are on average largely unperturbed. We conclude that the observed changes in relaxation rates result primarily from changes in the dynamic rather than structural properties of the various clusters.

In the native structure, Trp^{62} is highly exposed to solvent; the crystal structure shows that its side chain is substantially disordered (*27*), and NMR measurements in solution reveal dynamic behavior (*35*). By contrast, in the denatured states (*36*), and particularly during the early stages of folding (*37*), NMR experiments indicate that this residue, like the other tryptophan residues, is largely inaccessible to solvent. Together with data from the hydrogen-exchange protection experiments, we conclude that in these nonnative states the β-domain residues Trp^{62} and Trp^{63} associate with a nativelike hydrophobic cluster in the α domain involving Trp^{108} and Trp^{111} that is itself strongly linked to the other regions of nonrandom structure. Thus nonnative interactions stabilize a nativelike core. Presumably, the replacement of the large hydrophobic tryptophan residue at position 62 by glycine results in the destabilization of this core. The resulting increase in dynamic flexibility could reflect an increased population of more extended structures in the ensemble of interconverting conformers. A polypeptide chain in such structures will undergo conformational averaging much more rapidly than in compact denatured states, where significant energetic barriers are known to exist.

Our results suggest that, within the ensemble of conformations representing the denatured states of lysozyme, there are long-range interactions that link clusters of residues that are not close together in sequence. These results are consistent with the hypothesis that steps that involve the reorganization of the species formed initially during the refolding of lysozyme are likely to be key determinants of the kinetics of the folding process (*29*, *30*). Although the folding of small proteins is dominated by the search for nativelike contacts, in the case of larger proteins, including those with multiple domains, species with at least some nonnative interactions can be important determinants of the folding process (*8*). Such interactions appear to be located primarily at the interface between the two structural domains, the region associated with the slowest step in the folding of lysozyme (*4*).

The avoidance of misfolding and potential aggregation of nonnative species is a key aspect of the folding and long-term stability of proteins. For example, single point mutations in human lysozyme are responsible for the occurrence of systemic disease in which large quantities of amyloid fibrils are deposited in a variety of internal organs (*38*). That a single amino acid replacement can perturb the nonnative state of the protein is of particular interest, as the aggregation of partially or completely unfolded species is the essential step in the development of the amyloid structures. Thus, although a residue such as tryptophan may be exposed in the native state for functional reasons, it could be buried in the early stages of folding to reduce the tendency of these transiently populated species to aggregate. Such a conclusion leads to the possibility that the sequence of a protein codes for structural characteristics other than those of the native fold.

References and Notes

1. P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **293**, 321 (1999).
2. S. E. Radford, C. M. Dobson, *Cell* **97**, 291 (1999).
3. P. Romero *et al.*, *Pac. Symp. Biocomput.* **3** (1998).
4. C. M. Dobson, *Philos. Trans. R. Soc. London B. Biol. Sci.* **356**, 133 (2001).
5. B. A. Shoemaker, P. G. Wolynes, *J. Mol. Biol.* **287**, 657 (1999).
6. B. A. Shoemaker, J. Wang, P. G. Wolynes, *J. Mol. Biol.* **287**, 675 (1999).
7. W. F. van Gunsteren, R. Burgi, C. Peter, X. Daura, *Angew. Chem. Int. Ed. Engl.* **40**, 351 (2001).
8. C. M. Dobson, A. Sali, M. Karplus, *Angew. Chem. Int. Ed. Engl.* **37**, 868 (1998).
9. D. Shortle, M. S. Ackerman, *Science* **293**, 487 (2001).
10. H. J. Dyson, P. E. Wright, *Nature Struct. Biol.* **5** (suppl.), 499 (1998).
11. Y. K. Mok, C. M. Kay, L. E. Kay, J. Forman-Kay, *J. Mol. Biol.* **289**, 619 (1999).
12. F. J. Blanco, L. Serrano, J. D. Forman-Kay, *J. Mol. Biol.* **284**, 1153 (1998).
13. D. Shortle, C. Abeygunawardana, *Structure* **1**, 121 (1993).
14. K. M. Fiebig, H. Schwalbe, M. Buck, L. J. Smith, C. M. Dobson, *J. Phys. Chem.* **100**, 2661 (1996).
15. L. J. Smith *et al.*, *J. Mol. Biol.* **255**, 494 (1996).
16. H. Schwalbe *et al.*, *Biochemistry* **36**, 8977 (1997).
17. K. B. Wong, S. M. Freund, A. R. Fersht, *J. Mol. Biol.* **259**, 805 (1996).
18. D. Neri, M. Billeter, G. Wider, K. Wuthrich, *Science* **257**, 1559 (1992).
19. G. Saab-Rincon, P. J. Gualfetti, C. R. Matthews, *Biochemistry* **35**, 1988 (1996).
20. I. J. Ropson, C. Frieden, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7222 (1992).
21. M. E. Hodsdon, C. Frieden, *Biochemistry* **40**, 732 (2001).
22. G. Merutka, H. J. Dyson, P. E. Wright, *J. Biomol. NMR* **5**, 14 (1995).
23. D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **5**, 67 (1995).
24. Supplementary material is available on *Science* Online at www.sciencemag.org/cgi/content/full/295/5560/1719/DC1.
25. H. A. Sober, R. A. Harte, *Handbook of Biochemistry; Selected Data for Molecular Biology* (Chemical Rubber Co., Cleveland, ed. 2, 1970).
26. D. R. Shortle, *Curr. Opin. Struct. Biol.* **6**, 24 (1996).
27. C. C. Blake *et al.*, *Nature* **206**, 757 (1965).
28. H. Schwalbe *et al.*, *Protein Sci.* **10**, 677 (2001).
29. A. Miranker, C. V. Robinson, S. E. Radford, R. T. Aplin, C. M. Dobson, *Science* **262**, 896 (1993).
30. S. E. Radford, C. M. Dobson, P. A. Evans, *Nature* **358**, 302 (1992).
31. D. M. Rothwarf, H. A. Scheraga, *Biochemistry* **35**, 13797 (1996).
32. T. Ueda, H. Yamada, H. Aoki, T. Imoto, *J. Biochem. (Tokyo)* **108**, 886 (1990).
33. T. Ueda, T. Ohkuri, T. Imoto, *Biochem. Biophys. Res. Commun.* **228**, 203 (1996).
34. T. Ueda *et al.*, *J. Mol. Biol.* **235**, 1312 (1994).
35. M. Buck *et al.*, *Biochemistry* **34**, 4041 (1995).
36. C. Lyon, J. Jones, C. Redfield, C. Dobson, P. Hore, *J. Am. Chem. Soc.* **121**, 6505 (1999).
37. P. Hore, S. Winder, C. Roberts, C. M. Dobson, *J. Am. Chem. Soc.* **119**, 5049 (1997).
38. D. R. Booth *et al.*, *Nature* **385**, 787 (1997).
39. M. C. Vaney, S. Maignan, M. RiesKautt, A. Ducruix, *Acta Crystallogr. D Biol. Crystallogr.* **52** (1996).

