Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 6437–6442, June 1996
Biochemistry

The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface

(protein–protein interaction/E-group colicins)

KIN-FU CHAK*, MARTIN K. SAFO†, WEN-YEN KU†‡, SHIH-YANG HSIEH*, AND HANNA S. YUAN†§

*Institute of Biochemistry, National Yang-Ming University, Taipei, Taiwan, Republic of China; †Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529, Republic of China; and ‡Graduate Institute of Life Science, National Defence Medical Center, Taipei, Taiwan, Republic of China

Communicated by Richard E. Dickerson, University of California, Los Angeles, CA, February 29, 1996 (received for review December 11, 1995)

ABSTRACT **The immunity protein of colicin E7 (ImmE7) can bind specifically to the DNase-type colicin E7 and inhibit its bactericidal activity. Here we report the 1.8-Å crystal structure of the ImmE7 protein. This is the first x-ray structure determined in the superfamily of colicin immunity proteins. The ImmE7 protein consists of four antiparallel α-helices, folded in a topology similar to the architecture of a four-helix bundle structure. A region rich in acidic residues is identified. This negatively charged area has the greatest variability within the family of DNase-type immunity proteins; thus, it seems likely that this area is involved in specific binding to colicin. Based on structural, genetic, and kinetic data, we suggest that all the DNase-type immunity proteins, as well as colicins, share a "homologous-structural framework" and that specific interaction between a colicin and its cognate immunity protein relies upon how well these two proteins' charged residues match on the interaction surface, thus leading to specific immunity of the colicin.**

E-group colicins (from E1 to E9) are plasmid-borne antibiotic-like bacteriocins that are active against sensitive *Escherichia coli* and closely related coliform bacteria (1). They bind to the vitamin B12 receptor, BtuB, and subsequently translocate across the outer and cytoplasmic membranes, inducing cell death (2). Three cytotoxic classes of E-group colicins have thus far been identified, including pore-forming colicins such as colicin E1 (3), RNase colicins such as colicins E3 (4) and E6 (5), and DNase colicins such as colicins E2 (6), E7 (7), E8 (8), and E9 (7). Colicin E7 (ColE7) is a nonspecific endonuclease that causes both single- and double-stranded breaks in the DNA of sensitive cells. Production of ColE7 is regulated by a "SOS" response operon that encodes ColE7, ImmE7 (immunity protein of colicin E7), and a lysis protein for transportation of the ColE7/ImmE7 complex. Immediately after production, colicin forms a complex with its coordinately produced ImmE and thus neutralizes its toxicity toward the host cell.

The DNase-type colicins contain almost identical sequences in their translocation and receptor recognition domains, which are located in the N-terminal two-thirds of the sequence. C-terminal endonuclease domains (T2A domain) of DNase-type colicins are ≈80% identical, and sequences of their corresponding immunity proteins are 60–70% identical (7). However, despite the high sequence identities in either colicins or immunity proteins, an immunity protein can only completely protect a cell from the action of its own cognate colicin (9). Neither the mechanism for the specific protein–protein interaction between colicins and immunity proteins nor the inhibition of toxicity incurred after the formation of colicin/ImmE complex has been explained yet. Thus the crystal structure of ImmE7 may provide invaluable information, expanding our limited knowledge of the specific interactions between proteins. Here we report the 1.8-Å crystal structure of the ImmE7 protein. A surface rich in acidic residues has been identified in the structure, and this area seems likely to be involved in specific binding to colicin.

METHOD

We have previously reported the crystallization of the ImmE7 protein in two different crystal forms (10). Crystals were grown by vapor diffusion methods from solutions containing ammonium phosphate as the precipitant. Chunky plates grew over the course of days, with typical dimensions of $0.5 \times 0.5 \times 0.04$ mm^3. Crystals used for structure determination belong to space group I222 with cell parameters a = 75.14 Å, b = 50.48 Å, and c = 45.36 Å, and one molecule per asymmetric unit. Diffraction data were collected on an RAXIS IIC Imaging Plate System (Molecular Structure, The Woodlands, TX). Two derivatives were used in solving the structure by multiple isomorphous replacement method (Table 1). Heavy atom derivatives were prepared by soaking native crystals in the corresponding heavy atom solution. Heavy atom sites in the derivatives were located by difference Patterson techniques and refined by HEAVY (11).

