Crystal Structure of the DNA Binding Domain of the Heat Shock Transcription Factor

Celia J. Harrison, A. Andrew Bohm, Hillary C. M. Nelson*

The structure of the DNA binding domain, determined at 1.8 angstrom resolution, contains a three-helix bundle that is capped by a four-stranded antiparallel β sheet. This structure is a variant of the helix-turn-helix motif, typified by catabolite activator protein. In the heat shock transcription factor, the first helix of the motif (α2) has an α-helical bulge and a proline-induced kink. The angle between the two helices of the motif (α2 and α3) is about 20° smaller than the average for canonical helix-turn-helix proteins. Nevertheless, the relative positions of the first and third helices of the bundle (α1 and α3) are conserved. It is proposed here that the first helix of the three-helix bundle be considered a component of the helix-turn-helix motif.

All cells respond to stress, such as heat shock or chemical or environmental perturbations, by rapidly expressing a small number of proteins known as the heat shock proteins. These proteins are thought to protect the cell from damage. In eukaryotes, the expression of these proteins is regulated by the heat shock transcription factor (HSF) (*1*), which is a trimeric DNA binding protein that recognizes a DNA sequence called the heat shock element (HSE) (*2, 3*). HSFs have been cloned and sequenced from many eukaryotes, including yeast, fruit fly, tomato, chickens, mice, and humans (*4–6*). Two highly conserved regions are found in all HSFs: the DNA binding domain and the trimerization domain. We have shown previously that yeast HSF exists as a trimer in solution and that it binds HSEs as a trimer (*2*). The DNA binding monomer is a small globular domain of approximately 90 residues (*7*). Sequence alignment programs have failed to show any similarity between the DNA binding domain of HSF and other known structural motifs (*8*). The model for DNA binding assumes each monomer contacts a 5-bp nGAAn box of the HSE (*3*). Although the DNA binding domain alone binds DNA with a lower affinity than the trimeric form of the protein, we have shown using deoxyribonuclease I footprinting that the domain recognizes HSEs specifically and has a footprint similar to that of the trimeric protein (*8*).

The DNA binding domain crystallized in the orthorhombic space group $P2_12_12_1$, with unit cell dimensions of a = 36.8, b = 34.3, and c = 67.0 Å and one molecule per asymmetric unit (*9*). One native data set to 2.7 Å was collected on an Enraf-Nonius FAST area detector, and another native data set to 1.8 Å was collected on a Rigaku RAXIS IIC Imaging Plate System (Table 1). Two derivatives were used in solving the structure: potassium hexachloroiridate was cocrystallized with the protein at a concentration of 1 mM and selenium was incorporated into the protein as selenomethionine (*10*).

We solved the structure by the method of multiple isomorphous replacement and refined the model using X-PLOR 3.1 (*11, 12*) (Table 2). Figure 1A shows sample electron density from a simulated annealing omit map. The current model includes side chains for all visible 88 amino acids, as well as 81 water and 2 acetate molecules, for a total of 873 nonhydrogen atoms. Four residues (269 to 272) remain untraced. The R factor is 18.9% for all data between 8.0 and 1.83 Å, and the value for R_T^{free} is 24.8%. The average B factor for all nonhydrogen atoms is 19.5 Å^2. The root-mean-square (rms) deviations from ideal bond lengths and angles are 0.010 Å and 1.45°, respectively. There are no Ramachandran outliers, and the positions of all bound water molecules and acetate ions are chemically reasonable.

In the DNA binding domain of HSF, the three α helices form a compact three-helical bundle (Fig. 1B). The four β strands form a single antiparallel β sheet, which closes off the hydrophobic core formed by the bundle. An extended loop, formed by residues 260 through 276, lies between the last two β strands and protrudes out from the molecule. Although residues 260 to 268 and 273 to 276 are well ordered, residues 269 to 272 have no clearly interpretable electron density. The loop is partly tethered by crystal contacts to a symmetry related molecule, and it is possible that it will adopt a different conformation when outside the crystal environment. The length of this loop differs in the various HSF sequences, with the only conserved residue being Gly^{267} at the turn of the loop (Fig. 2).