40. We thank the Massachusetts Institute of Technology (for start-up funds and Rosenblith stipend), the Karl Winnacker-Foundation, the Fonds der Chemischen Industrie, the Bundesministerium für Bildung und Forschung (BMBF), and the Alfred P. Sloan Foundation for support. We dedicate this report to Gerhard Quinkert on the occasion of his 75th birthday.

2 November 2001; accepted 22 January 2002

A Functional Screen for the Type III (Hrp) Secretome of the Plant Pathogen *Pseudomonas syringae*

David S. Guttman,[1]*† Boris A. Vinatzer,[2]*† Sara F. Sarkar,[1] Max V. Ranall,[2] Gregory Kettler,[2] Jean T. Greenberg[2]

Type III secreted "effector" proteins of bacterial pathogens play central roles in virulence, yet are notoriously difficult to identify. We used an in vivo genetic screen to identify 13 effectors secreted by the type III apparatus (called Hrp, for "hypersensitive response and pathogenicity") of the plant pathogen *Pseudomonas syringae*. Although sharing little overall homology, the amino-terminal regions of these effectors had strikingly similar amino acid compositions. This feature facilitated the bioinformatic prediction of 38 *P. syringae* effectors, including 15 previously unknown proteins. The secretion of two of these putative effectors was shown to be type III–dependent. Effectors showed high interstrain variation, supporting a role for some effectors in adaptation to different hosts.

The bacterial type III secretion system is responsible for some of the most devastating diseases of animals and plants. This remarkable system enables a bacterium to strategically inject proteins directly into the host cytoplasm or its extracellular milieu, and thereby subvert host cellular processes (*1*–*3*). The type III apparatus is required for patho-

genesis and is highly conserved across a broad range of Gram-negative bacterial pathogens. Less conserved is the repertoire of proteins exported (the type III secretome) (*1–3*). Little is known about the function or mechanism of action of phytopathogenic effectors, although several are known to enhance the growth rate and transmission potential of the pathogen (*4, 5*). Avirulence (*avr*) genes are a class of phytopathogenic effectors that restrict host range (*3*). Most Avr proteins are thought to be secreted through the type III apparatus. Plants with the appropriate cognate resistance (*R*) genes recognize Avr proteins and mount a defense response characterized by a type of programmed cell death (PCD) called the hypersensitive response [HR (*6*)].

Variation in the constellation of effector genes among related pathogenic strains may facilitate adaptation to new hosts and permit the rapid evolution of novel pathogen specificities (*4, 7*). To critically test this hypothesis, a comprehensive analysis of the identity and function of effector genes is necessary. To achieve this end, we devised an in vivo screen exploiting the modular nature of effectors (*5, 8–11*) and the well-characterized intracellular interaction, between the COOH-terminus of the AvrRpt2 effector and its cognate R protein, Rps2 (*12*). Because the COOH-terminal HR-inducing domain of AvrRpt2 (lacking the NH_2-terminal secretion signal) is a good in vivo reporter for type III secretion (*5, 13*), we were able to devise a transposon containing the DNA coding for AvrRpt2$^{81-255}$ to capture type III secretion signals from unknown effector genes, also known as *hop* genes [*h*rp/hrc *o*uter *p*rotein (*14*)] (Web fig. 1) (*15*). Insertions of the *avrRpt2$^{81-255}$* transposon that create translational fusions with *hop* genes generate *P. syringae* strains that induce the HR in an Rps2-dependent manner upon infection of the model host plant, *Arabidopsis thaliana*. The screen relied on the type III secretion signal and the endogenous promoter of the *hop* gene and was thus highly specific. During infection at least one type III–utilizing pathogen delivers one set of effectors to the extracellular environment, and another set to the host cell interior (*16*). The precise mechanism governing the final location of an effector is largely unknown. On the basis of the data reported here, the final destination of the Hop::AvrRpt2$^{81-255}$ fusions, inside plant cells, appears to be driven largely by the AvrRpt2$^{81-255}$ moiety, although a specific translocation signal for AvrRpt2 is not known. This enabled the capture of type III delivery signals from a broad variety of effectors that are likely to have a number of different sites of action.

