determined from the extent of ﬂuorescence quenching observed at 344 nm when spectra of a mixture of proteins were compared with the sum of the individual protein spectra at the same concentration. Interaction-dependent ﬂuorescence quenching was determined in 10 mM Tris-HCl buffer, pH 7.0, 8.0, or 9.0, containing 100 mM NaCl at protein concentrations varying from 1 mM to 0.5 mM.

Determination of the Apparent Dissociation Constant (Kd)—HEWL activity assay was performed at 25 °C in 100 mM potassium phosphate, pH 6.4, using 0.125 mg ml⁻¹ Micrococcus lysodeikticus (Sigma) as substrate. Inhibition studies were carried out by monitoring the change in turbidity associated with the lysis of M. lysodeikticus cells as described previously (9). One unit of HEWL activity was deﬁned as the amount of enzyme causing a decrease in extinction of 0.001 per min at 450 nm. Ki value was determined according to the slow tight binding competitive inhibition model (with no conformational change) (10, 11). The following equation was used,

\[
K_{d} = \frac{K_{t}}{V_{t}} \left( \frac{K_{t}}{V_{t}} + \frac{1}{V_{V}} \right) \left( \frac{K_{t}}{V_{t}} + \frac{1}{V_{E}} + \frac{1}{V_{I}} \right) - \frac{K_{t}}{V_{t}}
\]

where Ki is the apparent dissociation constant, Et is the total enzyme (HEWL) concentration, Vt is the total inhibitor (Ivy) concentration, Vt is the velocity of a given concentration of Ivy, and Vt is the velocity in the absence of inhibitor.

HEWL (70 nm) was pre-incubated with Ivy (0–200 nm) at room temperature for 15 min prior to the addition of the M. lysodeikticus cells. The Ki value was determined by ﬁtting the experimental data onto the Vt/VT curves computed from the above equation (see Fig. 4).

Analysis of the Speciﬁcity of Ivy—The effect of Ivy on C-type lysozyme activity was determined in the presence of 0.5 mM lysozyme, 1 mM Ivy mixture was found to diﬀer signiﬁcantly from that expected when adding the intrinsic ﬂuorescence emission spectra of the individual proteins. An overall 20% quenching of the ﬂuorescence was measured with maximal quenching at 344 nm (Fig. 2). The shape and maximum of the spectrum are consistent with at least one relatively exposed tryptophan being quenched in the lysozyme-Ivy complex. The examination of the concentration dependence of the quenching spectrum showed that the shape of this spectrum was independent of the protein concentration between 1 nM and 1 μM. No reliable Ki value measurement could be obtained because of the insuﬃcient intrinsic ﬂuorescence intensity in the nM concentration range.

Exogenous HEWL is usually added prior to sonication to help the disruption of the E. coli cell wall according to the usual extraction protocol for recombinant proteins. HEWL is then removed during the subsequent puriﬁcation steps. In the case of Ivy, a succession of anomalies led us to suspect a strong interaction between the two proteins. After the puriﬁcation step by metal chelating chromatography on a nickel resin, SDS polyacrylamide gel electrophoresis analyses of the eluted proteins revealed the presence of two bands of nearly identical molecular mass, around 15 kDa, thus close to the predicted value for the mature form of Ivy (without signal peptide). Mass spectrometry and N-terminal sequencing clearly indicated that these fractions consisted of a mixture of two proteins present in equivalent quantities and identiﬁed one of them as the mature Ivy protein (molecular mass = 15.04 kDa) and the other as the exogenous HEWL (molecular mass = 14.3 kDa).

Preliminary results suggested the existence of a speciﬁc interaction between the two proteins. During metal chelating chromatography on a nickel resin, HEWL could only be retained if Ivy (extracted in the absence of lysozyme) had ﬁrst been trapped on the column (see “Experimental Procedures”). The SDS polyacrylamide gel electrophoresis analysis of the eluted fractions conﬁrmed the co-elution of Ivy and HEWL. Finally, the eﬀect of the increase of HEWL concentrations on the IEF behavior of Ivy also suggested a strong interaction (Fig. 1). In the absence of HEWL, the IEF migration of Ivy exhibited two close bands at pI 7.0 and pI 6.7 (for a theoretical pI of 6.74). These two bands most likely correspond to the dimeric and monomeric forms of Ivy in solution. HEWL alone was found to migrate at pI 10. The increase of HEWL concentrations resulted in a decrease of intensity of the Ivy bands at pI 7.0 and pI 6.7 (Fig. 1) and the appearance of an extra band of material reverse migrating into the gel wells, at pI 10.