The second and third helices (α2 and α3) comprise the helix-turn-helix motif (*13*). This region lies in the most conserved part of the HSF DNA binding domain, with 12 of 27 amino acids identical in all known HSFs (Figs. 1B and 2). This region has a sequence similar to that of a portion of the yeast suppressor gene for flocculation (*SFL-1*) and to that of a portion of the yeast suppressor gene for cell wall β-glucan assembly (*SKN-7*), which suggests that the two proteins encoded by these genes may also contain a helix-turn-helix motif (*14*).

Surprisingly, helix α2 has a proline-induced kink as well as an α-helical bulge, which results in major perturbations of the helix (Fig. 3A). The helix is kinked by 29° centered around the proline, compared to the average of 26° for proline-induced kinks (*15*). The proline at residue 238 is found in all known HSFs. A corresponding proline is found in SKN-7, but not in SFL-1.

Immediately before the proline is an α-helical bulge (Fig. 3A). The bulge is caused by a "looping out" of the helix between two hydrophobic residues, Phe^{232} and Leu^{237}, which are buried in the hydrophobic core. The phenylalanine and leucine are conserved in all HSFs. Normally, α helices within proteins have two or three amino acids between buried residues of the helix. The DNA binding domain of HSF has four amino acids between the two anchoring hydrophobic residues in the helix,

Table 1. Data statistics. Data sets for the two derivatives were collected on a FAST area detector. Cell dimensions for these two derivatives were identical to those of the native crystal.

Derivative	Resolution (Å)	Total observations*	Unique reflections	Completeness (%)	R_{sym} (%)†
Native 1‡	2.7	7,981	2,373	92	4.3
Native 2§	1.8	40,007	7,654	96	6.6
Selenium‡	2.5	4,333	2,930	89	4.3
K_2IrCl_6‡	2.6	6,914	2,354	80	4.3

*All reflections $F/\sigma(F) > 0.0$ were used; F, structure factor amplitude. †R_{sym} (%) = $100\,\Sigma|\langle I_i\rangle - I_i|/\Sigma I_i$, where $\langle I_i\rangle$ is the average of I_i (the intensity at any given position) over all symmetry equivalents. ‡Data collected at 8°C on a FAST area detector with a Rigaku rotating anode operating at 50 kV and 100 mA. §Data collected at 8°C on a RAXIS IIC imaging plate with a Rigaku rotating anode operating at 50 kV and 150 mA.

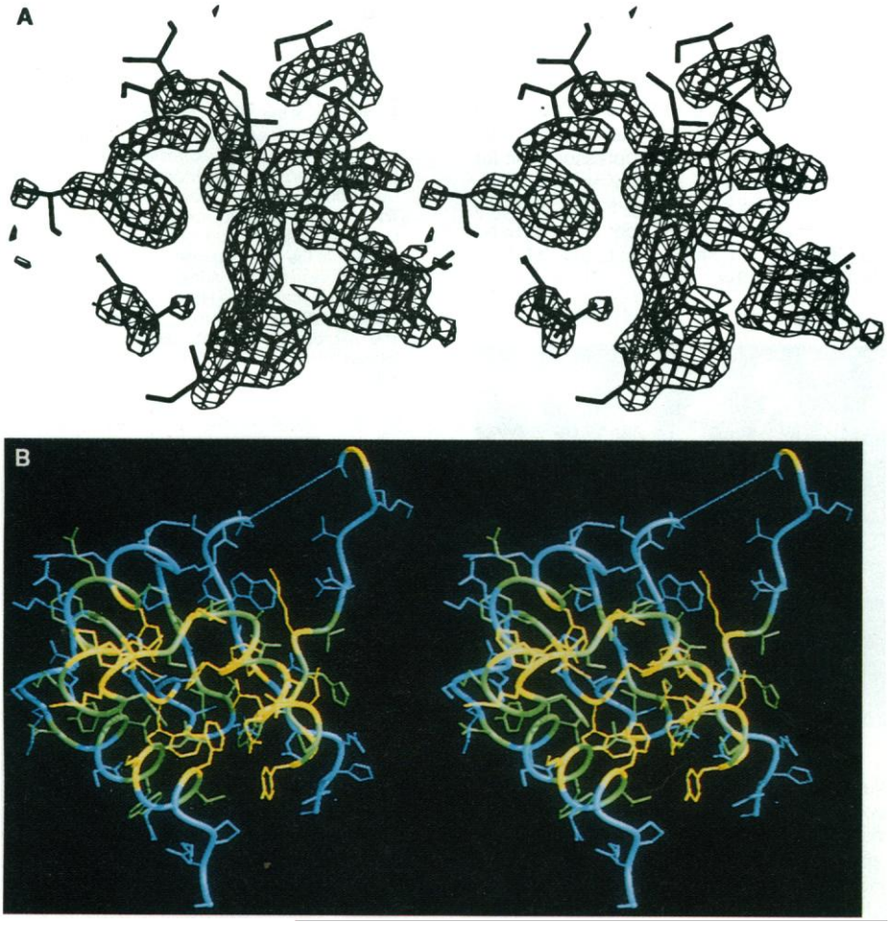
C. J. Harrison and H. C. M. Nelson, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720.
A. A. Bohm, Structural Biology Division of the Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, CA 94720.
*To whom correspondence should be addressed.

