Arginine Residues as Stabilizing Elements in Proteins†,‡

Nadir T. Mrabet,*,§,×,‖ Annemie Van den Broeck,§,◇ Ilse Van den brande,§,◇ Patrick Stanssens,§,◇ Yves Laroche,§,◇ Anne-Marie Lambeir,§,◇ Gaston Matthijssens,§,◇ John Jenkins,⊥,# Mohamed Chiadmi,⊥,△ Herman van Tilbeurgh,§,⊥,○ Felix Rey,⊥,▽ Joël Janin,⊥ Wim J. Quax,⊖ Ignace Lasters,∞,◇ Marc De Maeyer,∞,◇ and Shoshana J. Wodak*,×

Protein Engineering Department, Plant Genetic Systems, 22 Jozef Plateaustraat, B-9000 Gent, Belgium, Laboratoire de Biologie Physicochimique, Université Paris-Sud, F-91405 Orsay, France, Research and Development, Gist-Brocades, Wateringseweg 1, NL-2611 XT Delft, The Netherlands, and Plant Genetic Systems and Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles, Avenue Paul Héger, CP160 P2, B-1050 Brussels, Belgium

Received February 25, 1991; Revised Manuscript Received October 16, 1991

ABSTRACT: Site-specific substitutions of arginine for lysine in the thermostable D-xylose isomerase (XI) from *Actinoplanes missouriensis* are shown to impart significant heat stability enhancement in the presence of sugar substrates most probably by interfering with nonenzymatic glycation. The same substitutions are also found to increase heat stability in the absence of any sugar derivatives, where a mechanism based on prevention of glycation can no longer be invoked. This rather conservative substitution is moreover shown to improve thermostability in two other structurally unrelated proteins, human copper, zinc-superoxide dismutase (CuZnSOD) and D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *Bacillus subtilis*. The stabilizing effect of Lys → Arg substitutions is rationalized on the basis of a detailed analysis of the crystal structures of wild-type XI and of engineered variants with Lys → Arg substitution at four distinct locations, residues 253, 309, 319, and 323. Molecular model building analysis of the structures of wild-type and mutant CuZnSOD (K9R) and GAPDH (G281K and G281R) is used to explain the observed stability enhancement in these proteins. In addition to demonstrating that even thermostable proteins can lend themselves to further stability improvement, our findings provide direct evidence that arginine residues are important stabilizing elements in proteins. Moreover, the stabilizing role of electrostatic interactions, particularly between subunits in oligomeric proteins, is documented.

Enhancing protein stability by rational design has been one of the great challenges of protein engineering. Site-directed mutagenesis combined with X-ray diffraction studies as well as denaturation experiments has already been very useful in determining the contributions of specific amino acids to protein stability (Alber, 1989). From these studies several tentative rules have emerged. The contribution of buried residues to protein stability is correlated to their hydrophobicity as derived from their free energy of transfer (Matsumura et al., 1988; Kellis et al., 1988). Protein stability is more strongly affected by residues in the well-packed protein interior than by those on the surface (Alber et al., 1987a; Matsumura et al., 1988; Kellis et al., 1988, 1989; Sandberg & Terwilliger, 1989). Substitutions such as X → Pro and Gly → X are found to be stabilizing, an effect attributed to the reduction in conformational entropy of the protein in the unfolded state (Matthews et al., 1987). On the basis of similar entropic arguments, addition of disulfide bridges and other cross-links is believed and shown to be stabilizing, with documented exceptions however [reviewed by Wetzel (1987) and Creighton (1988)]. Substantial contributions to protein stability are likewise provided by electrostatic interactions. These include salt bridges (Perutz, 1978), hydrogen bonds (Wilkinson et al., 1983; Alber et al., 1987b), and interactions of charged groups with the so-called "α-helix dipole" (Mitchinson & Baldwin, 1986; Sali et al., 1988; Nicholson et al., 1988) and of polar amino acid side chains with α-helix ends (Serrano & Fersht, 1989).

In analyzing protein stability, irreversible processes that trap proteins in the denatured state must be considered along with reversible ones. Protein inactivation resulting from enzymatic or chemical modifications (Ahren & Klibanov, 1985; Harding, 1985) has been studied, and ways to prevent it by specific amino acid substitutions have been suggested (Brot & Weissenbach, 1983; Rosenberg et al., 1984; Estell et al., 1985; Casal et al., 1987).

This study describes an effective approach to enhancing the heat stability of D-xylose isomerase (XI;[1] D-xylose ketol-

† This work was performed as part of a protein engineering project also supported by Amylum NV, Aalst, Belgium.

‡ The coordinates of the three D-xylose isomerase structures discussed here have been deposited in the Brookhaven Protein Data Bank with the entry names 1XIM for wild type, 2XIM for K253R, and 3XIM for the K309R/K319R/K323R mutant.

* Address correspondence to either author.

§ Protein Engineering Department, Plant Genetic Systems.

× Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles.

‖ Present address: LEGG/URA-CNRS 457, Université de Nancy I, 54506 Nancy, France.

◇ Present address: Corvas International, Gent, Belgium.

⊥ Laboratoire de Biologie Physicochimique, Universite Paris-Sud.

Present address: AFRC Institute of Food Research, Shinfield, Reading RG2 9AT, England.

△ Present address: ABCx and Laboratoire de Biologie Physicochimique, Université Paris-Sud, 91405 Orsay, France.

○ Present address: LCCMB, Faculté de Médecine Nord, Marseilles, France.

▽ Present address: Gibbs Laboratories, Howard Hughes Medical Center, Harvard University, Cambridge, MA 01238.

⊖ Research and Development, Gist-Brocades.

∞ Plant Genetic Systems, Université Libre de Bruxelles.

[1] Abbreviations: XI, D-xylose isomerase; CuZnSOD, copper, zinc-superoxide dismutase; GAPDH, D-glyceraldehyde-3-phosphate dehydrogenase; SDH, D-sorbitol dehydrogenase; TIM, triose-phosphate isomerase; MOPS, 4-morpholinepropanesulfonic acid; EPPS, 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid; MES, 4-morpholineethanesulfonic acid; TEA, triethanolamine; EDTA, ethylenediaminetetraacetic acid; EGTA, [ethylenebis(oxyethylenenitrilo)]tetraacetic acid; DTT, dithiothreitol; bp, base pair(s); LB, Luria broth; IPTG, isopropyl β-D-thiogalactopyranoside; PBS, phosphate-buffered saline; Tris, tris(hydroxymethyl)aminomethane; kDa, kilodaltons; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

0006-2960/92/0431-2239$03.00/0

isomerase; EC 5.3.1.5) from *Actinoplanes missouriensis* by site-directed mutagenesis. D-Xylose isomerases are enzymes which convert xylose into xylulose. Due to their ability to also use glucose as a substrate to catalyze its isomerization into the sweeter fructose, they have been widely used in the food industry for the manufacture of high-fructose syrups. This process is carried out in industrial bioreactors at 60–65 °C where the half-life of the enzyme is of the order of several weeks. In addition to being profitable to industry, enhancing XI stability is also scientifically challenging. Indeed, protein engineering efforts aimed at increasing thermal stability have generally dealt with small single-chain proteins exhibiting reversible denaturation. In contrast, XI, an already thermostable enzyme, is large (M_r 172 000), is tetrameric, and undergoes irreversible thermal denaturation. In this context, analyzing and rationalizing protein stability properties is therefore significantly more complex.

In devising approaches to improve the stability of XI, processes of irreversible thermal denaturation have been considered in priority. Relevant pathways to analyze involve chemical modifications by sugar components of the high-fructose syrups and, in particular, the nonenzymatic glycation (Bookchin & Gallop, 1968; Bunn et al., 1978; Brownlee et al., 1984; Furth, 1988) of lysine residues.

On the basis of the heat stability analysis and the crystal structure of wild-type XI, selected lysine residues have been replaced with either arginine or glutamine. The mutant proteins were purified, and their heat stability was measured in the presence and in the absence of glucose. The three-dimensional structures of the arginine-containing mutants were also determined at high resolution by X-ray diffraction.

To further assess the conclusions drawn from these studies, specific lysines have been substituted with arginines in two other, structurally unrelated proteins, human CuZnSOD (EC 1.15.1.1) and *Bacillus subtilis* GAPDH (EC 1.2.1.12), and their effect on heat stability has been analyzed.

Detailed biochemical, structural, and model building analyses of the three wild-type enzymes and their engineered variants provide evidence that arginine is not only effective in preventing glycation but also confers intrinsic stability. The general implication that arginine residues may be stabilizing elements in proteins is discussed.