An initial multiple isomorphous replacement electron density map was calculated at 3.0-Å resolution and improved by the program DM (12), using a combination of solvent flattening, histogram mapping, Sayer's equation, and skeletonization. The multiple isomorphous replacement map showed continuous electron density with well-defined side chains for almost the entire molecule, except for a weak density region at residues 56–64. Refinement by simulated annealing using X-PLOR (13) decreased the R-factor to 0.308 for data between 8.0 and 2.2 Å, with 85% residues present. The DM-improved multiple isomorphous replacement phases were combined with model phases by SIGMAA (14), and the weak density region appeared clearly in the (2Fo-Fc) map. The model was further rebuilt and refined by X-PLOR, with resolution slowly increased to 1.8 Å. The current model consists of 85 residues (3–87) and 60 water molecules. The first two amino acids are presumably disordered in the crystal structure. Fig. 1 shows a sample final omit (2Fo-Fc) map, and except for the side chains of Asn-5, Glu-39, Lys-43, Asp-59, and Asn-60, all residues are well-defined. The average B factor for the nonhydrogen atoms in the protein molecule is 29.9 Å^2, and the Ramachandran plot shows no violations of accepted backbone torsion angles.

Abbreviations: ColE7, colicin E7; ImmE, immunity protein of E-group colicin.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1CEI, R1CEISF). This information is embargoed until March 3, 1997.

§To whom reprint requests should be addressed. e-mail: mbyuan@ccvax.sinica.edu.tw.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Summary of crystallographic analysis

	Native	$Pt(CH_3)_3I$	K_2PtCl_4
Structure Determination			
Heavy atom concentration, mM		Saturation	5
Soaking time, days		1	5
Resolution, Å	1.8	2.2	2.5
Reflections, unique/total	7917/31,416	4622/15,586	3138/11,312
Data completeness	95.2	98.6	97.7
R_{merge}*, %	5.4	7.8	7.7
R_{iso},† %		22.9	18.0
R_{cullis},‡ 15.0–3.0 Å		0.58	0.61
Phasing power, centric/acentric		1.72/2.11	0.85/1.04
Mean overall figure of merit, 15.0–3.0 Å		0.55§	
Structure refinement			
Resolution, Å	6.0–1.8		
R-factor, %	18.7		
Observations, $F \geq 2\sigma(F)$	7376		
Total number of protein atoms	681		
Total number of water molecules	60		
r.m.s.d. bond lengths, Å	0.010		
r.m.s.d. bond angles, degree	1.132		

*$R_{merge} = \frac{\sum_h \sum_i |I_{h,i} - \langle I_h \rangle|}{\sum_h \sum_i I_{h,i}}$, where $\langle I_h \rangle$ is the mean intensity of the i observations for a given reflection

†$R_{iso} = \sum (|Fder| - |Fnat|) / \sum |Fnat|$

‡$R_{cullis} = \sum [|Fh| - (|Fph| - |Fp|)] / \sum |Fh|$, for centric reflections.

§Value is 0.75 after density modification.

RESULTS

The structure of the ImmE7 protein is shown in Fig. 2. The molecule comprises four α-helices, designated helices 1–4, and one turn of a 3_{10}-helix at the N-terminus. The detailed secondary structure is illustrated in Fig. 3. Positions of the three long α-helices are similar to those of ImmE9, whose secondary structure has been assigned by NMR spectroscopic studies (16). The whole molecule is topologically similar to a four-helix bundle structure such as the calcium-binding protein ICaBP (17), with four antiparallel helices linked to each other in an up-and-down manner. However, the orientations and positions, as well as helix lengths, are quite different. The biggest variations are between helix 3 and 4; helix 3 is only a very short, one-turn α-helix, and helix 4 is shifted downward and tilted away from the molecule to pack against the N- and C-termini of helices 1 and 2.