and consequently Glu^{235} bulges out (*16*). Naturally occurring α-helical bulges have been seen in other proteins, including mercuric ion reductase and ribonucleotide reductase (*17*). In both of these cases, the extra amino acids appear to be required for activity. Studies with helix insertion mutants of staphylococcal nuclease and lysozyme show that the inserted amino acid is accommodated as a bulge (*18, 19*). At present, we do not know the functional significance of the α-helical bulge or proline kink in the α2 helix of HSF. Given the presence of these two features, it is not surprising that secondary structure programs failed to predict a helix-turn-helix motif in HSF.

Table 2. Phasing statistics. In the case of the selenomethionine derivative, selenium and sulfur differ by only 18 electrons, but the derivative has an occupancy of at least 95%. In the case of the iridium data set, we estimate that the iridium was incorporated into the crystal with ~20% occupancy for the major site and ~10% occupancy for the minor site. Difference Patterson maps were calculated and inspected for the peaks. Peaks corresponding to two selenium sites and one major and one minor iridium site were picked by HASSP, an automated heavy atom search and superimposition program (*28*). Single isomorphous replacement phases were calculated for each derivative with the package PROTEIN (*29*), and the origins were correlated by difference Fourier techniques. Heavy atom parameters were refined and multiple isomorphous replacement phases calculated within PROTEIN. The handedness of the structure was determined by the stereochemistry of the α helices.

Derivative	Resolution (Å)	Unique reflections*	R_{iso} (%)†	Number of sites	R_{cullis}‡	Phasing power§
Iridium	2.7	2,005	4.3	2	0.68	1.51
Selenium	2.7	2,095	9.8	2	0.66	1.75

*Unique reflections from 10.0 to 2.7 Å, with 2,244 native reflections phased of 2,246 total. †$R_{iso} = 100\ \Sigma|F_{PH}^2 - F_P^2|/\Sigma\ (F_{PH}^2 + F_P^2)$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively. Isomorphous differences are reported after local scaling. ‡$R_{cullis} = \Sigma||F_{PH} \pm F_P| - F_H\ (calc)|/\Sigma|F_{PH} \pm F_P|$ for centric reflections. §Phasing power is the rms heavy atom structure factor amplitude divided by the residual lack of closure error.



Helix-turn-helix proteins can be divided into two classes: an all α-helical class and an α helix and β sheet (α + β) class. Our structure places HSF in a family of similar α + β DNA binding proteins that use variations on the helix-turn-helix motif to contact DNA. These proteins [catabolite activator protein (CAP), histone H5, HNF-3/*fork head* motif, biotin repressor, and HSF] share an overall compact structure of three α helices and a three- or four-stranded β sheet (*20–23*). HSF and CAP share an α1-β1-β2-α2-α3-β3-β4 sequence of secondary structure elements. HNF-3/*fork head* motif, histone H5, and biotin repressor share an α1-β1-α2-α3-β2-β3 sequence of secondary structure elements.