Experimental Procedures

Materials

All reagents, except where indicated, were of the highest grade available from Merck. Cacodylic acid was from Serva. MOPS and EPPS were from Janssen Pharmaceutica. Other chemicals and their sources were as follows: NADH, SDH, and DNA polymerase (Klenow fragment), Boehringer Mannheim; T4 DNA ligase and [α-^{35}S]dATPαS, Amersham; polynucleotide kinase, New England Biolabs; modified T7 DNA polymerase (Sequenase), U.S. Biochemical Corp.; restriction endonucleases and nucleoside triphosphates, Pharmacia; IPTG, Research Organics Inc. Buffers were prepared with fresh Milli-Q water from the acid forms (except for TEA and Tris) and by adding a calculated amount of a stock solution of NaOH (10 N; stored in the presence of Chelex-100) to give the desired pH at a given temperature, as indicated. Buffers and incubation mixtures were all filter-sterilized after preparation and/or again immediately before use through 0.2-μm Millex filters (Millipore) rinsed with the same buffers. Chelex-100 was from Bio-Rad and was used in the column mode as recommended by the manufacturer to demetalate buffers where indicated; after demetalation, buffers were stored at 4 °C in the presence of small amounts of Chelex-100 and again filter-sterilized before use as described above.

Methods

Oligonucleotide-Directed Mutagenesis. Construction of mutations was carried out by oligonucleotide-directed mutagenesis according to the gapped-duplex DNA method using the pMa/pMc phasmid vector system (Stanssens et al., 1989). Nucleotide sequence analysis made use of the dideoxy chain-termination method (Sanger et al., 1987).

Overproduction of Recombinant Proteins. All DNA manipulations followed standard procedures (Maniatis et al., 1982).

(*i*) *Cloning and Expression of Xylose Isomerase.* Isolation, cloning, and expression of the *A. missouriensis* (DSM43046) XI gene by complementation of the xylose isomerase deficient *Escherichia coli* strain AB1886 (Howard-Flanders et al., 1966) will be described elsewhere.[2] Wild-type and mutant enzymes were produced in soluble form in *E. coli* strain K527, a xylose isomerase negative derivative of *E. coli* K514 (Colson et al., 1965) harboring the XI-overproducing plasmid pMa5-I, which also specifies resistance to ampicillin (Stanssens et al., 1989). Cells were grown overnight at 37 °C in a medium composed of 1% tryptone, 1% NaCl, 0.5% yeast extract, and 100 mg/mL ampicillin and centrifuged. The cell pellet was resuspended in a minimal volume of 50 mM Tris-HCl, 0.1 mM $CoCl_2$, 10 mM $MgCl_2$, 200 mM KCl, 5% glycerol, and 5 mM EDTA, pH 8.0, and lysozyme was added to a final concentration of 1 mg/mL. After being allowed to stand for 20 min at 0 °C, the cells were lysed using a French press and centrifuged at 23000 *g* for 30 min, and the resulting supernatant was diluted with an equal volume of 5% streptomycin sulfate. Incubation was maintained for 3 h at 4 °C and was followed by centrifugation (30 min at 23000 *g*). The resulting supernatant has a usual protein concentration of 18–22 mg/mL as determined with the Bio-Rad protein assay (Bradford, 1976) using bovine serum albumin as a standard.

(*ii*) *Cloning and Expression of the Human Copper, Zinc-Superoxide Dismutase.* The plasmid pSOD11 which directs the overexpression of the human CuZnSOD gene (designated *sod*) has been described (Hallewell et al., 1985). For purposes of mutagenesis, the *tac* promoter–*sod* expression cassette was excised from pSOD11 (a 1260-bp fragment running from an *Acc*I site, made blunt-ended by treatment with Klenow, to the unique *Sal*I site) and inserted into both pMa5-8 and pMc5-8 (between the *Sma*I and *Sal*I sites of the multicloning site). The resultant plasmids have been designated pMa5-sod11 and pMc5-sod11, respectively. The *E. coli* WK6 *lacI*^{q} strain (Zell & Fritz, 1987) was used as the host for production of wild-type and mutant forms of CuZnSOD. Cells were grown at 30 °C in TB medium (Tartof & Hobbs, 1988) supplemented with 100 ppm of $CuSO_4$, 0.46 ppm of $ZnSO_4$, and the appropriate antibiotic, to an OD_{650nm} = 0.5–1.0, and expression of the *sod* gene was induced by addition of 0.1 mM IPTG. After an induction period of 16 h, the cells (final OD_{650nm} = 15–25) were harvested by centrifugation. Under these conditions, efficient synthesis of a soluble 16-kDa protein is achieved as shown by SDS–PAGE of total cellular extracts prepared by sonication.

(*iii*) *Cloning and Expression of* D-*Glyceraldehyde-3-phosphate Dehydrogenase from B. subtilis.* The plasmid pBSgap-1 (Viaene & Dhaese, 1989), which consists of a partial *Sau*3A fragment containing the entire GAPDH coding region (designated *gap*) from *B. subtilis* inserted in the *Bam*HI site

[2] P. Stanssens et al., manuscript in preparation.

of pUC18, was a gift of Dr. Patrick Dhaese (Rijkuniversiteit, Gent, Belgium). High-level production of GAPDH was obtained by putting the gene under the transcriptional control of the *tac* promoter. The expression vector pMc5-gap was obtained by ligation of the following three gel-purified fragments: (i) a *tac* promoter fragment, obtained by opening pMT416 (Hartley, 1988) with *Bam*HI, treatment with Klenow, and subsequent digestion with *Eco*RI; (ii) the 1255-bp *Dra*I–*Xba*I fragment of pBSgap-1 where the first restriction site is located 16 nucleotides upstream from the coding region, whereas the second site is found in the pUC18 polylinker; and (iii) the large *Eco*RI–*Xba*I fragment of pMc5-8. The complementary pMa5-gap1 vector was constructed by transferring the *Eco*RI–*Xba*I *tac–gap* module onto pMa5-8. For production of wild-type and mutant forms of GAPDH, WK6 transformants were grown at 37 °C in TB medium and induced as described for CuZnSOD. The recombinant GAPDHs were present in the soluble fraction and accumulate up to at least 30% of the total protein content.

Purification of Recombinant Proteins and Enzymatic Assays. Protein concentrations were determined with the Bio-Rad protein assay (Bradford, 1976) using bovine serum albumin as a standard.

(*i*) *Xylose Isomerase.* Purification of the wild-type and mutant forms of XI and preparation of the apoenzymes will be described in detail elsewhere.[3] In brief, the purification protocol successively involved heating for 30 min at 70 °C [but only 50 °C for low-stability mutants such as K253Q (Lys-253 → Gln, described below)] followed by centrifugation; ammonium sulfate fractionation of the resulting supernatant; phenyl-Superose chromatography with reverse ammonium sulfate gradient; gel permeation on Sephacryl S-200 HR; Mono-Q anion-exchange chromatography with a NaCl gradient from 0 to 0.6 M; dialysis against 10 mM TEA, pH 7.2, made 10 mM EDTA (final pH is about 6); and finally dialysis against 5 mM MES, pH 6.0. Importantly, ammonium sulfate solutions and all chromatographic buffers were made 10 mM EDTA to ensure purification of xylose isomerase in the *apoenzyme* form. Metal contamination can be assessed by monitoring the level of residual cobalt, a metal ion which binds to XI with high affinity.[4] Absence of divalent metals was systematically assessed by electrothermal atomic absorption spectrometry on a Varian SpectrAA 30/40 Zeeman analyzer. Under the conditions described above, residual cobalt is found not to exceed 5×10^{-4} atom/$mol_{tetramer}$.

Unless otherwise mentioned enzymatic activity using D-xylose as substrate was measured by means of a modification of the SDH (L-iditol:NAD oxidoreductase, EC 1.1.1.14) assay (Kersters-Hilderson et al., 1987). Briefly, the rate of enzyme-catalyzed formation of D-xylulose at 35 °C was measured by following the kinetics of oxidation of NADH (0.15 mM) by SDH (≃0.02 μM) in 50 mM TEA, pH 7.5, 10 mM $MgSO_4$, 1 mM EGTA, and 100 mM xylose. The final XI concentration in this assay was ≃2.5 μg/mL and was precisely determined.

Enzymatic activity for D-glucose at 60 °C was determined using the SDH assay as described by Callens et al. (1986).