The four helices wrap around a central hydrophobic core and stabilize folding of the entire soluble protein. Side chains of Ile-7, Tyr-10, Phe-15, Leu-18, Leu-19, Ile-22, Leu-34, Leu-37, Leu-38, Phe-41, Ile-44, Leu-53, Ile-54, Ile-68, Val-69, Ile-72, and Phe-84 are involved in hydrophobic interactions (see Fig. 4*a*). These positions are most frequently occupied by hydrophobic residues in each of the DNase-type immunity proteins. We observe three important hydrogen bonds: (*i*) between Glu-12 (OE1) and Lys-73 (NZ), (*ii*) between Thr-45 (O) and Arg-76 (NH2), and (*iii*) between Glu-46 (OE2) and Arg-76 (NH2). These bonds tie helices 1, 2, and 4 together, and are located at the bottom of helices 1 and 2 as viewed in Fig. 4*b*. Helix 4 acts like a platform which holds down the starting points for helices 1 and 2. The five residues involved in hydrogen bonding are conserved in all the DNase-type immunity proteins. Combining all these observations, we predict that the structures of ImmE2, ImmE8, and ImmE9 must bear a high resemblance to ImmE7.

The ImmE7 protein possesses two long and mostly extended N-terminal and C-terminal tails. At the end of the N-terminal tail, a 3_{10} helix is found in which Ile-7 (O) hydrogen bonds to Tyr-10 (N). The remaining N-terminal tail packs against helices 1 and 2 and makes several hydrogen bonds with the residues in helix 1, helix 2, or the C-terminal tail, including Lys-4 (O) to His-40 (NE2), Ser-8 (OG) to Gly-83 (O), Asp-9 (O) to Lys-84 (NZ), and Tyr-10 (O) to Lys-85 (N). The C-terminal tail closes the hydrophobic core at the bottom of the molecule. Temperature factors for the two terminal tails and the two loops linking helices 1/2 and helices 3/4 are higher than average, especially the long loop from residues 56 to 64. Therefore we describe these two loops as variable loop 1 (average B-factor for nonhydrogen atoms, 38.4 Å^2) and variable loop 2 (average B-factor for nonhydrogen atoms, 43.5 Å^2). Because variable loops 1 and 2 make contacts with the neighboring symmetrically related molecules, crystal packing forces may fix their conformation (18), and these two loops might have different structures in solution.

DISCUSSION

The main challenge with immunity proteins at present is to determine how they exert their specific binding function, and to locate their specificity-determining region. It has been shown that immunity proteins protect the host cell from colicin action by binding to the colicin's C-terminal toxicity domain (T2A domain; refs. 19–21). Using site-directed mutagenesis to generate mutants in the T2A domain of the colicin E9, Curtis and James proposed that the protein–protein interaction region resides between Ser-525 and Val-546 (22). It is noteworthy that this region of E-group colicins has many positively charged residues. This region of the T2A domain appears to be the specific target for the highly negatively charged immunity protein. Therefore, genetic studies suggest that interactions between colicins and immunity proteins rely on the attractive electrostatic forces. Based on the crystal structure of the ImmE7 protein, we calculated the solvent accessible area for each residues by program 3D PROFILE (23). Except for Glu-46 and Glu-74, all acidic residues with a solvent accessible area >80 Å^2 are located at or near the two variable loops, including Glu-23, Asp-31, Asp-32, Glu-39, Asp-59, Asp-62, Asp-63, and Glu-66 (see Fig. 5*b*). Gene-fusion experiments (25) between ImmE9 and ImmE8 and chemical modification of Cys-23 in ImmE9 have demonstrated that the specificity-determining