When the α1 and α3 helices of HSF are superimposed with a typical α + β helix-turn-helix protein such as CAP, it is obvious that the α2 helices are in quite different positions (Fig. 3B). Biotin repressor is similar to CAP, but histone H5 and HNF-3/*fork head* motif also have displaced α2 helices. For all of these proteins, α1 and α3 are more conserved in their relative positions than α2 and α3. This emphasizes the importance of α1 and α3 in the helix-turn-helix motif. In addition to a variation in the interhelical angle, there are differences in the length of the turn between the helices of the helix-turn-helix motif. CAP and biotin repressor have three amino acids in the turn, with glycine at the first position in the turn. The other proteins have longer turns: HSF has five amino acids, histone H5 has six amino acids, and HNF-3/*fork head* motif has eight amino acids. We would place the three eukaryotic proteins (HSF, HNF-3/*fork head* motif, and histone H5) in a special class of helix-turn-helix variants that have a noncanonical α2-α3 interhelical angle and turn (*24*).

The different modes of DNA binding for the α + β class of helix-turn-helix proteins illustrate the generality of the helix-turn-helix motif. HSF is the only known trimeric DNA binding protein. Although histone H5 and HNF-3/*fork head* motif bind DNA as monomers, CAP and biotin repressor bind

Fig. 1. (**A**) Stereo view of sample $F_o - F_c$ (2.2σ) electron density at 1.8 Å for the refined model. A simulated annealing omit map was calculated with all the side chains from the hydrophobic core omitted. (**B**) Stereo view of the tertiary structure of the DNA binding domain of yeast HSF (*30*). Residues that are identical in all known HSFs are colored yellow, the most highly conserved residues within all HSFs are colored green, and other residues are colored blue. The secondary structure elements are α1 (residues 197 to 207), β1 (residues 216 to 217), β2 (residues 224 to 226), α2 (residues 229 to 240), α3 (residues 246 to 255), β3 (residues 258 to 259), and β4 (residues 277 to 281).

DNA as dimers. If we use biochemical evidence to estimate the number of base pairs bound by each monomer, HSF contacts 5 bp (*3*), HNF-3/*fork head* motif and CAP each contact around 14 bp, and biotin repressor appears to contact 18 bp (*25*). Histone H5 is thought to bind to DNA nonspecifically (*22*). We have used gel permutation assays to show that the DNA is not appreciably bent by trimers of HSF (*26*). The dimeric CAP and the monomeric HNF-3/*fork head* motif are both known to bend DNA, which accounts for their more extensive protein-DNA interfaces (*20, 23*). HSF, CAP, and HNF-3/*fork head* motif have most of their contacts with the major groove of DNA.

At this time, protein-DNA cocrystal structures of the α + β class of helix-turn-helix proteins are available for only HNF-3/*fork head* motif and CAP (*20, 23*). Like all other helix-turn-helix proteins, CAP and HNF-3/*fork head* motif use α3 as the major determinant for DNA recognition. These two proteins also use the NH_2-terminal ends of both α1 and α2 to make phosphate backbone contacts to DNA. On the basis of these studies, we suggest that α3 serves as the primary recognition helix of the DNA binding domain of HSF. This is supported by preliminary genetic evidence on *Saccharomyces cerevisiae* HSF (*27*). At present, we have no information on possible α1 and α2 DNA contacts, although it is likely that HSF uses α1 as HNF-3/*fork head* motif does. There is also the possibility that the extended loop is involved in binding to DNA. The analogous loop in the HNF-3/*fork head* motif cocrystal structure, part of the "winged helix" motif, contacts the DNA (*23*). Alternatively, the loop could be involved in protein-protein interactions between monomers.