(*ii*) *CuZnSOD.* Lysozyme (0.5 mg/mL) was added to the cell suspension which was then made 0.1 mM EDTA. After 20 min on ice, the cells were lysed using a French press, and the resulting homogenate was clarified by centrifugation. Streptomycin sulfate (0.5%) was added to the supernatant, and after 3 h at 4 °C, the insoluble material was removed by centrifugation. The resulting supernatant was then made 0.1 mM $CuSO_4$ and 0.05 mM $ZnSO_4$ and incubated 2–4 h at 25 °C before being heated to 70 °C for 30 min. After elimination of the insoluble material by centrifugation, the clarified lysate was dialyzed twice at 4 °C against 50 volumes of 10 mM Tris-HCl, pH 8.0, and chromatographed onto Q-Sepharose Fast Flow (Pharmacia) by means of a gradient from 40 to 200 mM CH_3COONa in 20 mM Tris-HCl, pH 8.0. The CuZnSOD-containing fractions, as assessed by SDS–PAGE and by native PAGE with Coomassie staining and with activity staining (Beauchamp & Fridovich, 1971), were pooled and made 1.4 M ammonium sulfate and chromatographed onto a Pharmacia phenyl-Superose HR 10/10 column equilibrated in 20 mM Tris-HCl and 1.4 M $(NH_4)_2SO_4$, pH 8.0. Under these conditions, CuZnSOD is not retained and elutes in the void volume. The latter was dialyzed at 4 °C against 50 volumes of 10 mM Tris-HCl, pH 8.0, with two buffer changes. The enzyme was then concentrated on a Mono-Q HR 5/10 run in 20 mM Tris-HCl, pH 8.0, including 2 μM $CuSO_4$ and 2 μM $ZnSO_4$, and eluted with a gradient from 0 to 160 mM CH_3COONa. The resulting enzyme was finally dialyzed twice against 10 mM potassium phosphate, pH 7.8, and shown to be homogeneously pure by SDS–PAGE. Enzymatic activity was determined using the pyrogallol assay (Marklund & Marklund, 1974) with minor modifications. Briefly, the inhibition of pyrogallol autoxidation by CuZnSOD was determined as follows. An assay mixture containing 0.90 mL of 52.7 mM Tris/cacodylic acid, 1.1 mM diethylenetriaminepentaacetic acid, pH 8.20 (TCDA buffer), and 0.050 mL of serial dilutions of CuZnSOD in TCDA buffer containing 1 mg/mL catalase (*Aspergillus niger*) was used. Following incubation at 25 °C for 10 min in a thermostated water bath, the mixture was vortexed at very high speed to achieve air equilibration immediately before the reaction was started by addition of 0.050 mL of 4.0 mM pyrogallol in 1 mM cacodylic acid. The stock solution of pyrogallol in 1 mM cacodylic acid was stable for 1 day when stored on ice in the dark. The increase in absorbance at 420 nm as a function of time, which measures the rate of autoxidation of pyrogallol, was determined at 25 °C over a period of 6 min during which time linearity was observed. The inhibition of pyrogallol oxidation as a function of added CuZnSOD was found to be linear up to an enzyme concentration of 250–300 ng/mL. Specific activities were 5000 and 5200 units/mg of protein for wild-type CuZnSOD and mutant K9R, respectively, where one unit represents the amount of enzyme necessary to achieve 50% inhibition.

(*iii*) *GAPDH.* WK6 *E. coli* cells harboring the *gap* gene were harvested by centrifugation after an overnight culture at 37 °C, washed with PBS, frozen and then thawed, and resuspended in 0.1 M TEA, pH 7.5, containing 2 mM DTT and 5 mM EDTA. Lysozyme was added (1 mg/mL), and after 20 min on ice, the cells were lysed using a French press. Cell debris was removed by centrifugation, and streptomycin sulfate was added to a final concentration of 0.5%. After 3 h at 4 °C, insoluble material was removed by centrifugation, and the supernatant was made 1 mM NAD^+. Partial (≃90%) purification of recombinant GAPDH was achieved by ammonium sulfate (75%) precipitation at 4 °C. About 95% of the GAPDH activity remained in the soluble fraction. Insoluble material was removed by centrifugation. The super-

[3] N. T. Mrabet, manuscript in preparation; P. Stanssens et al., manuscript in preparation.

[4] J. Snauwaert, unpublished experiments. The high-affinity constant measured by equilibrium fluorescence spectroscopy is 6.9×10^5 M^{-1}. A second binding cobalt binding site is also observed, with a lower affinity constant (2.2×10^3 M^{-1}).

natant was made 100% ammonium sulfate and stored at −20 °C as such. Immediately before use, the protein was thawed on ice, then collected by centrifugation, dissolved in 1 mM DTT, 1 mM EDTA, and 1 mM NAD^+ in 0.1 M TEA, pH 7.8 (TDNE buffer), and finally desalted by passage through a PD10 column equilibrated in the same buffer. The purified enzyme was shown to behave identically with authentic GAPDH isolated from *B. subtilis* strain BR151 on both native and SDS–PAGE. Enzymatic activity measurements were performed as previously described (Misset et al., 1986).

Heat Inactivation Assays. (*i*) *Xylose Isomerase.* Heat inactivation experiments were performed using the apoenzymes at a concentration of 1 mg/mL without or with added divalent metal as indicated. After equilibration in Chelex-treated MOPS (50 mM, pH 7.2 at 25 °C; $\Delta pH/\Delta t = -0.011$) in the absence or presence of 10 mM $MgSO_4$ or $CoCl_2$, the enzyme solution was drawn up into a 1-mL Hamilton gas-tight syringe mounted with a 4-cm-long Teflon needle (Hamilton; internal diameter = 0.46 mm). The syringe body was snugly fitted into a glass mantle connected to a Lauda RM6 circulating water bath set at the desired temperature. Under such conditions, temperature equilibration of the enzyme solution from 25 to 85 °C was achieved in about 1 min. Aliquots were withdrawn at appropriate times (after discarding the solution residing in the Teflon needle) and immediately cooled to 0 °C before being assayed for residual activity by means of the SDH assay. The decay was found to obey first-order kinetics; the data (up to 15 time points for a 5-h incubation) were fitted satisfactorily ($r^2 > 0.95$) to monoexponential decay using the GraphPad (San Diego, CA) software.

(*ii*) *CuZnSOD.* After equilibration for 18 h at 4 °C in 20 mM potassium phosphate, pH 7.8, in the presence of 25 μM $CuSO_4$ and 25 μM $ZnSO_4$, wild-type CuZnSOD and mutant K9R (Lys-9 → Arg) were diluted to 0.2 mg/mL in the same buffer and brought to 85 °C for the indicated time intervals (Figure 8) using the experimental arrangement described for xylose isomerase. Heat inactivation was stopped by cooling the samples to 0 °C by adding 50 volumes of ice-cold TCDA buffer including 1 mg/mL catalase. Residual activity was measured with the pyrogallol assay described above.

(*iii*) *GADPH.* The temperature dependence for heat inactivation of wild-type Bsu-GAPDH and mutants G281K (Gly-281 → Lys) and G281R (Gly-281 → Arg) was determined using enzyme mixtures containing 40 units/mL GAPDH in TDNE buffer and incubated for 15 min at the temperatures indicated. The inactivation process was stopped by adding 9 volumes of ice-cold solution containing 1 mg/mL bovine serum albumin in TDNE buffer, and residual activity was determined as described (Misset et al., 1986). To follow the kinetics of inactivation at 75 °C of mutants G281K and G281R, residual activity as a function of time was determined under identical conditions, by means of the experimental arrangement described for xylose isomerase.

Nonenzymatic Glycation of Xylose Isomerase. The nonenzymatic glycation experiments were performed on apoenzymes, made free of residual EDTA by extensive buffer exchange in 50 mM MES, pH 5.5, using Amicon centrifugal-flow cell concentrators, Centricon PM-30, and subsequently equilibrated in 12.5 mM potassium phosphate buffer, pH 7.7, by passage through a PD-10 column (Pharmacia). α-D-Glucose prepared in the same buffer was then added to a final concentration of 250 mM. To eliminate the effect of fructose that can be produced by enzymatic isomerization of glucose, the mixtures were prepared in the absence of metal. The final protein concentration was 1 mg/mL. Incubation was performed at 60 °C in a Lauda RM6 thermostated water bath. Aliquots of the mixture were withdrawn at appropriate times and immediately cooled to 0 °C. To circumvent sugar interference with the SDH assay described above, residual XI activity was determined by measuring the increase in absorbance at 278 nm due to the production of xylulose at 35 °C in 50 mM TEA, pH 7.5, 10 mM $MgSO_4$, and 1 mM EGTA, in the presence of 100 mM xylose, using an extinction coefficient, $\epsilon_{278nm} = 10.3\ M^{-1}\ cm^{-1}$, for xylulose. The final XI concentration in the assay was ≃1 μg/mL and was precisely determined. Half-lives were computed by fitting the inactivation kinetics data to monoexponential decay as described above.

X-ray Crystallography of Xylose Isomerase. A detailed description of the X-ray crystallographic procedures is to be published elsewhere.[5] In brief, purified protein aliquots (25 mg/mL) were made 0.1 M sodium phosphate, pH 7.1, and 1 M ammonium sulfate and placed in hanging drops over pits containing 1.2 M ammonium sulfate. Crystals were grown in a period of weeks at 18 °C. Metal ions and ligands were added by soaking the crystals in solutions of 0.1 M MOPS, pH 6.9, and 1.7 M ammonium sulfate, containing the desired metal salt and ligand, for a few hours before mounting in glass capillaries. Diffraction data were collected on films with synchrotron radiation at LURE-DCI (Orsay, France), using monochromatic radiation near 1.4 Å and a rotation camera. Crystals were cooled to 4 °C. Each pack represented a 0.8° oscillation about the *c* axis. Digitized films were processed with program DENZO (Otwinowsky, Yale University, New Haven, CT) and the CCP4 package (CCP4, Daresbury, England). Crystals of wild-type XI complexed with xylitol (see below) belong to trigonal space group $P3_221$ with cell dimensions $a = b = 143.45$ Å and $c = 231.5$ Å and have one complete tetramer per asymmetric unit. The crystals of the mutants described here are isomorphous, with cell parameters very close to those described for wild-type *A. missouriensis* XI.