We chose to study effectors from *P. syringae* pv. *maculicola* strain ES4326 (*Pma*ES4326) because the *A. thaliana* host responses elicited by this strain are very well characterized (*17–19*). Approximately 75,000 independent transposon insertion strains were screened (*20*). Twenty-five independent HR-inducing isolates with fusions to AvrRpt2$^{81-255}$ identified 13 different *hop* genes (Table 1) that, aside from *hopPmaA*, were chromosomally located (*21*).

Seven Hop proteins had widely varying degrees of similarity to known phytopathogenic effectors or *P. syringae* virulence factors (Table 1). These included a putative ortholog and an apparent paralog (HopPmaB) of AvrPphE (*22, 23*), the latter having a eukaryotic NH_2-myristoylation motif at amino acid 2 that may facilitate targeting to host

[1]Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada. [2]Department of Molecular Genetics and Cell Biology, University of Chicago, 1103 East 57th Street, Chicago, IL 60637, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: guttman@botany.utoronto.ca

Table 1. Characterization of effectors identified by in vivo screening in *Pma*ES4326 and functional genomic analysis of *Pto*DC3000. All *hop* genes were sequenced to completion, with the exception of *avrE$_{Pma}$*. For this effector all characterizations were based on the *Pto*DC3000 homolog (AAF1499) except for the Blast analysis which used the incomplete (inc) sequence (40% missing). BLASTP and PSI-BLAST queries were done using the nr database at the National Center for Biotechnology Information (*47*). HopPtoO and HopPtoP were identified in the unfinished genome sequence of *Pto*DC3000 and confirmed to be secreted by creating fusions with AvrRpt2$^{81-255}$. *holPmaN* was amplified by PCR from a cosmid with primers designed on *hopPmaL*. Additional information is presented in Web table 1 and Web note 4 (15).

Effectors	N*	Acc†	Homology (accession number)	BLASTP (bits/E value)	Size	%GC‡	Int site§ (aa)	Linkage‖	%D 50aa	Loc¶
HopPmaA	7	AF458040	*E. coli* Hypothetical Protein (BAA36749)	118/3e-25	774	54.45	284		18	C
HopPmaB	1	AF458041	*Pph#* AvrPphE (AAA67930)	47.8/4e-04	347	60.16	328	EEL	24	C
			Pma AvrPphE$_{Pma}$ (AF458042)	59.7/4e-08						
AvrPphE$_{Pma}$	1	AF458042	*Pph#* AvrPphE (AA67930)	375/1e-103	362	58.02	307	EEL	16	C
HopPmaD	2	AF458043	*Ppi*** AvrPpiG1 (CAC16700)	59.4/7e-08	260	58.98	131	*hopPtoP$_{Pma}$*‖‖	8	M
			Xcv†† AvrBst (AAD39255)	44.8/0.002						
HrpW$_{Pma}$	1	AF458044	*Pto*‡‡ HrpW (AAF71503)	367/1e-100	424	56.47	137	CEL	16	–
AvrE$_{Pma}$	1	AF458405	*Pto*‡‡ AvrE (AAF71499)	712/0.0	1795	58.06	inc	CEL	16	C
HopPmaG	1	AF458045	*E. coli* Lytic Transglycosylase, 35k, Precursor (S65868)	97.8/2e-19	413	54.75	14	*hopPmaL*	16	S
HopPmaH	1	AF458046	*Bacillus subtilis* Pectate Lysase PelB (E69674)	214/3e-54	540	56.99	226		16	–
HopPmaI	4	AF458047	*Legionella pneumophila* Chaperone Protein DnaJ (P50025)	67.4/2e-10	432	60.67	315		14	C
HopPmaJ	2	AF458048	no homology	–	149	59.55	48		18	C
HopPmaK§§	1	AF458049	no homology	–	308	49.30	56		24	C
HopPmaL	2	AF458050	*Pph#* VirPphA (AAD47203)	135/7e-31	385	58.20	250	*hopPmaG*	20	C
HopPtoA1$_{Pma}$	1	AF458051	*Pto*‡‡ HopPtoA1 (AAF71504)	550/1e-155	484	60.34	291	CEL	24	C
HolPmaN		AF458391	*Pph#* VirPphA (AAD47203)	43.5/6e-04	155	63.25			18	C
			Pma HopPmaL (AF45805)	133/3e-31						
HopPtoO		AF458392	*Gallus gallus* NAD(P)(+)—Arginine ADP-Ribosyltransferase 2 Precursor (P55807)	51.2/1e-05	283	46.50	200		24	C
HopPtoP		AF458393	*B. subtilis* phage SPBc2 Putative Transglycosylase (NP046584)	53.9/2e-06	324	59.60	111	*hopPmaD*‖‖‖	20	C
			HrpW (AAF71503)	47.4/3e-04						

*Number recovered. †Accession number. ‡The mean %GC for *Pto*DC3000 = 58.56%. §Number of amino acids from the *hop* start codon to the integration site of the *avrRpt2$^{81-255}$* containing transposon. In the case of multiple recoveries, only the smallest fusion presented. ‖EEL, Exchangeable Effector Locus of the hrp cluster; CEL, Conserved Effector Locus of the hrp cluster (*44*). ¶Based upon the subcellular localization program TargetP (www.cbs.dtu.dk/services/TargetP-1.0) (*38*). Chloroplast (C), mitochondrial (M), secretory pathway (S), no localization (–). #*P. syringae* pv *phaseolicola*. **P. syringae* pv *pisi*. †† *Xanthomonas campestris* pv *vesicatoria*. ‡‡*P. syringae* pv *tomato*. §§*hopPmaK* resides in an operon downstream of *chpPmaK (chaperone of hrp outer protein PmaK)*, a gene encoding a homolog of the DspB chaperone (*49*). ‖‖‖*hopPmaD* is linked to *hopPtoP$_{Pma}$*.

cell membranes. HopPmaD was a member of the *Xanthomonas campestris* AvrBsT and *Yersinia pestis* YopJ effector family present in many plant and animal pathogens. Members of this family induce PCD in both plant and animal hosts. Although the similarity between HopPmaD and AvrBsT is weak, all the amino acids implicated in inducing PCD were conserved (*24*). The first 20 amino acids of HopPmaD were very similar to those found in the AvrPto effector (*25*), suggesting recombination may generate new *hop* genes. HopPmaL was similar to VirPphA, a *P. syringae* virulence protein not previously known to be secreted (*26*). HolPmaN (Hop-like) was a truncated, presumably nonfunctional protein similar to HopPmaL that was found through polymerase chain reaction (PCR) analysis of *Pma*ES4326. Orthologs of the harpin HrpW effector and the virulence protein AvrE were also recovered.