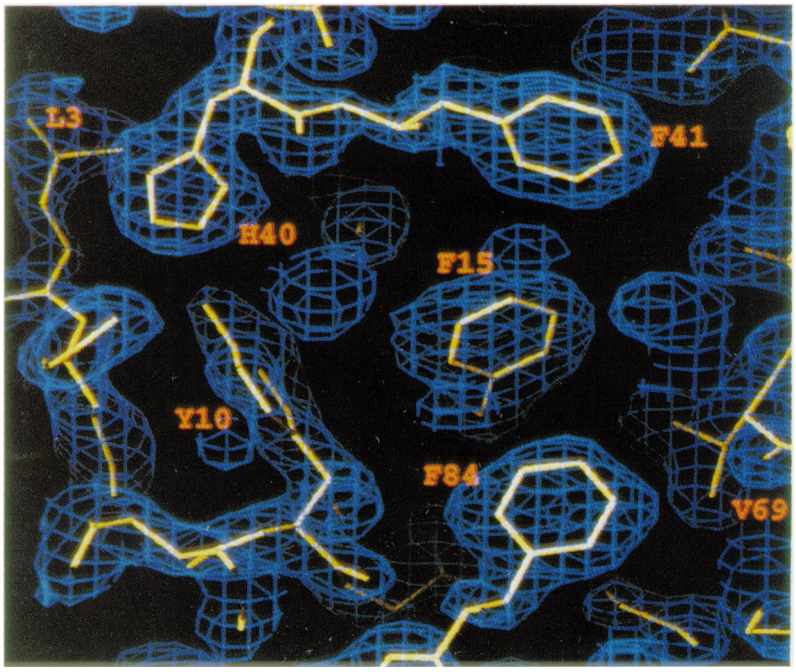

FIG. 1. Representative omitted electron density map of ImmE7 protein. A view of the hydrophobic core among Tyr-10, Phe-15, His-40, Phe-41, and Phe-84 illustrates the close packing. The map was calculated by omitting Tyr-10, Phe-15, and Phe-84 from the x-ray model, simulated annealing refinement was done with a 3-Å spherical shell of fixed atoms surrounding the omitted region, and it was contoured at 1.0 rms above the average electron density.

region is located at positions 16–43, which covers the protruding areas of helix 1, helix 2, and variable loop 1. These results suggest that exposed charged residues located in the protruding region may play an important role in recognizing colicin.

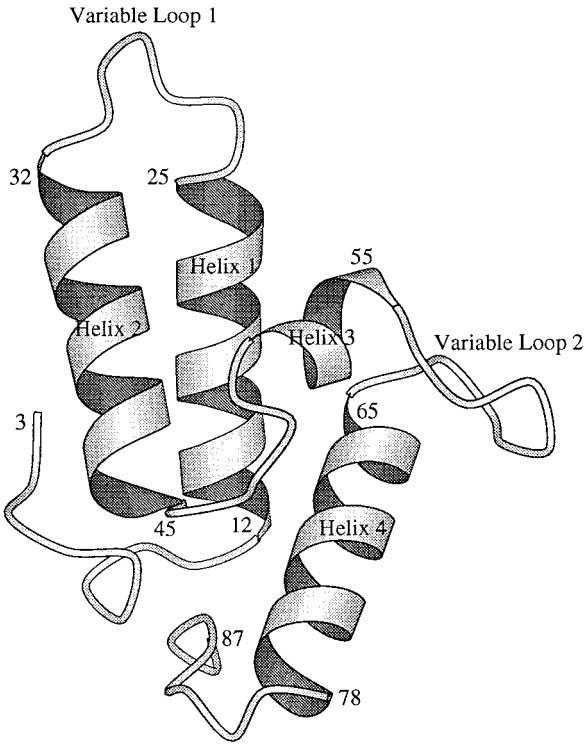

FIG. 2. Schematic drawing of the molecular structure of ImmE7 protein. The protein consists of mainly four α-helices, labeled helices 1–4. The numbers in the figure indicate the starting and ending residues for each helix and the N- and C-termini. Loops 1 and 2 are more flexible and are thus denoted as variable loops 1 and 2. Helices 1 and 2 protrude out the molecule, and together with variable loops 1 and 2 are probably involved in protein–protein interactions.