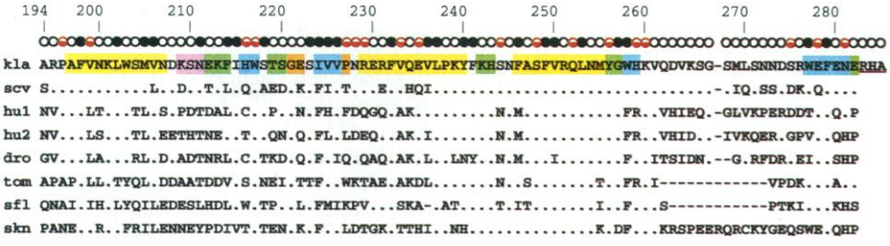


Fig. 2. Secondary structure and solvent accessibility of the *K. lactis* DNA binding domain. The numbering corresponds to the residue position within the *K. lactis* HSF gene. Line 1 shows the solvent accessibility for individual residues (*31*). White circles indicate more than 40% solvent-accessible surface; half-filled red circles indicate between 20 to 40% solvent accessibility; and black circles indicate residues with less than 20% solvent accessibility. Line 2 shows the sequence of the *K. lactis* HSF (kla) (*4*). The underlined residues at the COOH-terminus of the *K. lactis* sequence were added during the cloning procedure. Residues are shaded according to secondary structure (*32*): yellow, α helix; pink, 3_{10} helix; green, turn; blue, β sheet; and orange, bend. Sequences below the *K. lactis* sequence are scv, *S. cerevisiae* HSF amino acids 171 to 259; hu1, human HSF1 amino acids 14 to 102; hu2, human HSF2 amino acids 6 to 94; dro, *Drosophila melanogaster* amino acids 45 to 132; tom, tomato Lp-HSF24 amino acids 6 to 83; sfl, yeast suppressor gene for flocculation *SFL-1* amino acids 80 to 168; and skn, yeast suppressor gene for cell wall β-glucan assembly *SKN-7* amino acids 83 to 172 (*5, 14*). HSF sequences from mice, chickens, and fission yeast were omitted for clarity (*6*). Dots indicate identity to the *K. lactis* sequence; dashes indicate the absence of a residue. Single-letter amino acid abbreviations used 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.

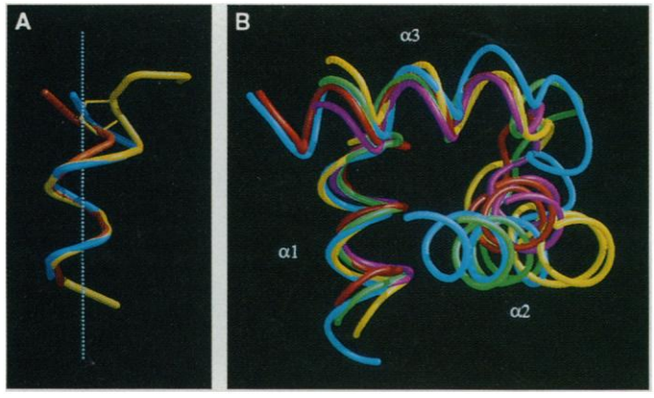


Fig. 3. (**A**) Superimposition of α-helical bulges (*30*). Residues 229 to 241 of HSF are shown in yellow; staphylococcal nuclease residues 122 to 130 are in blue (*17*); and a perfect α helix is shown in orange. Pro^{238} (side chain shown) causes a 29° kink. The helical axis for HSF residues 229 to 236 is shown in white. The inserted glycine in the staphylococcal nuclease structure is between residues 126 and 127. (**B**) Superimposition of the three α helices for the α + β helix-turn-helix proteins (*30*). We have omitted the β sheets from the figure because these sheets are in different orientations with respect to the α-helical bundles. CAP (red) and biotin repressor (magenta) have the best overlap for the α2 helix and have interhelical angles between α2 and α3 of 114° and 106°, respectively. Histone H5 (green) and HNF-3/*fork head* motif (blue) have displaced α2 helices and interhelical angles between α2 and α3 of 126° and 136°, respectively. The angle between α2 and α3 in HSF (yellow) is 98° when measured with the NH_2-terminal end of the α2 helix and 77° when measured with the COOH-terminal end of the α2 helix.