(*i*) *Wild-Type Xylose Isomerase Structure.* The crystal structure of unligated *A. missouriensis* XI was previously solved at 2.8-Å resolution (Rey et al., 1988). A model built from a 2.8-Å electron density map phased with a single isomorphous derivative and noncrystallographic symmetry was used as the starting point for crystallographic refinement of the structure of the wild-type enzyme complexed with xylitol and Co^{2+} for which 2.2-Å data were available, using the Konnert–Hendrickson least-squares refinement procedure (Hendrickson, 1985). Noncrystallographic symmetry restraints were applied during refinement along with standard stereochemical restraints. These restraints, implemented in the PROLSQ refinement program by Leslie (1987), keep equivalent atoms of the four subunits near positions related by the 222 symmetry of the tetramer, which is not a symmetry of the crystal. Refinement proceeded automatically until the *R* factor reached 0.23, at which stage minor errors in the model became apparent in the electron density map. These were corrected with FRODO (Jones, 1985), and further refinement was applied by progressively adding solvent molecules and removing noncrystallographic symmetry restraints. The final model of the wild-type enzyme/xylitol complex has a *R* factor of 0.152 for 113 145 independent reflections (93% of observed and 85% of expected) and excellent stereochemistry (Table I). Analysis of the results by the method of Luzatti indicates that the rms error in atomic positions is less than 0.35 Å. In

[5] J. Jenkins, manuscript in preparation.

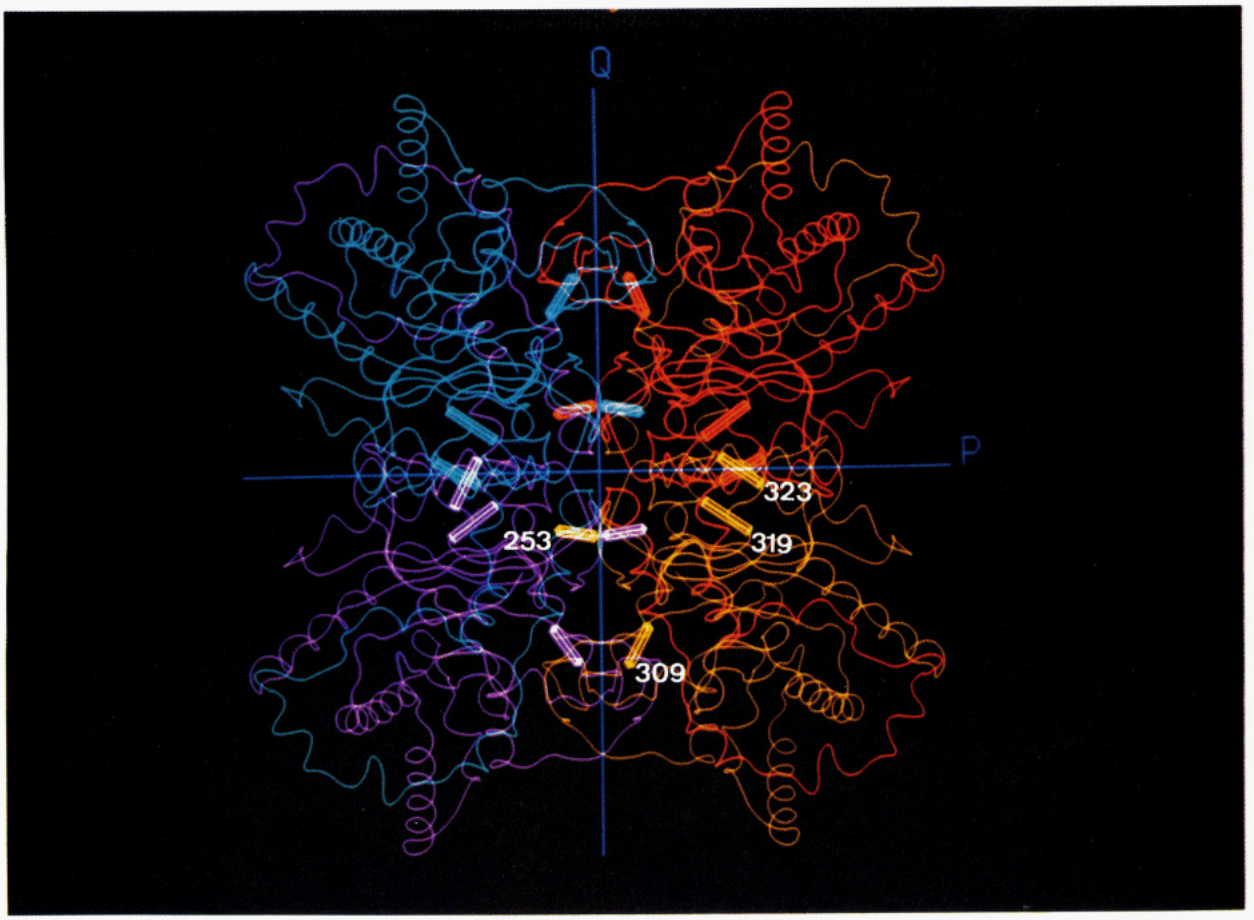

FIGURE 1: Ribbon representation of the D-xylose isomerase tetramer. Each subunit is colored differently. The 2-fold symmetry axes, labeled P and Q, are drawn in blue; the R axis is perpendicular to the plane of projection. The lysines subjected to mutagenesis are represented by tiny cylinders drawn around the line joining the C_α to the N_ζ atoms. For clarity the lysines are labeled only in the yellow subunit.

the crystal, the 222 symmetry of the tetramer is almost exact even though it is not crystallographic. This is verified by superimposing the C_α tracings from all four subunits: the rms distance of equivalent C_α atoms from the mean position is only 0.12 Å. For all atoms, it is 0.23 Å. Due to redundant information, the average structure is known to better than 0.2 Å.

Each of the polypeptide chains folds into a typical TIM barrel comprising residues 2–328 which consists of eight parallel β-strands connected by as many α-helices. The remainder of the polypeptide chain (residues 329–394) forms a large loop composed of three α-helices which embraces the adjacent subunit within a *P*-axis-related dimer. The metal ions and active site are located at the C-terminal end of the β-barrel. The residues implicated in binding metal or substrates belong to several β-strands and connecting loops. These loops are also involved in extensive intersubunit contacts (Rey et al., 1988). This molecular organization is also observed for the D-xylose isomerases from *Arthrobacter* (Henrick et al., 1987), *Streptomyces albus* (Dauter et al., 1989), and *Streptomyces rubiginosus* (Carrell et al., 1984). Superimposing our model onto that of the *Arthrobacter* enzyme reveals that C_α positions differ by not more than 0.58 Å rms except for a stretch of eight residues (277–285) in the loop connecting helix α_7 to strand β_8. This loop is longer by one residue in *Arthrobacter* and its conformation is very different. Otherwise, the two proteins are extremely similar.

(*ii*) *Structure of the K253R Mutant.* Crystals of the K253R (Lys-253 → Arg) mutant enzyme were obtained as described above. The K253R data were collected in the presence of Mg^{2+} and the substrate xylose to a resolution of 2.3 Å. Crystallographic refinement using PROLSQ (Hendrickson, 1985) started from the wild-type model and proceeded smoothly to a *R* factor of 0.149 for 108 519 independent reflections with no manual intervention, except for the replacement of the mutated side chain. Statistics on the stereochemistry are similar to those of the wild-type structure (Table I). The rms difference from that structure is 0.10 Å on all main-chain atoms; displacements larger than 0.5 Å affect only a few side chains on the surface of the molecule, remote from the site of mutation. No other changes were observed except for those involving the mutated residue, as described under Results.

(*iii*) *Structure of the Triple Mutant K309R/K319R/K323R.* Diffraction data on the structure of the triple mutant (Lys-309 → Arg/Lys-319 → Arg/Lys-323 → Arg) were collected in the presence of Co^{2+} and the inhibitor sorbitol, to a resolution of 2.3 Å. Refinement also proceeded smoothly to a *R* factor of 0.149 for 95 937 independent reflections. Statistics on stereochemistry are similar to those of the wild-type structure (Table I). Differences in atomic positions relative to the wild-type structure are very small (0.10 Å rms for main-chain atoms), except for loop 277–285 which was rebuilt. This loop is poorly ordered both in the wild-type and in the triple mutant enzymes, and it has different conformations in homologous D-xylose isomerases as discussed below.

Molecular Modeling. Molecular modeling was used as an aid in rationalizing observed properties and in guiding site-specific mutation experiments.