Fig. 5 shows the electrostatic surface potential of the ImmE7 protein, displayed by the program GRASP (24). Negatively charged surface (potential <10 kT) is deep red, and positively charged surface (potential >10 kT) is deep blue. A strongly acidic region is found on the top half of the molecule, especially in the protruding area near variable loop 1. The bottom half of the molecule is more neutral and hence is shown mostly in white. This confirms the previous analysis of solvent accessible area, which indicates that the ImmE7 probably interacts with ColE7 by the protruding surface near variable loop 1. Variable loop 2 also shows a deep red negative surface, and hence we cannot exclude the possibility that this region may also participate in protein–protein interactions, although this has not been confirmed by genetic studies.

Even though the specificity-determining region of ImmE9 has been identified by gene fusion experiments as located at positions 16–43, the single-site mutation studies cannot distinguish which residues are responsible for specificity. Changing residues of ImmE9 in this region to those of the equivalent ImmE8 did not affect the phenotype conferred by the mutant protein, with one exception. Mutating Val-34 to Asp, resulted in an additional increase in low-level ColE8 immunity (25). A similar result has been found in the mutation of the T2A domain of ColE9. Substitution of a single residue cannot change the phenotype of ColE9, but changing four additional residues to the corresponding amino acids of ColE8 brought the phenotype of the mutant protein closer to that of ColE8 (22).

Kinetic studies show that the dissociation constants between ColE9 and immunity proteins are in the order of 10^{-14} M for ImmE9, 10^{-8} M for ImmE2, 10^{-6} M for ImmE8, and 10^{-4} M for ImmE7 (26). This demonstrates not only that ColE binds to its own immunity protein with the highest affinity but also that it binds less strongly to other immunity proteins. From this and from the mutational studies, we believe that the DNase-type immunity proteins, as well as the colicins, must share a "homologous-structural framework" and that the specific in-

```
                  Helix 1               Helix 2
              _____________         _____________
       1        11        21        31        41
       |         |         |         |         |
ImmE7  MELKNSISDYTEAEFVQLLKEIEKENVAATDDVLDVLLEHFVKITEHPDG
ImmE2  ====H==========LEFV=K=CRAE.G==EEDDNK=VRE=ERL======
ImmE8  ============T==KKIIED=INCE.G.DEKKQ=DN====ISV======
ImmE9  ====H==========L==VTT=CNAD.TSSEEE=VKLVT==EEM====S=

       Helix 3          Helix 4
        ____        _____________
       51        61        71        81
       |         |         |         |
ImmE7  TDLIYYPSDNRDDSPEGIVKEIKEWRAANGKPGFKQG -87
ImmE2  S======R=D=E===================S===== -86
ImmE8  S======EG=N=====AVI============S===== -85
ImmE9  S======KEGD====S===NTV=Q=======S===== -86
```

FIG. 3. Amino acid sequences of ImmE7 protein from *E. coli* and multiple sequence alignment of the four known DNase-type immunity proteins. The location of helices 1 (positions 12–25), 2 (positions 32–45), 3 (positions 52–55), and 4 (positions 65–78) are displayed above the sequence. Secondary structure is that defined by the program PROCHECK (15). The amino acids that are identical to ImmE7 are denoted with an equal sign, and the gaps are denoted with a period.

teraction is determined by how well these two proteins' charged residues match on the interaction surface. In terms of a key and lock analogy, not only must the key fit into the lock (homologous framework), but the teeth of the key must also