REFERENCES AND NOTES

1. P. K. Sorger, *Cell* **65**, 363 (1991); R. I. Morimoto, K. D. Sarge, K. Abravaya, *J. Biol. Chem.* **267**, 21987 (1992); R. I. Morimoto, *Science* **259**, 1409 (1993).
2. P. K. Sorger and H. C. Nelson, *Cell* **59**, 807 (1989).
3. O. Perisic, H. Xiao, J. T. Lis, *ibid.*, p. 797; H. Xiao, O. Perisic, J. T. Lis, *ibid.* **64**, 585 (1991). HSEs are best described as a series of inverted repeats of a 5-bp sequence, referred to as an nGAAn box. For example, a three-box site would have the sequence 5′-nGGAnnTTCnnGAAn-3′. A typical HSE contains between two and six nGAAn boxes, and a given heat shock promoter can have several HSEs.
4. B. K. Jakobsen and H. R. Pelham, *EMBO J.* **10**, 369 (1991).
5. P. K. Sorger and H. R. Pelham, *Cell* **54**, 855 (1988); G. Wiederrecht, D. Seto, C. S. Parker, *ibid.*, p. 841; J. Clos *et al.*, *ibid.* **63**, 1085 (1990); K. D. Scharf *et al.*, *EMBO J.* **9**, 4495 (1990); S. K. Rabindran, G. Giorgi, J. Clos, C. Wu, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6906 (1991); T. J. Schuetz, G. J. Gallo, L. Sheldon, P. Tempst, R. E. Kingston, *ibid.*, p. 6911.
6. K. D. Sarge, V. Zimarino, K. Holm, C. Wu, R. I. Morimoto, *Genes Dev.* **5**, 1902 (1991); A. Nakai and R. I. Morimoto, *Mol. Cell. Biol.* **13**, 1983 (1993); G. J. Gallo, H. Prentice, R. E. Kingston, *ibid.*, p. 749.
7. A fragment of the *Kluyveromyces lactis* HSF gene encoding amino acids 194 to 282 (*4*) was cloned into a derivative of the expression vector pET3-b [F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, *Methods Enzymol.* **185**, 60 (1989)]. The resulting plasmid, pHN212, adds an NH_2-terminal methionine and three amino acids (Arg-His-Ala) at the COOH-terminal end. This fragment of the *K. lactis* HSF is analogous to residues 171 to 259 in *S. cerevisiae* HSF.
8. K. Flick, B. Drees, H. Nelson, unpublished results.
9. The expression plasmid, pHN212, was transformed into BL21(DE3) cells, and the transformants were induced as described (*7*). Briefly, the protein was purified to homogeneity by a three-column protocol consisting of heparin–agarose ion exchange, sulfopropyl ion exchange, and reversed-phase chromatography. Analysis of typical batches of protein by electrospray-ionization mass spectrometry showed that the protein fragment, with an experimental and calculated molecular weight of 10,933 daltons, is missing the NH_2-terminal methionine (C. Harrison and D. King, unpublished results). A crystal form suitable for x-ray analysis was found with the sparse matrix method [J. Jancarik and S.-H. Kim, *J. Appl. Crystallogr.* **24**, 409 (1991)]. Crystals were grown by vapor diffusion at 8°C against 100 mM sodium acetate (pH 4.6), 200 mM ammonium acetate, and 25 to 30% PEG4000 (Fluka Chemie AG) at a final protein concentration of 15 mg/ml.
10. The expression strain BL21(DE3) is not auxotrophic for methionine. To incorporate the selenomethionine, we supplemented minimal media with L-selenomethionine (50 μg/ml), as well as with threonine, phenylalanine, leucine, isoleucine, lysine, and valine (100 μg/ml). The latter amino acids directly or indirectly inhibit methionine biosynthesis. The native protein contains three methionines, and electrospray-ionization mass spec-