The model of human CuZnSOD is based on the crystal

Table I: Crystallographic Data on Wild-Type D-Xylose Isomerase and Engineered Mutants[a]

XI enzyme (ligands)	wild type (xylitol–Co^{2+})	K253R (D-xylose–Mg^{2+})	triple[b] (D-sorbitol–Co^{2+})
data collection			
resolution (Å)	2.2	2.3	2.3
no. of crystals	6	5	5
film packs	52	54	44
reflections: total	255106	240939	254909
reflections: independent	122049	108522	95937
R(merge) (%)	6.0	7.3	5.9
refinement			
R factor (%)	15.2	14.9	14.9
reflections	113145	108519	95937
protein atoms	12260	12268	12276
solvent	881	919	871
av *B* factor (Å^2)	21.1	18.2	18.8
distances (Å)			
bond lengths (1–2)	0.011	0.011	0.011
bond angles (1–3)	0.038	0.038	0.039
planar groups (1–4)	0.037	0.037	0.037
nonbonded	0.266	0.267	0.267
planarity (Å)	0.010	0.009	0.010
chiral volumes (Å^3)	0.127	0.128	0.128

[a] Crystals were soaked in 0.1 M sodium phosphate, pH 7.1, 1.7 M ammonium sulfate, and 0.2 M xylitol or in 0.5 M D-xylose or D-sorbitol and 10 mM $CoCl_2$ or 25 mM $MgSO_4$, as indicated. *R*(merge) = $\sum_i\sum_j|I_{ij} - I_i|/\sum_i\sum_j I_{ij}$, over equivalent reflections. Other details are given under Experimental Procedures. [b] Triple designates the XI triple mutant, K309R/K319R/K323R.

structure of the highly homologous (83% sequence identity) protein from the bovine species (Tainer et al., 1982; Brookhaven Protein Data Bank identifier code 2SOD), while that of *B. subtilis* D-glyceraldehyde-3-phosphate dehydrogenase is built on the crystal structure of the *Bacillus stearothermophilus* enzyme, strain NCA 1503 (Skarzynski et al., 1987; Brookhaven Protein Data Bank identifier code 1GD1), which displays 80% sequence identity with the protein from *B. subtilis*, strain BR151.

To rationalize the observed effect of the Lys → Arg mutation at position 9 in human CuZnSOD, changes introduced in amino acids were limited to the mutation at this position which is a lysine in both bovine and human wild-type enzymes. There is no other amino acid change within a 7-Å radius from this residue. Moreover, the Lys-9 side chain is well ordered with a mean atomic temperature factor less than 15 Å^2. Interactions made by Lys-9 in the bovine structure were therefore taken to represent those made in the human protein. Modeling of the Lys → Arg mutant was performed using a procedure analogous to that described previously (Wodak et al., 1986). Using the BRUGEL package (Delhaise et al., 1985), Lys-9 was replaced by an arginine, and the mutant side-chain conformation was determined by exhaustive search for stable conformations as a function of its freely rotating torsion angles varying in steps of 5°, evaluating the total energy of the otherwise rigid protein at each step. The lowest energy conformation obtained by this search was then subjected to 1000 steps of Conjugate Gradient energy minimization: all side chains with at least one atom within a 15-Å distance of the mutated side chain were allowed to move freely, while all other protein atoms were restrained to remain near their starting positions in the bovine crystal structure. The computations did not consider explicit water molecules except for certain well-defined water positions. A distance-dependent dielectric constant was used to mimic the presence of bulk solvent (Warshel & Levitt, 1976). The final protein conformation obtained by this two-step procedure was used to analyze interactions made by the mutated side chain with the protein.

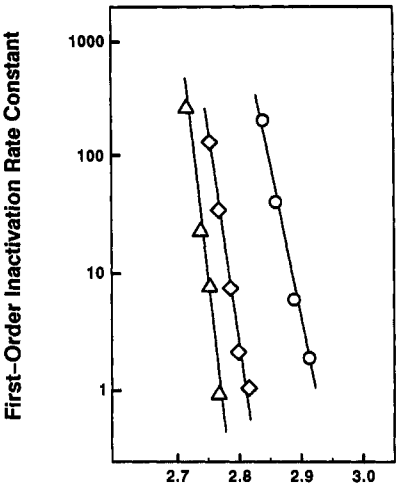

FIGURE 2: Arrhenius plots for the kinetics of heat inactivation of wild-type D-xylose isomerase: (○) apoenzyme, no metal added; (◇) apoenzyme + 10 mM $MgSO_4$; (△) apoenzyme + 10 mM $CoCl_2$. The heat inactivation assays were done in 50 mM MOPS (pH 7.2 at 25 °C). Each data point was derived by fitting the inactivation kinetics data at each temperature to monoexponential decay (see Experimental Procedures).

Interactions involving the side chains at residues at position 281 in *B. subtilis* mutants were studied in a similar fashion. Here, the *B. stearothermophilus* GAPDH crystal structure, which contains a well-ordered arginine at position 281, was used to analyze the interactions made by Arg-281 in the G281R mutant. Lysine was then substituted into the same structure, and the lowest energy conformation of this and all other side chains within a radius of 5 Å from residues at positions O201, P281, Q201, and R281 (see Figure 10) was derived from a systematic search, followed by energy minimization. The same strategy was utilized to examine the effect of replacing Arg-52 and Ser-202 of the *B. stearothermophilus* enzyme by lysine and asparagine, respectively, which are present at those positions in GAPDH from *B. subtilis*. These substitutions were analyzed because they represent the only changes within a 7-Å radius from each member of the ion pair formed between neighboring subunits by residues Arg-281 and Glu-201 in the GAPDH tetramer from *B. stearothermophilus*.

Analogous systematic search procedures were used to determine which of the positions occupied by lysines in the crystal structure of wild-type xylose isomerase from *A. missouriensis* could not accommodate arginine substitutions without requiring significant structural rearrangements.

RESULTS

Heat Inactivation of Wild-Type Xylose Isomerase. Figure 1 illustrates the overall organization of the XI molecule (Rey et al., 1988). The protein is made up of four identical subunits (labeled A–D) and organized into a dimer of dimers (labeled AB and CD). *A. missouriensis* XI inactivates irreversibly above 50 °C at neutral pH in the absence of divalent cations. Addition of Mg^{2+} or Co^{2+} ions prevents denaturation at temperatures up to 80–85 °C. The stabilizing effect of divalent metal ions is apparent in the Arrhenius plots of the kinetics of thermal inactivation (Figure 2). Glucose, also a substrate for xylose isomerase, has an opposite effect: metal-free XI inactivates faster at 60 °C in the presence of glucose than in its absence (Figure 3), a behavior also observed with other sugar substrates such as fructose, xylose, or, even more so, ribose but not with the active-site specific inhibitor xylitol, a nonreducing sugar (N. T. Mrabet, unpublished experiments). Moreover, extensive dialysis to eliminate unbound sugar does not restore the enzymatic activity.[6] We therefore hypothesized

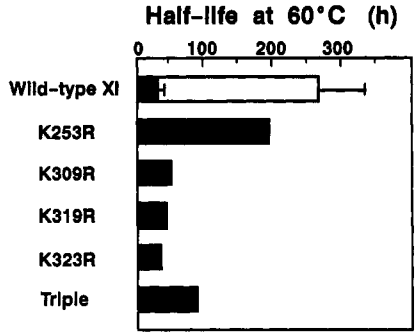

FIGURE 3: Glucose-induced inactivation of wild-type and mutant D-xylose isomerases. Half-lives at 60 °C in the presence of 250 mM α-D-glucose (solid bars) were computed by fitting the inactivation kinetics data to monoexponential decay (see Experimental Procedures). The effect of incubating wild-type XI under the same conditions but in the absence of glucose is also shown (open bar). The error bars represent one standard deviation (SD) for independent experiments: WT, $n = 4$ in the absence of glucose, and $n = 3$ in the presence of glucose. Triple designates the triple mutant, K309R/K319R/K323R.

that inactivation results from the nonenzymatic glycation of important amino groups in the XI molecule, rather than due to alteration of protein conformation produced by substrate binding. Such chemical modification has been subject to extensive studies and shown to occur in a reaction consisting of Schiff base formation between the sugar aldehyde function and both N-terminal and lysyl N^{ϵ}-amino groups; this aldimine linkage is then converted into a stable ketoamine by Amadori rearrangement, with potential alteration of biological activity (Bookchin & Gallop, 1968; Bunn et al., 1978; Brownlee et al., 1984; Furth, 1988). Glycation might be significant in industrial processes where xylose isomerase is used to produce fructose from glucose. An attractive protein engineering approach to prevent such chemical modification is to substitute arginine for lysine. Three independent chemical properties of the guanidinium group of arginine make it a sensible choice: (i) maintenance of a net positive charge; (ii) reduced chemical reactivity due to a very high pK_a; and (iii) occurrence of resonance stabilization and concurrent possibility to provide extended H-bond patterns (Riordan, 1979).

Identification of the Mutation Sites in Xylose Isomerase. Each XI subunit contains 20 lysines (Amore & Hollenberg, 1989). Five of them, at positions[7] 100, 118, 149, 183, and 253, have solvent-accessible surface areas less than 5 Å^2 and are thus buried in the tetramer (Miller et al., 1987). Lysines 100 and 253, which also reside in the interdimer interface and are engaged in a number of electrostatic interactions, were obvious candidates for substitution. Model building with the BRUGEL package (Delhaise et al., 1985) suggested that position 253, but not position 100, could accommodate an arginine side chain. Lysines at positions 309, 319, and 323 also attracted our attention. They are located on the solvent-exposed side of helix α_8 of the TIM barrel, and they form a cluster of positively charged residues with nearby arginines at positions 313, 321, and 326. The resulting positive electrostatic potential is expected to lower the pK_a of the lysine N^{ϵ}-amino groups, thereby increasing their reactivity toward open-chain sugar carbonyls.