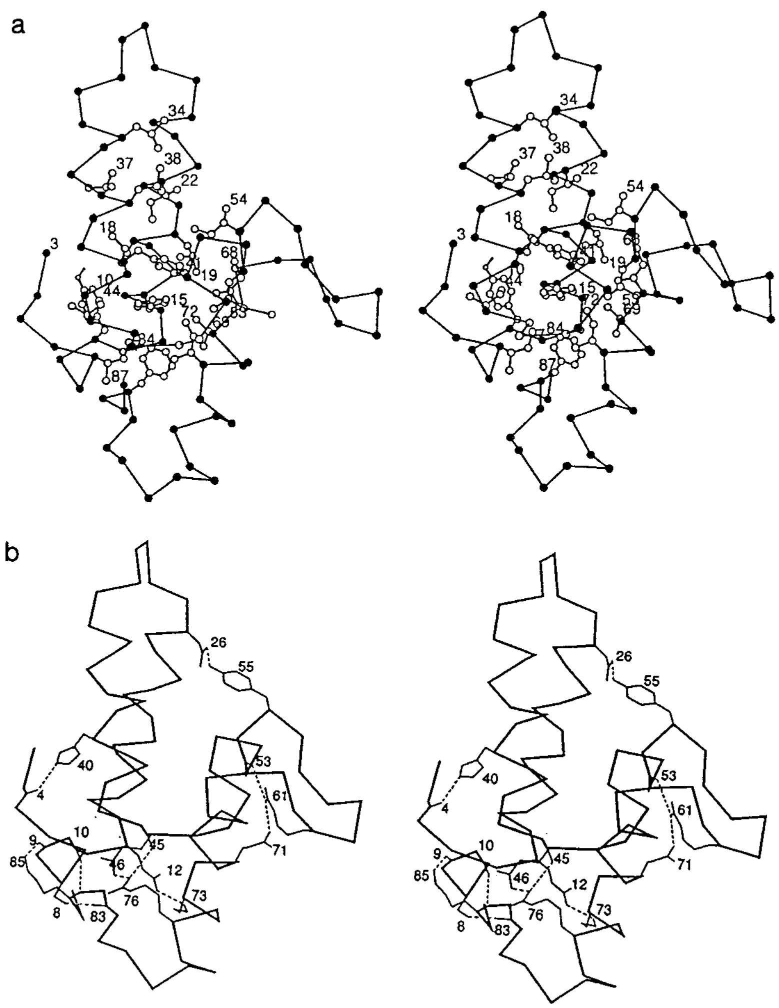

FIG. 4. (*a*) Stereo drawing illustrating the distribution of the buried hydrophobic residues in ImmE7 protein. The main chain is presented only by straight lines between α-carbon atoms (filled circles). The side-chains pointing toward the interior of the four-helix bundle with a solvent-accessible area <30 Å^2 are shown (open circles). These hydrophobic residues stabilize the core of the protein. (*b*) Stereoview of the hydrogen-bonding network in ImmE7 protein. Side chains shown are those whose hydrogen bonds help hold the four helices in a fixed geometry. Only α-carbon backbone is shown with exceptions of residues 4, 9, 10, 45, 53, and 83. The carbonyl groups of these residues are also shown with their Cα-atom, because they are involved in the hydrogen bonding.