The HEWL-Ivy interaction was further studied by ﬂuorescence spectroscopy. The ﬂuorescence spectrum of a 1 μM lysozyme, 1 μM Ivy mixture was found to diﬀer signiﬁcantly from that expected when adding the intrinsic ﬂuorescence emission spectra of the individual proteins. An overall 20% quenching of the ﬂuorescence was measured with maximal quenching at 344 nm (Fig. 2). The shape and maximum of the spectrum are consistent with at least one relatively exposed tryptophan being quenched in the lysozyme-Ivy complex. The examination of the concentration dependence of the quenching spectrum showed that the shape of this spectrum was independent of the protein concentration between 1 nM and 1 μM. No reliable Ki value measurement could be obtained because of the insuﬃcient intrinsic ﬂuorescence intensity in the nM concentration range.

EXPERIMENTAL PROCEDURES

E. coli ORFan Encodes a Potent Inhibitor of C-type Lysozyme

RESULTS AND DISCUSSION

FIG. 1. Isoelectrofocusing analysis of pure Ivy complexed with various concentrations of HEWL.
Finally, co-crystallization experiments and the subsequent analysis of the crystal content demonstrated the presence of both proteins, thus suggesting a specific and stable interaction between the two molecules (data not shown). The determination of the complex three-dimensional structure is currently in progress.

In addition to its specific physical interaction with HEWL, Ivy is also a potent inhibitor of HEWL enzymatic activity (Fig. 3). Preliminary experiments showed that in the presence of 1 μg/ml of Ivy, the addition of 1 μg/ml of HEWL produced a nonlinear kinetic with an upward concavity (Fig. 3a, curve b). In contrast, the pre-incubation of Ivy with HEWL for 15 min resulted in kinetic exhibiting a slight downward concavity (Fig. 3a, curve c). These results suggest a slow binding kinetic model for the Ivy-HEWL interaction. In addition, near-complete inhibition is reached for a range of Ivy concentrations comparable with the concentration of HEWL (see Figs. 3b and 4), indicating that Ivy behaves as a slow tight binding inhibitor. A $K_i$ value of about 1 nM was thus estimated by fitting the experimental data (Fig. 4) with Morrison’s equation corresponding to this model (see Eq. 1 and Refs. 10 and 11).

The previous experiments demonstrated the potent inhibitory activity of Ivy on hen egg white lysozyme. We then explored the effect of Ivy on the related proteins of increasing evolutionary divergence. We selected a set of lysozyme and lysozyme-like proteins based on structural similarity using the MMDB data base (14). Using HEWL as initial query (MMDB accession number 1151), λ phage lysozyme (root mean square deviation, 1.3 Å; 21.1% identity), and chitinase (root mean square deviation, 1.9 Å; 11.1% identity) were selected as representatives of structural homologs with low sequence similarity. The inhibitory effect of Ivy was thus tested on the two proteins. Ivy was found to cause a weak inhibition of λ phage lysozyme (Fig. 5). The activity was only reduced by 15% at a molar ratio of 200:1, Ivy:λ phage lysozyme. We found no inhibitory effect of Ivy on chitinase from M. lysodeikticus (15). Ivy was found to strongly inhibit the lysozyme activity in saliva (Fig. 6). Around 50 μg/ml of Ivy is sufficient to observe a decrease of 50% of the activity, which was fully abolished for an Ivy concentration of 0.5 mg per ml of saliva.