Five previously unknown Hops (including two identified using bioinformatics; see below) contained regions most similar to proteins found in animal pathogens or nonpathogenic bacteria. HopPmaA had similarity to a VT1- or VT2-Sakai prophage protein of unknown function in pathogenic *E. coli* (*27, 28*). HopPmaG and HopPtoP were similar to different transglycosylases. These enzymes act on the bacterial peptidylglycan layer; one such enzyme is required for flagellar formation (an assembly requiring a type III–like apparatus) in *Salmonella enterica* (*29*). HopPmaG and HopPtoP may similarly facilitate assembly of the type III apparatus. HopPmaH showed high similarity to a plant cell wall–degrading pectate lyase, an enzyme important for *P. marginalis* virulence (*30*). HopPmaI had COOH-terminal similarity to a DnaJ domain; such domains interact with the molecular chaperone Hsp70 and alter its substrate binding (*31*). HopPmaI also had an ATP- or GTP-binding site motif A (P-loop) at amino acid 118, and three full and one partial proline-rich tandemly repeated sequences (*32*). These latter repeats are present nearly five times in the *Pto*DC3000 homolog (*32*). The variable numbers of repeats could be important for adaptation to different hosts, similar to some *Xanthomonas campestris* effectors with repeated segments (*33*). The *hopPmaI*$^{1\text{-}315}$*::avrRpt2*$^{81\text{-}255}$ insertion strain had reduced virulence (Fig. 1, A and B). Finally, HopPtoO had a myristoylation motif at amino acid 2 and similarity to a region of an NAD(P)(+)-arginine ADP-ribosyltransferase 2 precursor that was also present in the *P. aeruginosa* effector ExoS (*34*).

All *hop::avrRpt2*$^{81\text{-}255}$ strains showed *RPS2*-dependent HRs (Web table 1), indicating that the Hop-AvrRpt2$^{81\text{-}255}$ fusions were not generally cytotoxic. Type III–dependent delivery was confirmed for each of the Hop::AvrRpt2$^{81\text{-}255}$ fusion proteins (Fig. 1, C and D and Web table 1). Additionally, all but one *hop* gene (*hopPmaJ*) had upstream sequences called "hrp boxes" found in promoter regions of many effector genes (*35*). Finally, the synthesis of most of the Hop::AvrRpt2$^{81\text{-}255}$ fusion proteins was undetectable in rich medi-