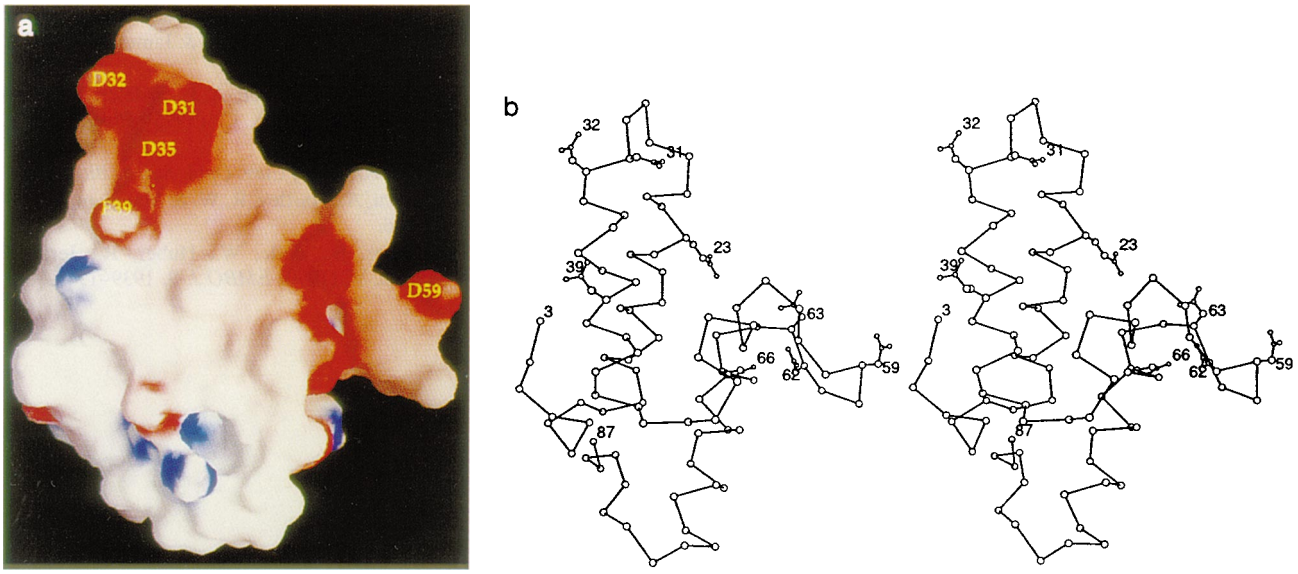

FIG. 5. (*a*) Electrostatic surface potential on the ImmE7 protein displayed with the program GRASP (24). Negative potentials (<10 kT) are displayed in deep red, and positive potentials (>10 kT) are displayed in deep blue. Solvent-accessible surfaces were calculated with a water probe radius of 1.4 Å. Obviously, the bottom half of the molecule is more neutral, and hence the surface is mostly in white. On the other hand, the top surface is rich in charged residues, and this surface seems likely to be involved in specific binding to colicin. (*b*) The stereoview showing the positions of the exposed acidic residues in the α-carbon backbone structure. Only the acidic residues (except Glu-46 and Glu-74) with solvent-accessible area >80 Å^2 are plotted. This figure is approximately in the same view as that in *a*.

completely match the pins of the lock (specific interactions) to increase the affinity of the binding and lead to the specific immunity of the colicin. Moreover, the specific interaction depends on many residues, such as the eight acidic residues identified on the top surface of ImmE7. Therefore a single-site mutation could probably weaken but not destroy the tight interactions. Increasing the numbers of mutated residues could gradually alter the specificity of an immunity protein toward the newly targeted colicin and vice versa. Evolution of distinct types of colicins and their cognate immunity proteins may be achieved by varying only the charged residues within the structural framework located in the T2A domain of colicin and the top surface of immunity proteins, respectively.

Most of the inhibitory proteins whose three-dimensional structures have been determined are either serine proteinase or cysteine proteinase inhibitors (27). Barstar is another example of an RNase inhibitor that inhibits the activity of barnase (28). Many of the serine proteinase inhibitors fall into a single class (the "small" inhibitors) and inhibit the target enzymes by binding across the active site, mainly through the use of a single exposed loop. Cysteine proteinase inhibitors such as stefin B bind to the cysteine proteinase papain in a similar fashion (29), with the main interactions provided by two hairpin loops and the amino terminal tail. On the contrary, barstar interacts with barnase by charged residues located mainly on an α-helix and adjacent loop segment. This leads us to suggest that ImmE7 may be a mixture of the above inhibitors and that not only the variable loops 1 and 2 but also the protruding area in helices 1 and 2 may be involved in protein–protein interactions. The amino acid sequences of the DNase-type immunity proteins are most variable in these regions. This is probably related to binding as well as the high specificity of the inhibitory function of the proteins.

CONCLUSION

The structure of ImmE7 protein provides a novel system with which to study the protein–protein interactions, and most interestingly, the mechanism of specific immunity. Many questions remain unanswered, such as how the immunity protein inhibits the activity of colicin and how it dissociates from very tight binding with colicin. However, the structure presented here offers a starting point for further crystallographic, biochemical, and genetic studies of immunity proteins. In particular, it should now be possible to determine systematically which regions of ImmE7's protein-binding surface mediate the very tight and specific interactions with ColE7. The structure also provides a framework for other immunity proteins of the same superfamily.