On a gel filtration column, Ivy is eluted with an apparent molecular mass of about 30 kDa, indicating that the predominant form in solution is a homodimer, as already suggested by IEF experiments. Fluorescence studies confirmed this model. The fluorescence emission spectrum of HEWL exhibits a broad peak with a maximum at 342 nm and a long wavelength tail typical of relatively exposed tryptophan residues. In contrast, the spectrum of Ivy shows a peak at 334 nm ~25% more intense on an absolute scale and 2.5 times more intense on a per tryptophan scale (Fig. 2). Such intense fluorescence and the relatively short wavelength of maximum emission both argue for tryptophans buried within apolar environments. Furthermore, the shape of the emission spectrum was found to be independent of the protein concentration in a broad 0.5 nm to 1 μm range and appears insensitive to change in pH between 7.0 and 9.0.
E. coli ORFan Encodes a Potent Inhibitor of C-type Lysozyme

and 9.0. The order of magnitude of the dimerization $K_d$ appeared much lower than $10^{-8} \text{ M}$, although no precise measurement could be made in this concentration range. Altogether, these biophysical results suggest that the Ivy homodimer is the physiologically active unit.

C-type lysozyme is an ancient protein whose origin goes back about 500 million years (16). It has long been recognized that bacteria might have evolved a resistance mechanism against an ubiquitous bactericidal enzyme should be advantageous to all murein-containing bacteria, in particular Gram-positive bacteria, $ykfE$/Ivy-like genes are expected to exist in many bacterial genomes, making its ORFan nature a paradox. Indeed, we detected a new putative ortholog of $E. coli$ Ivy within the recently published genome of Pseudomonas aeruginosa (1). However, the two protein sequences only share 30% of identical residues in their most similar region, indicating a fast divergence rate (Fig. 7). It is thus likely that the genes of the $ykfE$/Ivy family are not detected in other bacteria because of their low sequence conservation. It is our hope that the knowledge of the three-dimensional structure of Ivy will allow the discovery of other inhibition mechanisms. We thank Dr. C. Cambillau for access to x-ray diffraction equipment and helpful discussions and Dr. Mary Berlin for quick validation of the Ivy name. We also acknowledge the helpful comments of anonymous referees concerning the analysis of the Ivy inhibition mechanism. We thank Dr. C. Evrard for the gift of the known signal peptide in Ivy.

Acknowledgments—We thank Dr. C. Cambillau for access to x-ray diffraction equipment and helpful discussions and Dr. Mary Berlin for quick validation of the Ivy name. We also acknowledge the helpful comments of anonymous referees concerning the analysis of the Ivy inhibition mechanism. We thank Dr. C. Evrard for the gift of the known signal peptide in Ivy.

Fig. 5. Comparative inhibition of HEWL (black circles) and λ phage lysozyme (open circles) activity by increasing concentrations of $ykfE$/Ivy.

Fig. 6. Inhibition of lysozyme activity of human saliva by increasing concentrations of $ykfE$/Ivy.

and 9.0. The order of magnitude of the dimerization $K_d$ appeared much lower than $10^{-8} \text{ M}$, although no precise measurement could be made in this concentration range. Altogether, these biophysical results suggest that the Ivy homodimer is the physiologically active unit.

C-type lysozyme is an ancient protein whose origin goes back about 500 million years (16). It has long been recognized that bacteria might have evolved a resistance mechanism against an ubiquitous bactericidal enzyme should be advantageous to all murein-containing bacteria, in particular Gram-positive bacteria, $ykfE$/Ivy-like genes are expected to exist in many bacterial genomes, making its ORFan nature a paradox. Indeed, we detected a new putative ortholog of $E. coli$ Ivy within the recently published genome of Pseudomonas aeruginosa (1). However, the two protein sequences only share 30% of identical residues in their most similar region, indicating a fast divergence rate (Fig. 7). It is thus likely that the genes of the $ykfE$/Ivy family are not detected in other bacteria because of their low sequence conservation. It is our hope that the knowledge of the three-dimensional structure of Ivy will allow the discovery of other Ivy homologues by the identification of a set of critical positions in the sequence beyond the twilight zone of sequence similarity.