Thermal Inactivation of Xylose Isomerase Mutants. Substitutions with arginine and with glutamine were performed at positions 253, 309, 319, and 323. Glutamine was used as a control of the effect of removing the positive charge to mimic

[6] N. T. Mrabet, manuscript in preparation.

[7] The *A. missouriensis* XI sequence is numbered starting with the N-terminal amino acid as residue number 2.

Table II: Kinetic Parameters of Wild-Type D-Xylose Isomerase and Engineered Mutants

xylose isomerase	k_{cat}(xylose) (s^{-1})	K_M(xylose) (mM)	k_{cat}(glucose) (s^{-1})	K_M(glucose) (mM)	K_M(Mg^{2+}) (mM)
wild type	17.3	4.8	24.9	217	0.08
K253R	17.8	5.3	19.6	177	0.10
K253Q	10.9	4.4	20.9	210	0.08
triple[b]	18.3	5.3	25.1	259	0.12

[a] Steady-state kinetic parameters are derived from fitting the data to the Michaelis–Menten equation (Michaelis & Menten, 1913) using the GraphPad software. Other methods are described under Experimental Procedures. [b] Triple designates the XI mutant, K309R/K319R/K323R.

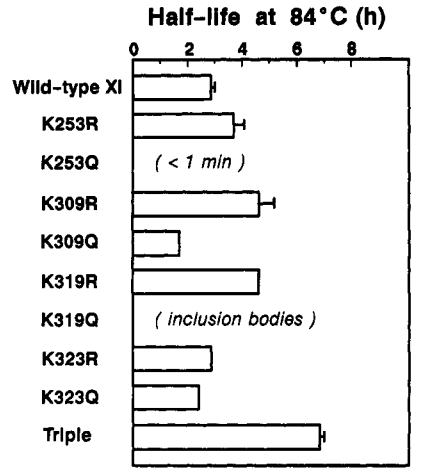

FIGURE 4: Heat inactivation of wild-type D-xylose isomerase and engineered mutants in the absence of glucose. The heat inactivation experiments in the absence of glucose were conducted at 84 °C as described under Experimental Procedures except that the apoenzymes were equilibrated here in Chelex-treated EPPS (50 mM, pH 7.5 at 84 °C) in the presence of 5.0 mM $MgSO_4$. Half-lives were computed as described in the legend to Figure 3. The error bars represent 1 SD for independent experiments: WT, $n = 8$; K253R, $n = 4$; K309R, $n = 2$; Triple, $n = 2$. Triple designates the triple mutant, K309R/K319R/K323R.

the charge loss produced by glycation. In addition, the triple mutant K309R/K319R/K323R was made. All mutants were well expressed in *E. coli* as soluble proteins, except K319Q which aggregates irreversibly in inclusion bodies and could not be evaluated. The soluble variants all display enzymatic properties similar to the wild type (Table II); this includes the mutants at position 253 which is relatively close to the enzyme's active site. In the presence of glucose at 60 °C, the arginine-containing mutants are more stable (Figure 3). The largest gain in stability is observed with K253R as its half-life increases about 6-fold. Under the same conditions, two of the single-point surface arginine mutants are also more stable, but to a lesser extent: half-lives are increased 1.6- and 1.4-fold for K309R and K319R, respectively. K323R shows no measurable improvement over wild type. Cumulation of the three arginine substitutions at the surface of the protein results in a stability gain by a factor of 3.

Figure 4 shows that, at 84 °C in the presence of Mg^{2+} but in the absence of sugar, K253R is also stabilized relative to the wild type, but the effect is less than with glucose: the half-life increases by 30% to 3.7 h. In contrast, the glutamine substitution is strikingly destabilizing, the half-life of K253Q being less than 1 min. The K309R and K319R mutants also show improved thermal stability under these conditions, while the corresponding glutamine mutants are destabilized to various degrees, and substitution at site 323 with either arginine or glutamine has no effect. Finally, the triple mutant K309R/K319R/K323R is more stable than either single mutant, with a half-life of 7 h. We thus see that even ther-

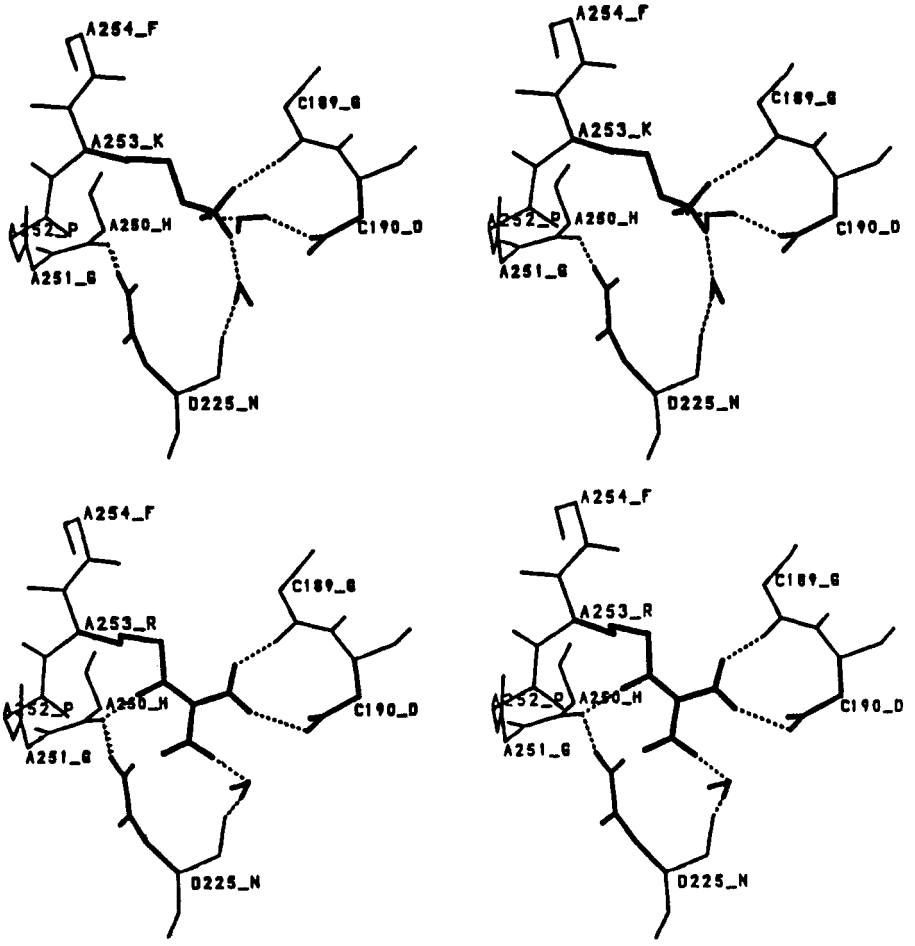

FIGURE 5: Crystal structures of wild-type D-xylose isomerase and mutant K253R. Stereoscopic representation of the wild-type (panel A, top) and K253R mutant (panel B, bottom) crystal structures around position 253. The lysine and arginine side chains are shown with heavy lines. Water molecules and other side chains are shown with light lines. For clarity, only the heavy atoms and the polar hydrogens are shown. Residues are identified by the chain label followed by the position and the amino acid 1-letter code. Dashed lines indicate the hydrogen-bonding network. Note that the side-chain amide of Asn-225 in subunit D donates a hydrogen bond to the carbonyl of Gly-251 of subunit A in the wild-type XI. This could fix the amide orientation, which would be maintained in the K253R variant. In the mutant, however, the amide $N_{\eta 2}$ is in van der Waals contact with the guanidinium group of Arg-253. This presumably unfavorable interaction can be relieved by flipping the amide at the expense of losing the bond to Gly-251.

mostable proteins can lend themselves to stability improvement either by single amino acid substitutions or through the cumulative effect of several sequence changes.

Crystal Structure of the K253R Mutant of Xylose Isomerase. Lysine 253 in the wild type lies at the end of a long loop which connects strand β_7 to helix α_7. This part of the loop runs along the *Q* axis of symmetry near the center of the tetramer and is in van der Waals contact with its counterpart in a neighboring subunit. Subunit packing in this region leaves large cavities filled by water molecules which are well-defined in the electron density map. The side chain of Lys-253 in subunit A extends into one such cavity and makes contacts with two other subunits involving Asn-185, Gly-189, and Asp-190 of subunit C across the *Q* axis of symmetry and Asn-225 of subunit D across the *R* axis (Figure 5A). The ϵ-NH_3 points toward the carboxylate of Asp-190 of subunit C, but the lysine side chain is too short to make a hydrogen bond: the electrostatic interaction occurs via a water molecule. A proper H-bond is formed instead with the main-chain carbonyl of Gly-189 of subunit C.