trometry confirmed that there was greater than 95% incorporation of selenomethionine at the three positions. Difference Patterson maps revealed only two selenium sites, later determined to be at Met^{205} and Met^{255}. A difference Fourier using multiple isomorphous replacement phases failed to reveal the third site, which should have been at Met^{269}. Subsequent determination of the structure showed that there is poor electron density for amino acids 269 to 272, which is in part of the extended loop.
11. A. T. Brunger, *X-PLOR, Version 3.1: A System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
12. Electron density maps used during tracing were generated with the complete 2.7 Å resolution data set (Table 2). Solvent-flattening routines [B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985)] were applied to improve the electron density for initial tracing. Despite a low experimental figure of merit (m = 0.41), solvent-flattened maps were of reasonable quality, allowing all of the secondary structure to be traced. Three α helices and four β strands were traced with the program PSFRODO [T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978)] on an Evans and Sutherland PS390 Graphics System. Conservative partial models were subjected to positional refinement with X-PLOR 3.1 (*10*). Partial model combination with the program SIGMAA [R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986)] improved the electron density in the untraced regions of the map. Partial model combination was continued until the crystallographic R factor was 35% and R_T^{free} [as described by A. T. Brunger, *Nature* **355**, 472 (1992)] was 43%. At this point, the quality of the partial model map and the quality of the $2F_o - F_c$ map were similar, and partial model combination was discontinued. The 2.7 Å model consisted of 85 of the 92 residues; the side chains of 16 of these residues were not modeled in the electron density. For refinement of the model to 1.8 Å, we used the RAXIS native data set (Table 1). The initial overall B factor was set to 16 $Å^2$, a value estimated from a Wilson plot of the native data between 1.9 and 2.7 Å. X-PLOR 3.1 (*10*) was used for all of the refinement, with the use of the force field parameters developed by R. A. Engh and R. Huber [*Acta Crystallogr.* **A47**, 392 (1991)]. As the refinement progressed through iterative cycles of simulated annealing and model building, electron density, which had not been seen previously, appeared for the solvent-exposed side chains. The extended loop was the only structural element that required some refitting in the simulated annealing omit maps.
13. S. C. Harrison, *Nature* **353**, 715 (1991); C. O. Pabo and R. T Sauer, *Annu. Rev Biochem.* **61**, 1053 (1992), R. B. Brennan, *Cell* **74**, 773 (1993).
14. A. Fujita *et al.*, *Gene* **85**, 321 (1989); J. L Brown, S North, H. Bussey, *J. Bacteriol.* **175**, 6908 (1993).
15. D J. Barlow and J M. Thornton, *J. Mol. Biol.* **201**, 601 (1988).
16. Both SFL-1 and SKN-7 contain the corresponding phenylalanine and leucine. SFL-1 has three amino acids between these two residues, whereas SKN-7 has four amino acids.
17. N Schiering *et al*, *Nature* **352**, 168 (1991); A. Åberg, P. Nordlund, H. Eklund, *ibid.* **361**, 276 (1993).
18. L. J. Keefe, J. Sondek, D. Shortle, E. E. Lattman, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3275 (1993).
19. D. W. Heinz, W. A. Baase, F. W. Dahlquist, B. W. Matthews, *Nature* **361**, 561 (1993); D W. Heinz and B. W. Matthews, *Protein Eng.*, in press.
20. S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **253**, 1001 (1991).
21. K. P. Wilson, L. M. Shewchuk, R. G. Brennan, A J. Otsuka, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A* **89**, 9257 (1992).
22. V. Ramakrishnan, J. T. Finch, V. Graziano, P L. Lee, R. M Sweet, *Nature* **362**, 219 (1993).
23. K. L. Clark, E. D. Halay, E. Lai, S. K. Burley, *ibid.* **364**, 412 (1993); E. Lai, K. L. Clark, S. K. Burley, J. E Darnell Jr., *Proc Natl. Acad. Sci. U.S.A.* **90**, 10421 (1993).
24. We differentiate this variation from that seen in the Oct-1 POU-specific domain and HNF-1 proteins, which are members of the all α-helical class of helix-turn-helix motifs. They both have a longer turn between the two helices of the motif, but they still have the same relative positions of α2 and α3 as the canonical motif from the all α-helical class [N. Assa-Munt, R. J. Mortishire-Smith, R. Aurora, W. Herr, P. E. Wright, *Cell* **73**, 193 (1993); B. Leiting *et al.*, *EMBO J.* **12**, 1797 (1993); T. A. Ceska *et al.*, *ibid.*, p. 1805].
25. R. H. Costa, D. R. Grayson, J. Darnell Jr., *Mol. Cell. Biol.* **9**, 1415 (1989); H. N. Liu-Johnson, M. R. Gartenberg, D. M. Crothers, *Cell* **47**, 995 (1986); J. Abott and D. Becket, *Biochemistry* **32**, 9649 (1993).
26. H. S. Rye, B. L. Drees, H. C. M. Nelson, A. N. Glazer, *J. Biol. Chem.* **268**, 25229 (1993).
27. S. Hubl and H. Nelson, unpublished results.
28. T. C. Terwilliger, S.-H. Kim, D. Eisenberg, *Acta Crystallogr.* **A43**, 1 (1987).
29. W. Steigemann, *PROTEIN: A Package of Crystallography Programs for Analysis of Proteins* (Max Planck Institute for Biochemistry, Martinsried, Germany, 1982).
30. Molecular graphics images for Figs. 1B, 3A, and 3B were produced with the MidasPlus software system from the Computer Graphics Laboratory, University of California, San Francisco [T. E. Ferrin *et al.*, *J. Mol. Graphics* **6**, 13 (1988)].
31. Solvent accessibility for individual residues was normalized to accessible surface area in a fully exposed residue in a Gly-X-Gly peptide [C. Chothia, *Nature* **253**, 304 (1975)], where X is any amino acid.
32. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
33. We thank A. de Vos for help, advice, and time on the FAST at Genentech; D. King for advice on purification and for electrospray-ionization mass spectrometry; R. Standeart for advice on selenomethionine incorporation; R. Stevens, J. Somoza, and T. Earnest for help and advice on crystallography; D. Heinz and S. Burley for sending us preprints of their manuscripts; and all of the members of the Nelson laboratory. This work was partially supported by NIH grant GM44086 (H.C.M.N.) and an award from the Pew Scholars Program in Biomedical Sciences (H.C.M.N.). Instrumentation was funded in part by funds provided from the Lucille P. Markey Charitable Trust through the Berkeley campus Program in Biomolecular Structure and Design. C.J.H. was supported in part by an NSF Graduate Research Fellowship and by U.S. Public Health Service predoctoral training grant GM08295. A.A.B. was supported by the director, Office of Energy Research, Office of Health Effects, of the U.S. Department of Energy under contract DE-AC03-76SF00098.