REFERENCES

1. Stover, C. K., Pham, X. Q., Erwin, A. L., Minogue, S. D., Warrenner, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalk, D. J., Lagroux, M., Garber, R. L., Golty, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lam, R., Smith, R., Spencer, D., Wong, K. G., Wu, Z., and Paulsen, I. T. (2000) Nature 406, 959–964
2. Fischer, D., and Eisenberg, D. (1999) Bioinformatics 15, 759–762
3. Alimi, J. P., Poirot, O., Lopes, F., and Claverie, J. M. (2000) Genome Res. 10, 959–966
4. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasser, J. D., Rode, C. M., Mayhew, G. F., et al. (1997) Science 277, 1453–1474
5. Wasinger, V. C., and Humphery-Smith, I. (1998) FEMS Microbiol. Lett. 169, 375–382
6. Pasquali, C., Frutiger, S., Wilkins, M. R., Hughes, G. J., Appel, R. D., Bairsch,
A., Schaller, D., Sanchez, J. C., and Hochstrasser, D. F. (1996) *Electrophoresis* 17, 547–555
7. Abergel, C., Monchois, V., Chenivesse, S., Jeudy, S., and Claverie, J. M. (2000) *Acta Crystallogr.* 56, 1694–1695
8. Gill, S. C., and von Hippel P. H. (1989) *Anal. Biochem.* 182, 319–326
9. Charlemagne, D., and Jolles, P. (1976) *C. R. Acad. Sci. (Paris)* 272, 2721–2723
10. Morrison, J. F. (1969) *Biochim. Biophys. Acta* 185, 269–286
11. Morrison, J. F., and Walsh, C. (1988) *Adv. Enzymol. Relat. Areas Mol. Biol.* 61, 201–301
12. Soumillion, P., Jespers, L., Vervoors, J., and Fastrez, J. (1995) *Protein Eng.* 8, 451–456
13. Hackman, R. H., and Golberg, M. (1964) *Anal. Biochem.* 8, 397–401
14. Wang, Y., Addess, V., Bocquet, C., and Verny, C. (1989) *Anal. Biochem.* 182, 319–326
15. Charlemagne, D., and Jolles, P. (1970) *C. R. Acad. Sci.* (Paris) 272, 1–27
16. Morrison, J. F. (1969) *Biochim. Biophys. Acta* 185, 269–286
17. Tachibana, M., Morioka, H., Machino, M., and Mizukoshi, O. (1986) *Auris Nasus Larynx* 13, 97–99
18. Hanamure, Y., and Lim, D. J. (1986) *Am. J. Otolaryngol.* 7, 410–425
19. McCormick, K. A. (1997) *Surv. Ophthalmol.* 42, 233–246
20. Schenkels, L. C., Veerman, E. C., and Nieuw Amerongen, A. V. (1995) *Crit. Rev. Oral Biol. Med.* 6, 161–175
21. Cole, A. M., Dewan, P., and Ganz, T. (1999) *Infect. Immun.* 67, 3267–3275
22. Kondo, I. R., Hanna, L., and Keshishyan, H. (1975) *Proc. Natl. Acad. Sci. U. S. A.* 97, 8856–8861
23. Walker, W. A. (1979) *Ciba Found. Symp.* 70, 201–219
24. Ved'mina, E. A., Pasternak, N. A., Shenderovich, V. A., Zhuravleva, T. P., and Andrusenko, I. T. (1979) *Antibiot. Khimioter.* 24, 746–750
25. Hancock, R. E., and Scott, M. G. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 8856–8861
26. Vollmer, W., and Tomas, A. (2000) *J. Biol. Chem.* 275, 20496–20501
27. Bukharin, O. V. (1994) *Zh. Mikrobiol. Epidemiol. Immunobiol.* 1, (Suppl.) 1–5
28. Bondarenko, V. M., Petrovskaya, V. G., and Iablokov, A. L. (1994) *Zh. Mikrobiol. Epidemiol. Immunobiol.* 1, (Suppl.) 27–28
29. Nemtseva, N. V. (1997) *Zh. Mikrobiol. Epidemiol. Immunobiol.* 1, (Suppl.) 22–25
30. Lentsner, A. A., Lentsner, Kh. P., and Toom, M. A. (1975) *Zh. Mikrobiol. Epidemiol. Immunobiol.* 8, 77–81
31. Bukharin, O. V., and Deriabin, A. A. (1989) *Zh. Mikrobiol. Epidemiol. Immunobiol.* 11, 16–19
32. Pearson, W. R. (2000) *Methods Mol. Biol.* 132, 185–219

E. coli ORFan Encodes a Potent Inhibitor of C-type Lysozyme 18441