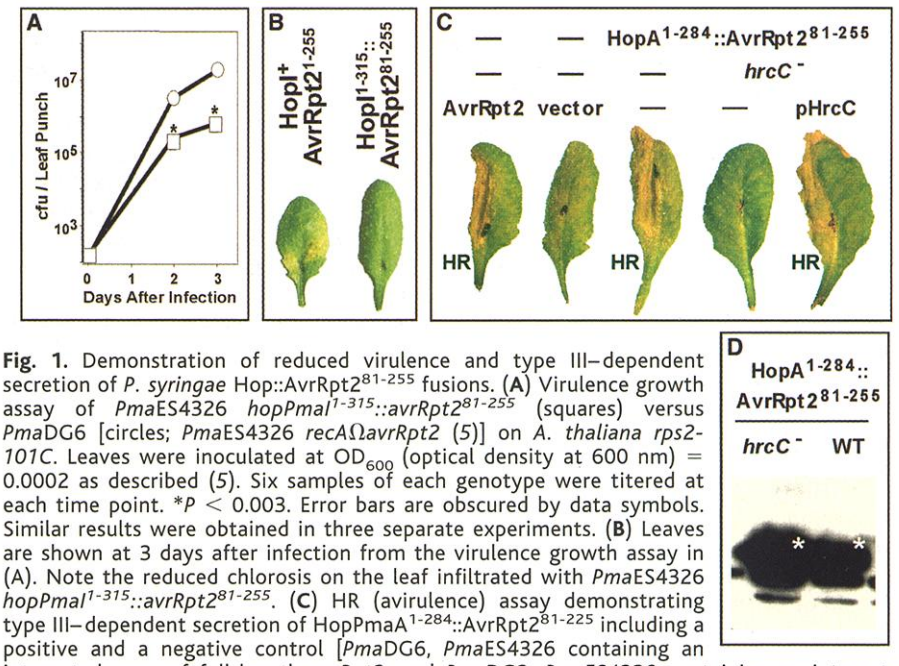

Fig. 1. Demonstration of reduced virulence and type III–dependent secretion of *P. syringae* Hop::AvrRpt2$^{81\text{-}255}$ fusions. (**A**) Virulence growth assay of *Pma*ES4326 *hopPmaI*$^{1\text{-}315}$*::avrRpt2*$^{81\text{-}255}$ (squares) versus *Pma*DG6 [circles; *Pma*ES4326 *recAΩavrRpt2* (*5*)] on *A. thaliana rps2-101C*. Leaves were inoculated at OD_{600} (optical density at 600 nm) = 0.0002 as described (*5*). Six samples of each genotype were titered at each time point. *$P < 0.003$. Error bars are obscured by data symbols. Similar results were obtained in three separate experiments. (**B**) Leaves are shown at 3 days after infection from the virulence growth assay in (A). Note the reduced chlorosis on the leaf infiltrated with *Pma*ES4326 *hopPmaI*$^{1\text{-}315}$*::avrRpt2*$^{81\text{-}255}$. (**C**) HR (avirulence) assay demonstrating type III–dependent secretion of HopPmaA$^{1\text{-}284}$::AvrRpt2$^{81\text{-}225}$ including a positive and a negative control [*Pma*DG6, *Pma*ES4326 containing an integrated copy of full-length *avrRpt2*, and *Pma*DG3, *Pma*ES4326 containing an integrated copy of a vector control, respectively (*5*)]. pHrcC is a *Pma*ES4326 cosmid clone (3B3) containing the *hrcC* gene (*20*). All other Hop::AvrRpt2$^{81\text{-}255}$ fusions behaved similarly (Web table 1). (**D**) Western blot of total protein extracted from a *Pma*ES4326 (WT) strain and a *Pma*ES4326 *hrcC*$^-$ strain both containing an integrated copy of *hopPmaA*$^{1\text{-}284}$*::avrRpt2*$^{81\text{-}255}$ (band corresponding to the fusion protein indicated by an asterisk) in order to verify that the HopPmaA$^{1\text{-}284}$::AvrRpt2$^{81\text{-}225}$ fusion protein is expressed to similar levels in the HR$^+$ and HR$^-$ strains (*48*). A background band at 90 kD served as a loading control (Web fig. 2).

Table 2. Analysis of NH_2-terminal amino acid composition of type III effectors. The Hop group (Hop) includes the proteins shown to be secreted (Table 1) as well as the known type III–secreted proteins: HopPtoB (ORF1 EEL) of *Pto*DC3000 (AAF71498), AvrRpt2 (CAA79815), AvrRpm1 (NP_114197), HopPsyA (AAF71481), AvrB (AAA25726), HrpZ (AAB00127), AvrPto (AAA25728), HrpA (AAB00126). The Flagella group (Fla) includes *Pto*DC3000 homologs (identified in the unfinished genome sequence at www.tigr.com) of flagellar type III–secreted proteins from *S. enterica* (n = 12). The control group (Con) includes a group of randomly selected proteins from *P. syringae* (n = 23). The *Salmonella* effector group (SalEff) includes *S. enterica* effectors (n = 24). The *Salmonella* control (SalCon) group includes a group of randomly selected proteins from *S. enterica* (n = 32). All of these loci are presented in Web table 4. *P* values (*P*) were determined using the nonparametric Kolmogorov Smirnov test. *P* values in bold indicate significance (≤0.05).

Protein (amino acids analyzed)	Mean % amino acid (standard error)				
	Ser	Asp	Leu	Lys	Asn
Hop(1–50)	17.13 (1.01)	3.58 (0.45)	6.17 (0.83)	2.52 (0.49)	6.00 (0.81)
Hop(51-end)	8.20 (0.36)	6.62 (0.32)	7.96 (0.50)	5.48 (0.64)	5.15 (0.35)
P	**<0.0001**	**0.0001**	**0.0104**	**0.0104**	0.240
Fla(1–50)	13.17 (1.24)	4.33 (0.85)	7.50 (1.13)	4.83 (0.83)	8.33 (1.01)
Fla(51-end)	10.26 (0.76)	5.75 (0.49)	7.64 (0.67)	3.69 (0.54)	5.83 (0.66)
P	0.100	0.249	0.249	0.100	**0.034**
Con(1–50)	5.57 (0.64)	5.74 (0.79)	12.17 (1.07)	3.65 (0.69)	3.04 (0.56)
Con(51-end)	6.21 (0.3)	4.96 (0.28)	10.69 (0.60)	3.94 (0.32)	3.19 (0.25)
P	**0.026**	0.591	0.428	**0.026**	0.059
SalEff(1-50)	10.87 (1.10)	2.96 (0.49)	8.35 (0.87)	6.87 (1.02)	4.78 (0.71)
SalEff(51-end)	7.13 (0.51)	4.78 (0.31)	10.62 (0.74)	5.86 (0.53)	5.17 (0.38)
P	**0.0013**	**0.0013**	0.0591	0.0591	0.1208
SalCon(1-50)	5.00 (0.40)	5.19 (0.62)	10.94 (0.72)	4.94 (0.64)	3.25 (0.52)
SalCon(51-end)	5.28 (0.27)	5.14 (0.32)	10.69 (0.63)	4.09 (0.38)	3.33 (0.28)
P	0.088	0.159	0.433	0.433	**0.022**

Table 3. *hop* genes distribution in selected *Pseudomonas syringae* strains and other Gram-negative pathogens. Genomic DNA extraction, Eco RI digest and Southern blot followed standard protocols (*45*). Probes were generated by PCR and corresponded to full-length *hop* genes (except for *hopPmaI*, which was missing the region homologous to *dnaJ*) from Table 1. *Pma*ES4326, *Pma*M5, *Pto*DC3000, *Pma*M3, *Pma*M1, *Pma*M2, *Pma*M6, *Pto*5034, *Pc*83-1, *Xcc*BP109, and *Xcc*750 infect *A. thaliana*. *Pto*DC3000 and *Pto*5034 also infect tomato. *Pph*3121, *Psy*61, and *Psy*B782 infect bean.