We thank Dr. L.-S. Kan for discussion. This research was supported by grants from the National Science Council of the Republic of China to H.S.Y. (NSC85–2311-B001–094), K.-F.C. (NSC85–2331-B010), S.-Y.H. (NSC84–2113-M001–013), and M.K.S (NSC85–2811-B001–039).

1. Witkin, E. M. (1976) *Bacteriol. Rev.* **40,** 869–907.
2. Pugsley, A. P. (1984) *Microbiol. Sci.* **1,** 168–175.
3. Cramer, W. A., Dankert, J. R. & Uratani, Y. (1983) *Biochim. Biophys. Acta* **737,** 173–179.
4. Bowman, C. M., Sidikaro, J. & Nomura, M. (1971) *Nat. New Biol.* **234,** 133–137.
5. Akutsu, A., Masaki, H. & Ohta, T. (1989) *J. Bacteriol.* **171,** 6430–6436.
6. Schaller, K. & Nomura, M. (1976) *Proc. Natl. Acad. Sci. USA* **68,** 3989–3993.
7. Chak, K.-F., Kuo, W.-S., Lu, F.-M. & James, R. (1991) *J. Gen. Microbiol.* **137,** 91–100.
8. Toba, M., Masaki, H. & Ohta, T. (1988) *J. Bacteriol.* **170,** 3237–3242.
9. Cooper, P. C. & James, R. (1984) *J. Gen. Microbiol.* **130,** 209–215.
10. Ku, W.-Y., Wang, C.-S., Chen, C.-Y., Chak, K.-F., Safo, M. K. & Yuan, H. S. (1995) *Proteins Struct. Funct. Genet.* **23,** 588–590.
11. Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987) *Acta Crystallogr.* **A34,** 1–5.
12. Cowtan, K. D. (1996) *Acta Crystallogr.* **852,** 43–48.
13. Brunger, A. T. (1992) X-PLOR: A System for x-ray Crystallography and NMR (Yale Univ. Press, New Haven, CT), version 3.1.
14. Read, R. J. (1986) *Acta Crystallogr.* **A42,** 140–149.
15. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
16. Osborne, M. J., Lian, L.-Y., Wallis, R., Reilly, A., James, R., Kleanthous, C. & Moore, G. R. (1994) *Biochemistry* **33,** 12347–12355.
17. Szebenyi, D. M. E. & Moffat, K. (1981) *Nature (London)* **294,** 327–332.

18. Kossiakoff, A. A., Randal, M., Guenot, J. & Eigenbrot, C. (1992) *Proteins Struct. Funct. Genet.* **14,** 65–74.
19. Ohno-Iwashita, Y. & Imahori, K. (1980) *Biochemistry* **19,** 652–659.
20. Wallis, R., Reilly, A., Barnes, K., Abell, C., Campbell, D. G., Moore, G. R., James, R. & Kleanthous, C. (1994) *Eur. J. Biochem.* **220,** 447–454.
21. Ohno-Iwashita, Y. & Imahori, K. (1982) *J. Biol. Chem.* **257,** 6446–6451.
22. Curtis, M. D. & James, R. (1991) *Mol. Microbiol.* **5,** 2727–2733.
23. Bowie, J. U., Luthy, R. & Eisenberg, D. (1991) *Science* **253,** 164–170.
24. Nicholls, A. & Honig, B. (1991) *J. Comp. Chem.* **12,** 435–445.
25. Wallis, R., Moore, G. R., Kleanthous, C. & James, R. (1992) *Eur. J. Biochem.* **210,** 923–930.
26. Wallis, R., Leung, K.-Y., Pommer, A. J., Videler, H., Moore, G. R., James, R. & Kleanthous, C. (1995) *Biochemistry* **34,** 13751–13759.
27. Bode, W. & Huber, R. (1991) *Curr. Opin. Struct. Biol.* **1,** 45–52.
28. Guillet, V., Lapthorn, A., Hartley, R. W. & Mauguen, Y. (1993) *Structure* **1,** 165–176.
29. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990) *EMBO J.* **9,** 1939–1947.