The crystal structure of the K253R mutant shows the arginine side chain in the same extended conformation as the lysine (Figure 5B). Like the lysine, Arg-253 has a well-ordered side chain (average *B* factor, 15 Å^2). Being longer, however, it can form a direct H-bond with Asp-190 of subunit C: the $N^{\eta 2}$ atom in the guanidinium group replaces the bridging water molecule. It also has additional H-bonding possibilities which are used to maintain the H-bond to the carbonyl of Gly-189 of subunit C. The N^{ϵ} atom forms a new intrachain H-bond to the carbonyl of His-250 of subunit A, which the lysine cannot satisfy in the wild type. As for the $N^{\eta 1}$ atom, it can form a hydrogen bond with one of two possible orientations of the amide group of Asn-225 of subunit D. All other H-bonds, particularly those involving active-site residues, are maintained as in the wild-type XI.

Crystal Structure of the K309R/K319R/K323R Triple Mutant of Xylose Isomerase. Helix α_8 where the substitutions are made is close to the N-terminal strand β_1. The sites involved do not participate in subunit contacts but instead are on the protein surface, exposed to solvent. In the wild-type crystal structure, the side chain of Lys-309 makes no obvious interaction and is disordered. In contrast, the crystal structure of the triple mutant indicates that the Arg-309 side chain makes at least two interactions: an ionic interaction with the carboxylate of Glu-306 in the same helix and a water-mediated H-bond to the main-chain oxygen of Asp-302 (not shown). At position 319, which is strictly invariant in all XI enzymes sequenced to date,[8] both the wild-type lysine and the mutant arginine are involved in polar interactions and their side chains are clearly positioned in the electron density maps, with av-

[8] P. Stanssens et al., manuscript in preparation.

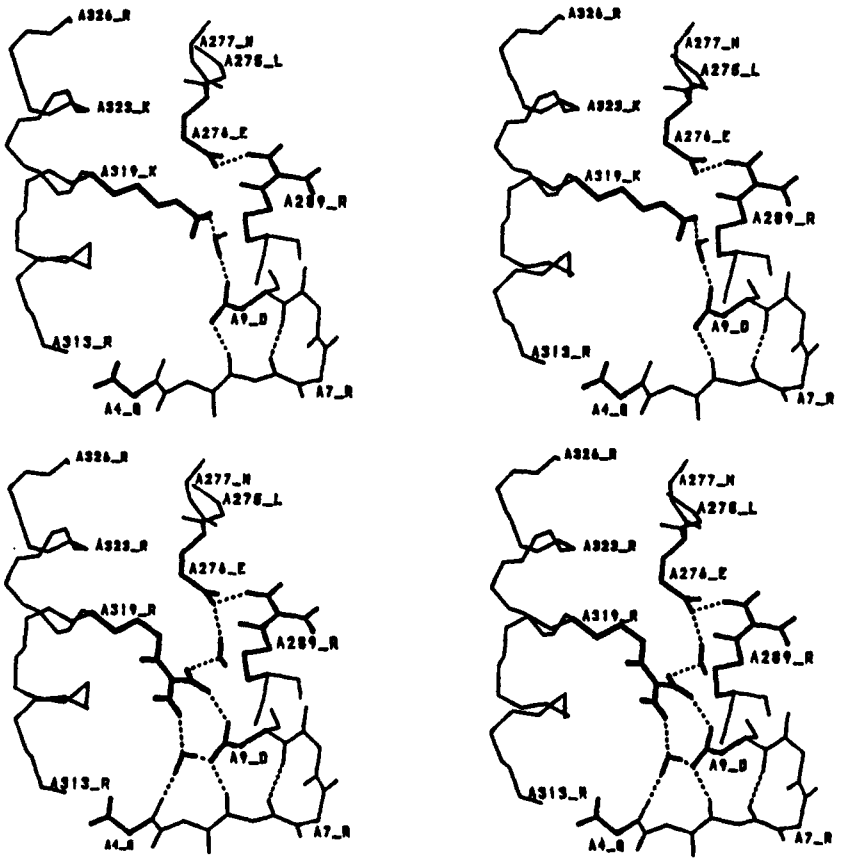

FIGURE 6: Crystal structures of wild-type D-xylose isomerase and mutant K319R. Stereoscopic representation of the wild-type (panel A, top) and mutant (panel B, bottom) crystal structures around position 319. The lysine and arginine side chains are shown with heavy lines. Water molecules and other side chains are shown with light lines. For clarity, only the heavy atoms and the polar hydrogens are shown, and only the backbone of the helix (residues 313–326) is represented. Dashed lines indicate the hydrogen-bonding network.

erage *B* factors of about 25 Å^2. Lys-319 makes a water-mediated ionic interaction with Asp-9 of strand β_1 which is far removed along the sequence (Figure 6A). This interaction is maintained in the mutant with the $N^{\eta 2}$ of Arg-319, and in addition, a direct H-bond is now formed with the $N^{\eta 1}$ (Figure 6B). Furthermore, new water-mediated hydrogen bonds are created with the carboxylate of Glu-276 and the main-chain oxygen of Gln-4. Not much can be said about the third mutation at position 323, as neither lysine nor the arginine at this position appear to participate in any interaction (not shown).

Other Proteins Also Stabilized by Engineered Arginines. (*i*) *Human CuZnSOD.* Superoxide dismutases are metalloenzymes which catalyze the dismutation of superoxide radicals into hydrogen peroxide (MacCord & Fridovich, 1969). Superoxide radicals are ubiquitous, potentially harmful byproducts of aerobic metabolism. Hence, SODs have been attributed a prominent superoxide-detoxifying function (Fridovich, 1975, 1989; Carlioz & Touati, 1986). CuZnSODs are unusually stable proteins (Forman & Fridovich, 1973; Malinowsky & Fridovich, 1979; Lepock et al., 1985; Roe et al., 1988). The crystal structure of the homodimeric bovine CuZnSOD has been solved to high resolution (Tainer et al., 1982). Despite a high (83%) sequence identity (Jabush et al., 1980), the human CuZnSOD is less stable and thus constitutes an interesting candidate for stability engineering by site-directed mutagenesis.

CuZnSODs have a lysine at position 9. This residue is conserved among vertebrates. The only known exception is provided by the enzyme from the swordfish where it is replaced by an arginine (Rocha et al., 1984). Lysine 9 in the human CuZnSOD model, which is based on the high-resolution crystal structure of the bovine enzyme (see Methods), is pointing out to the solvent and is not involved in interactions with the protein (Figure 7A). Modeling the mutation into arginine suggests that several new hydrogen bonds can be made between the guanidinium group and residues of the turn connecting strands β_1 and β_2 in the same subunit: these involve the main-chain oxygens of Gly-12 and Val-14 (Figure 7B). The CuZnSOD K9R mutant was engineered by site-directed mutagenesis and shown to have wild-type enzymatic activity, and its thermostability was compared to that of the wild type by measuring rates of inactivation at 85 °C. As shown in Figure 8, the decay obeys first-order kinetics for both enzymes. The half-life for the wild-type CuZnSOD is 17 min. The K9R mutant enzyme has improved heat stability as its half-life is increased nearly 6-fold to 98 min.

(*ii*) *B. subtilis GAPDH.* The 3D structures of the homotetrameric GAPDH enzymes from *B. stearothermophilus* (Bst-GAPDH) (Skarzynski et al., 1987) and from lobster (Moras et al., 1975) have been previously determined. These two proteins differ significantly in their thermostability, and considerable debate has been generated by attempts to rationalize this difference on structural grounds (Biesecker et al., 1977; Perutz, 1978; Argos et al., 1979; Walker et al., 1980; Skarzynski et al., 1987; Menéndez-Arias & Argos, 1989). Recently, the gene coding for GAPDH in *B. subtilis*, strain BR151 (Young et al., 1969), was cloned and sequenced (Viaene & Dhaese, 1989). Comparison of the deduced amino acid sequence[9] with that of the *B. stearothermophilus* enzyme

[9] To follow the residue numbering of *B. stearothermophilus* GAPDH, the N-terminal amino acid of the sequence from the *B. subtilis* enzyme is also numbered 0.