16 September 1993; accepted 7 December 1993

inhA, a Gene Encoding a Target for Isoniazid and Ethionamide in *Mycobacterium tuberculosis*

Asesh Banerjee, Eugenie Dubnau, Annaik Quemard, V. Balasubramanian, Kyung Sun Um, Theresa Wilson, Des Collins, Geoffrey de Lisle, William R. Jacobs Jr.*

Isoniazid (isonicotinic acid hydrazide, INH) is one of the most widely used antituberculosis drugs, yet its precise target of action on *Mycobacterium tuberculosis* is unknown. A missense mutation within the mycobacterial *inhA* gene was shown to confer resistance to both INH and ethionamide (ETH) in *M. smegmatis* and in *M. bovis*. The wild-type *inhA* gene also conferred INH and ETH resistance when transferred on a multicopy plasmid vector to *M. smegmatis* and *M. bovis* BCG. The InhA protein shows significant sequence conservation with the *Escherichia coli* enzyme EnvM, and cell-free assays indicate that it may be involved in mycolic acid biosynthesis. These results suggest that InhA is likely a primary target of action for INH and ETH.

Despite the availability of effective chemotherapies, tuberculosis is responsible for one in four avoidable adult deaths in developing countries (*1*). Infection with drug-sensitive strains of *M. tuberculosis* can be effectively cured with a combination of INH, rifampicin, and pyrazinamide (*2*). However, the particular susceptibility and increased mortality of the disease among individuals infected with human immunodeficiency virus (HIV) pose a serious threat to tuberculosis control programs (*3*). Moreover, the emergence of multidrug-resistant strains of *M. tuberculosis* (MDR-TB) has resulted in fatal outbreaks in many countries, including the United States (*4*). Strains of MDR-TB, some of which are resistant to as many as seven drugs, are deadly to both HIV^- and HIV^+ individuals (*5*).

INH was first reported to be an effective antituberculosis drug in 1952, displaying particular potency against *M. tuberculosis* and *M. bovis* (*6*). Mutants resistant to INH have emerged since then (*7*), and today such mutants account for as many as 26% of the clinical *M. tuberculosis* isolates in certain U.S. cities (*5*). Some INH-resistant strains are associated with a loss of catalase activity (*8*), and deletions of the catalase-peroxidase gene (*katG*) correlate with INH resistance in certain *M. tuberculosis* isolates (*9*). Furthermore, transfer of the wild-type

A. Banerjee, E. Dubnau, A. Quemard, V. Balasubramanian, K. S. Um, W. R. Jacobs Jr., Howard Hughes Medical Institute, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461.
T. Wilson, D. Collins, G. de Lisle, AgResearch, Wallaceville Animal Research Center, Upper Hutt, New Zealand.

*To whom correspondence should be addressed.