Strain	Effector														
	HrpW	HopJ	HopP	HopI	AvrE	HopPtoA1	HopH	HopO	HopL	AvrPphE	HopG	HopA	HopK	HopD	HopB
*Pma*ES4326*	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
*Pma*M5	x	x	x	x	x	x	x	x	x	x	x	x	x	–	–
Pph NPS3121†	x	x	x	x	x	x	x	x	x	x	x	x	–	–	–
Pto DC3000‡	x	x	x	x	x	x	x	x	x	x	x	–	–	–	–
Pma M3	x	x	x	x	x	x	x	x	x	x	x	–	–	–	–
Pma M1	x	x	x	x	x	x	x	x	x	x	–	–	–	–	–
Pma M2	x	x	x	x	x	x	x	x	x	x	–	–	–	–	–
Pma M6	x	x	x	x	x	x	x	x	x	x	–	–	–	–	–
Pto 5034	x	x	x	x	x	x	x	x	x	x	–	–	–	–	–
Ps Cit7	x	x	x	x	x	x	x	–	x	–	–	–	–	–	–
Psy B728a§	x	x	x	–	x	x	x	x	x	–	–	–	–	–	–
Psy 61	x	x	x	x	x	x	x	x	–	–	–	–	–	–	–
P. viridiflava PV5	x	–	x	–	x	x	x	x	x	–	–	–	–	–	–
P. chicorii 83-1	x	x	x	x	–	–	–	–	–	–	–	–	–	–	–
P. fluorescence	x	x	x	–	–	–	–	–	–	–	–	–	–	–	–
P. putida PPS75	x	x	–	x	–	–	–	–	–	–	–	–	–	–	–
P. stutzeri	x	x	x	–	–	–	–	–	–	–	–	–	–	–	–
P. aeruginosa PA01	x	x	–	–	–	–	–	–	–	–	–	–	–	–	–
P. aeruginosa PA11	x	x	–	–	–	–	–	–	–	–	–	–	–	–	–
P. aeruginosa PA12	x	x	–	–	–	–	–	–	–	–	–	–	–	–	–
R. solanacearum‖	x	x	–	–	–	–	–	–	–	–	–	–	–	–	–
Xcc 750¶	x	x	–	–	–	–	–	–	–	–	–	–	–	–	–
Xcc BP109	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B. cepacia#	x	–	–	–	–	–	–	–	–	–	–	–	–	–	–

**P. syringae* pv. *maculicola* †*P. syringae* pv. *phaseolicola* ‡*P. syringae* pv. *tomato* §*P. syringae* pv. *syringae* ‖*Ralstonia solanacearum* ¶*Xanthomonas campestris* pv. *campestris* #*Burkholderia cepacia*

um and induced by minimal medium, similar to what has been found for other *P. syringae* effectors (Web fig. 2 and Web table 1) (*36, 37*). The only exceptions were HopPmaA$^{1\text{-}284}$::AvrRpt2$^{81\text{-}255}$ which was synthesized in rich medium and was further induced in minimal medium, and HopPmaJ$^{1\text{-}48}$::AvrRpt2$^{81\text{-}255}$ which was produced constitutively.

Pseudomonas syringae effectors have exceptionally high Ser content in the NH_2-terminal 50 amino acid secretion regions compared to the rest of the protein (Tables 1 and 2). A low Asp, Leu, and Lys content in the effectors' NH_2-termini was also observed. These NH_2-terminal amino acid biases resemble chloroplast and mitochondrial targeting sequences (*38*). Indeed, most *P. syringae* effectors are predicted to localize to chloroplasts (Table 1 and Web table 3). This similarity could reflect a common mechanism used by the type III apparatus and organelle import complexes to recognize secretion signals and to facilitate secretion. Alternatively, the similarities could either indicate a common evolutionary origin for secretion and organelle targeting mechanisms, signify an analogous functional requirement for these sequences for proper targeting of effectors to host organelles [one *E. coli* effector localizes to mitochondria (*39*)], or simply be coincidental. *Salmonella enterica* effectors also show NH_2-terminal Ser and Asp biases (Table 2), suggesting a conserved function for this feature. The overall amino acid composition of *P. syringae* effectors was also biased, showing a high overall Ser and Asn content and a low overall Leu, Ile, and Val content (Web table 2). Some features of this amino acid bias were shared with *P. syringae* homologs of flagellar secreted proteins from *S. enterica* and with *S. enterica* effectors (Web table 2). Some amino acid biases may be important for the unique requirement of effectors and flagellar components to be unfolded during the delivery process and to refold upon reaching their final destinations (*40*).

We estimated the total number of effectors in a single strain of *P. syringae* (*Pto*DC3000) based upon the characteristic NH_2-terminal amino acid biases (high Ser and low Asp) and the conserved hrp box promoter element (*35*) (Web table 3). Preliminary *Pto*DC3000 sequence data was obtained from The Institute for Genomic Research (TIGR; www.tigr.org). The proposed secretome includes 38 proteins, of which 15 are putative novel effectors. Two of these latter ORFs (HopPtoO and HopPtoP, Table 1) were tested and shown to be delivered to *A. thaliana* (Web table 1 and Web fig. 2). *Pseudomonas syringae* may contain additional effector genes induced in different environmental contexts and thus lack hrp boxes. Indeed, *P. syringae* may interact with hosts from at least two kingdoms, as it contains homologs of the *Photorhabdus luminescens* insectidical toxin complex genes *tcaA-C*, *tcbA*, *tccA-C*, and *tcdA* (www.tigr.org) (*41*).

A core set of 10 of the 15 confirmed effectors was present in many commonly studied *P. syringae* strains causing diseases in different hosts (Table 3). Some of these effectors are also present in related Pseudomonads and other plant pathogens. A third of the effectors studied showed a high degree of variability, even among strains with similar host specificity. Several effectors also appear to be duplicated and some duplicated *hop* genes may have been inactivated by base changes, transposon insertions or genetic rearrangements (see *holPmaN* example above and the *avrPphD*$_{Pto}$, *hopPtoO2* and *holPtoU2* examples in Web table 3). *hop* gene inactivation or loss may contribute to pathogen fitness by allowing the pathogen to evade the host surveillance mechanism. In summary, *hop* genes represent a highly dynamic set of genes in *P. syringae* populations.

We have shown that the type III secretome of *P. syringae* is likely to contain more effectors than in any other pathogen characterized so far (*2*). The principle of our in vivo functional screen can be adapted to any type III–utilizing bacterial species. This approach, coupled with the ability to predict effectors based on their biased amino acid composition and mode of regulation affords the possibility of making rapid progress in understanding the mechanisms of action and the functions of effectors in diverse pathogens.

References and Notes

1. J. E. Galan, A. Collmer, *Science* **284**, 1322 (1999).
2. C. Hueck, *Microbiol. Mol. Biol. Rev.* **62**, 379 (1998).
3. G. R. Cornelis, F. Van Gijsegem, *Annu. Rev. Microbiol.* **54**, 735 (2000).
4. D. W. Gabriel, *Phys. Mol. Plant Pathol.* **55**, 205 (1999).
5. D. S. Guttman, J. T. Greenberg, *Mol. Plant-Microbe Interact.* **14**, 145 (2001).
6. J. T. Greenberg, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 525 (1997).
7. A. Collmer, *Curr. Opin. Plant Biol.* **1**, 329 (1998).
8. S. A. Lloyd, M. Norman, R. Rosqvist, H. Wolf-Watz, *Mol. Microbiol.* **39**, 520 (2001).
9. M. P. Sory, A. Boland, I. Lambermont, G. R. Cornelis, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11998 (1995).
10. M. P. Sory, G. R. Cornelis, *Mol. Microbiol.* **14**, 583 (1994).
11. M. B. Mudgett, B. J. Staskawicz, *Mol. Microbiol.* **32**, 927 (1999).
12. M. J. Axtell, T. W. McNellis, M. B. Mudgett, C. S. Hsu, B. J. Staskawicz, *Mol. Plant-Microbe Interact.* **14**, 181 (2001).
13. M. B. Mudgett *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13324 (2000).
14. J. R. Alfano, A. Collmer, *J. Bacteriol.* **179**, 5655 (1997).
15. Web note 1. Web tables and figures, as well as details of experimental procedures are available on *Science* Online at www.sciencemag.org/cgi/content/full/295/5560/1722/DC1.
16. V. T. Lee, O. Schneewind, *Mol. Microbiol.* **31**, 1619 (1999).
17. J. T. Greenberg, A. Guo, D. F. Klessig, F. M. Ausubel, *Cell* **77**, 551 (1994).
18. J. Glazebrook, E. E. Rogers, F. M. Ausubel, *Annu. Rev. Genet.* **31**, 547 (1997).
19. ———, *Genetics* **143**, 973 (1996).
20. Isolates were grown overnight to saturation, pooled into groups of eight, diluted by 1/10 and 1/20, and infiltrated into *A. thaliana* ecotype Columbia (*RPS2*$^+$). Each dilution was individually tested on half of each of two replicate leaves of a single plant. Positive and negative HR controls were performed on each day. Plants were scored for HR by at least two people. Pools that induced an HR were deconvoluted down to a single HR-inducing colony. Each positive isolate was retested at least four times to confirm its ability to induce the HR. The region of insertion was sequenced in those strains that were scored as positive by either inverse or anchor PCR (*42, 43*). The complete sequence of *hop* genes was obtained from a *Pma*ES4326 cosmid library kindly provided by F. M. Ausubel (Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA). This library was screened by PCR and colony filter hybridization for *hop*-containing clones following standard protocols. Most cosmids were positive for only one *hop* gene each. One cosmid contained *hopPmaB* as well as $avrPphE_{Pma}$. Sequence upstream of $avrPphE_{Pma}$ had homology with *hrpK*, indicating that this cosmid is derived from the EEL (*44*). One cosmid contained $hrpW_{Pma}$, $avrE_{Pma}$, and $hopPtoA1_{Pma}$. Two cosmids contained *hopPmaD* as well as $hopPtoP_{Pma}$. DNA was extracted from positive cosmids and subcloned into pBS SK+ (Stratagene). Subclones were screened by PCR and PCR products spanning *hop* genes were sequenced and assembled using Lasergene (DNASTAR). Cosmid DNA was also digested with Eco RI and Hind III, run on a 0.8% agarose gel in 1× TBE, alkaline transferred to a positively charged nylon membrane, and probed with ^{32}P-radiolabeled cosmid DNA following standard protocols (*45*). Conservation of restriction fragments between clones hybridizing to the same cosmid DNA was used to identify overlapping cosmids. A cosmid containing *hopPmaG* overlapped by four hybridizing fragments with a cosmid containing *hopPmaL*.
21. Web note 2 (*15*).
22. J. Mansfield, C. Jenner, R. Hockenhull, M. A. Bennett, R. Stewart, *Mol. Plant-Microbe Interact.* **7**, 726 (1994).
23. C. Stevens *et al.*, *Mol. Microbiol.* **29**, 165 (1998).
24. K. Orth *et al.*, *Science* **290**, 1594 (2000).
25. J. M. Salmeron, B. J. Staskawicz, *Mol. Gen. Genet.* **239**, 6 (1993).
26. R. W. Jackson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10875 (1999).
27. K. Makino *et al.*, *Genes Genet. Syst.* **74**, 227 (1999).
28. K. Yokoyama *et al.*, *Gene* **258**, 127 (2000).
29. T. Nambu, T. Minamino, R. M. Macnab, K. Kutsukake, *J. Bacteriol.* **181**, 1555 (1999).
30. C. H. Liao, D. E. McCallus, W. F. Fett, Y. G. Kang, *Can. J. Microbiol.* **43**, 425 (1997).
31. W. L. Kelley, *Trends Biochem. Sci.* **23**, 222 (1998).
32. Three full repeats of 37 or 38 amino acids (38 amino acids for the first repeat and 37 amino acids for the second and third) and a fourth partial repeat of 27 amino acids. The repeat sequence is RPGGAeQQAR-PETPPRsRPQtNsAPPPP-kAEPRPsSG (*46*), starting at amino acid 194. Capital letters represent identities, small letters represent similar amino acids (based on PAM250), and a dash indicates a one amino acid insertion of a P in the first repeat. Repeat regions were identified with the aid of Dotlet (www.isrec.isb-sib.ch/java/dotlet/Dotlet.html) using a 15–amino acid window. Partially completed *Pto*DC3000 genome sequence obtained from TIGR (www.tigr.com).
33. K. Herbers, J. Conrads-Strauch, U. Bonas, *Nature* **356**, 172 (1992).
34. A. M. Barbieri, Q. Sha, P. Bette-Bobillo, P. D. Stahl, M. Vidal, *Infect. Immun.* **69**, 5329 (2001).
35. R. W. Innes, A. F. Bent, B. N. Kunkel, S. R. Bisgrove, B. J. Staskawicz, *J. Bacteriol.* **175**, 4859 (1993).
36. K. van Dijk *et al.*, *J. Bacteriol.* **181**, 4790 (1999).
37. L. G. Rahme, M. N. Mindrinos, N. J. Panopoulos, *J. Bacteriol.* **174**, 3499 (1992).
38. O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, *J. Mol. Biol.* **300**, 1005 (2000).
39. B. Kenny, M. Jepson, *Cell. Microbiol.* **2**, 579 (2000).
40. C. E. Stebbins, J. E. Galan, *Nature* **414**, 77 (2001).
41. D. Bowen *et al.*, *Science* **280**, 2129 (1998).
42. H. Ochman, F. J. Ayala, D. L. Hartl, *Methods Enzymol.* **218**, 309 (1993).
43. D. S. Guttman, D. Charlesworth, *Nature* **393**, 263 (1998).
44. J. R. Alfano *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4856 (2000).
45. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1989).
46. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
47. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
48. Web note 3 (*15*).
49. S. Gaudriault, L. Malandrin, J. P. Paulin, M. A. Barny, *Mol. Microbiol.* **26**, 1057 (1997).

50. D.S.G. would like to acknowledge and thank A. Liu, J. Stavrinides, and L. Chu for their help with the screen. D.S.G. was supported by grants from NIH (NRSA GM020024) the Canadian National Science and Engineering Research Counsel (NSERC), the University of Toronto Connaught Fund, and the Canadian Foundation for Innovation. J.T.G. gratefully acknowledges funding by the Consortium for Biotechnology Research, Dow, Syngenta, the Pew Foundation, the University of Chicago Faculty Research Fund, and the University of Chicago's Division of Biological Sciences under the Research Resources Program for the Medical School of the Howard Hughes Medical Institute. Sequencing of *Pseudomonas syringae* pv *tomato* DC3000 was accomplished with support from the NSF. We thank A. Driks and L. Mets for helpful discussions.

7 November 2001; accepted 22 January 2002

Colorectal Cancer in Mice Genetically Deficient in the Mucin Muc2

Anna Velcich,[1]* WanCai Yang,[1] Joerg Heyer,[2] Alessandra Fragale,[1] Courtney Nicholas,[1] Stephanie Viani,[1] Raju Kucherlapati,[3] Martin Lipkin,[4] Kan Yang,[4] Leonard Augenlicht[1]

The gastrointestinal tract is lined by a layer of mucus comprised of highly glycosylated proteins called mucins. To evaluate the importance of mucin in intestinal carcinogenesis, we constructed mice genetically deficient in Muc2, the most abundant secreted gastrointestinal mucin. Muc2$^{-/-}$ mice displayed aberrant intestinal crypt morphology and altered cell maturation and migration. Most notably, the mice frequently developed adenomas in the small intestine that progressed to invasive adenocarcinoma, as well as rectal tumors. Thus, Muc2 is involved in the suppression of colorectal cancer.

Mucins are highly glycosylated proteins that are the major component of the mucus that lubricates and protects underlying intestinal epithelium (*1*). Alterations of mucin expression and glycosylation have been observed in human colon cancer specimens (*2*), but the role of these proteins in tumorigenesis remains unclear. To evaluate the importance of mucin in the early stages of intestinal carcinogenesis, we generated mice genetically deficient in Muc2, the most abundant secreted gastrointestinal apomucin, which is the unglycosylated form of mucin (*3, 4*). Targeted inactivation of the *Muc2* gene was achieved by replacing a genomic fragment (*5*) harboring exons 2 to 4 of *Muc2* with a *phosphoglycerate kinase-neomycin* (*PKG-Neo*) cas-

[1]Department of Oncology, Albert Einstein Cancer Center/Montefiore Medical Center, 111 East 210 Street, Bronx, NY 10467, USA. [2]Memory Pharmaceuticals Corporation 100, Philips Parkway, Montvale, NJ 07645, USA. [3]Harvard Medical School, Brigham & Women's Hospital, 20 Shattuck Street, Boston, MA 02115, USA. [4]Strang Cancer Prevention Center, 1300 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: velcich@aecom.yu.edu