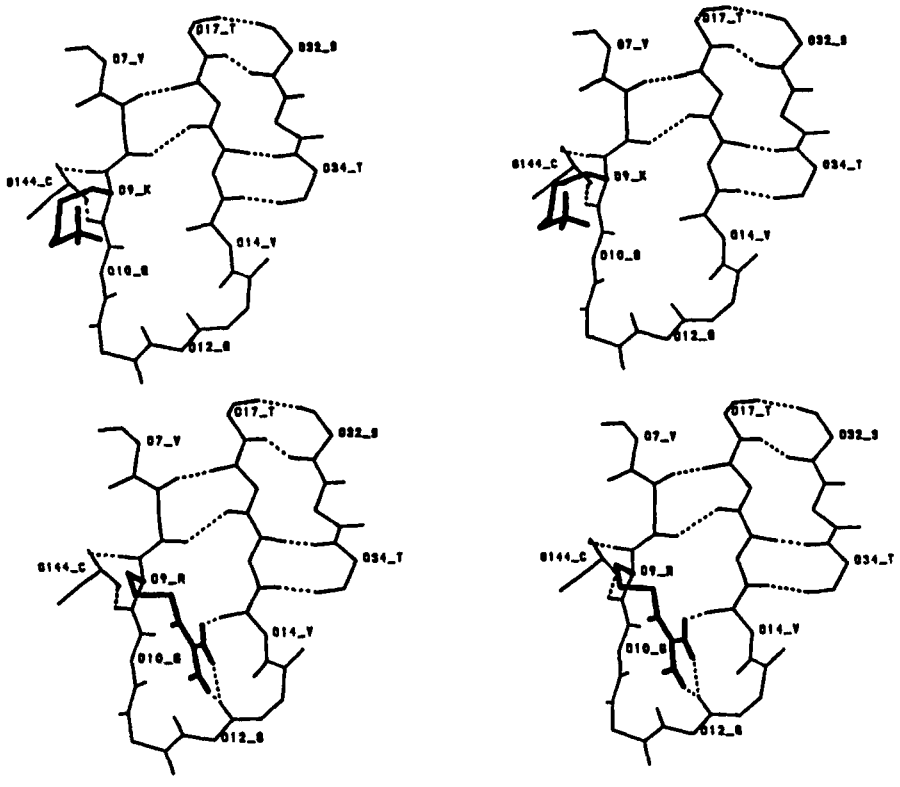

FIGURE 7: Modeled structures of human copper, zinc-superoxide dismutase and mutant K9R. (Panel A, top) Stereoview of the modeled human CuZnSOD around lysine O9 using the bovine crystal structure as a template. Only main-chain atoms are shown except for lysine and arginine which are shown in full atomic detail with heavy lines. Also shown is the conserved cysteine 144 (numbered 146 in the human enzyme) whose main-chain atoms hydrogen-bond to those of residue 9. Cys-144 also establishes an invariant disulfide bridge to Cys-57. This disulfide bridge strengthens the interactions between subunits in the CuZnSOD dimer and appears to stabilize the enzyme active site [reviewed in Getzoff et al. (1989)]. Dashed lines indicate the hydrogen-bonding network. (Panel B, bottom) same view as in (A) but showing the modeled arginine at O9.

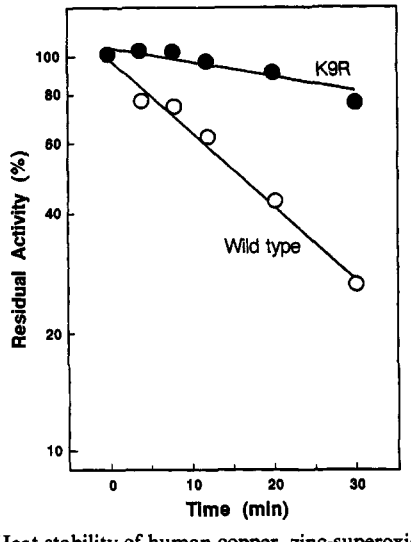

FIGURE 8: Heat stability of human copper, zinc-superoxide dismutase and mutant K9R. The kinetics of heat inactivation of wild-type CuZnSOD (○) and mutant K9R (●) at 85 °C are shown. Each data point is derived from the slope of the line generated by linear regression analysis using at least three dilutions with CuZnSOD concentrations ranging from 20 to 200 ng/mL. Specific activities were 5000 and 5200 units/mg for wild-type and mutant CuZnSOD, respectively, where one unit of enzymatic activity represents the amount of CuZnSOD which causes 50% inhibition of pyrogallol autoxidation. Other details are given under Experimental Procedures.

(Branlant et al., 1989, and references cited therein) reveals 80% identity. A remarkable difference, however, occurs at position 281 where the arginine of Bst-GAPDH is replaced by a glycine. Arginine 281 had been previously implicated in stabilizing the enzyme from *B. stearothermophilus* by forming ionic interactions with Glu-201 of adjacent subunits (Biesecker et al., 1977; Perutz, 1978), a residue also present in *B. subtilis* GAPDH (Bsu-GAPDH).

A soluble recombinant Bsu-GAPDH was expressed at high levels in *E. coli* WK6. Yet, initial attempts to isolate the protein failed to recover sufficient enzymatic activity. Moreover, rapid inactivation took place upon standing at 4 °C. Purification could nevertheless be achieved at 0–2 °C by adding DTT, EDTA, and the cofactor NAD^+ (see Experimental Procedures); these additives, previously shown to act as stabilizing agents for GAPDHs (Amelunxen & Carr, 1982), were thus always present in the study described here. Two mutant enzymes were engineered where Gly-281 was substituted with either lysine (G281K) or arginine (G281R). The mutations do not affect enzymatic activity. Residual activities after 15-min incubation at temperatures from 25 to 80 °C (Figure 9A) show that the midpoint of temperature stability is approximately 50 °C for wild-type Bsu-GAPDH but markedly increases to 72 °C for G281K and 75 °C for G281R. To assess the relative stability of the mutants, the kinetics of heat inactivation of G281K and G281R was followed at 75 °C (Figure 9B). The enzymatic decay obeys first-order kinetics. The half-life of G281K is 5 min. Arginine in place of lysine at position 281 increases the half-life of the enzyme 3.6-fold to 18 min. By comparison, wild-type Bsu-GAPDH has a half-life of 19 min but at 50 °C.

Interactions in the Arg-substituted Bsu-GAPDH can be readily analyzed using the Bst-GAPDH crystal structure (Skarzynski et al., 1987). Within 7 Å of residues 281 and 201, there are only two additional amino acid differences at positions 52 and 202 which are respectively Arg and Ser in Bst-GAPDH and Lys and Asn in Bsu-GAPDH. The side

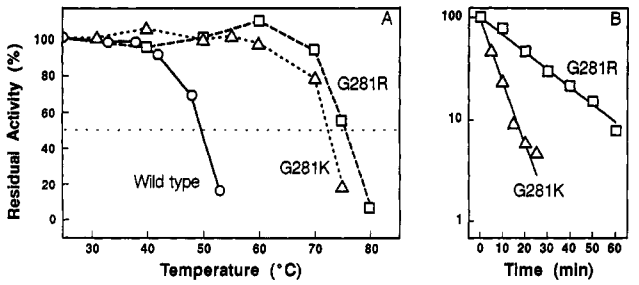

FIGURE 9: Heat stability of mutants G281R and G281K of *B. subtilis* D-glyceraldehyde-3-phosphate dehydrogenase. (Panel A) Temperature dependence for heat inactivation of wild-type *B. subtilis* GAPDH (○) and mutants G281K (△) and G281R (□) in TDNE buffer. (Panel B) Kinetics of inactivation at 75 °C of mutants G281K (△) and G281R (□). Specific activities were 48.4, 69.0, and 54.2 units/mg for wild type, G281K, and G281R, respectively, where one unit is defined as the amount of enzyme required to cause the disappearance of 1 mmol of NADH/min at 25 °C. Other details are provided under Experimental Procedures.

chain of residue 52 points into the solvent and is far from position 281. This suggests that the change from Arg to Lys at this position should have a negligible effect on the protein conformation as substantiated by our modeling analysis. Residue 202 is closer to the ionic pair of interest and more intimately involved in subunit contacts. However, it is located in a rather large pocket and remains quite accessible to solvent, while modeling suggests that a Ser → Asn change at this position would not entail alterations in the subunit contacts. Recalling that Bst-GAPDH is organized as a dimer of dimers, labeled OP and QR, we observe that Arg-281 in subunit P is involved in interdimer interactions to Glu-201 in subunit Q, and reciprocally. Moreover, Arg-281 of subunit R forms water-mediated ionic bonds with the carboxylate of Glu-201 in both subunits O and Q (Figure 10A). Model building suggests that the shorter lysine side chain can make a hydrogen bond across the interdimer interface only through a bridging water molecule (Figure 10B).

DISCUSSION

A fundamental question in biology pertains to the molecular basis for stability of protein tertiary and quaternary structures. This study shows that a multidisciplinary "protein engineering" approach is able to unravel yet another aspect of this complex problem. It highlights the role of a specific amino acid in the context of electrostatic stabilization in proteins.

Although the significance of electrostatic interactions to protein stability has been appreciated for quite some time (Perutz, 1978; Mrabet et al., 1986b; Rogers, 1989), only recently has their contribution been directly assessed by means of protein engineering (Wilkinson et al., 1983; Alber et al., 1987; Nicholson et al., 1988; Serano & Fersht, 1989; Serrano et al., 1990; Dao-pin et al., 1991), and much remains to be learned about how they are influenced by specific chemical groups. Analysis of the specific substitutions performed in this study provides valuable insights into this question.

Substituting Arginine for Lysine To Prevent Chemical Modification. Our strategy stems from the observation that glucose inactivates D-xylose isomerase from *A. missouriensis* and from the knowledge that lysine ϵ-amino groups have often been implicated in forming covalent adducts with glucose by nonenzymatic glycation [see, e.g., Brownlee et al. (1984) and Furth (1988)]. It consists of substituting arginine for lysine at selected positions where glycation is expected to disrupt packing and/or electrostatic interactions that are critical to protein stability. On this basis, four lysines have been mutated into arginine at positions 253, 309, 319, and 323 in XI, without affecting the enzymatic activity. The single substitution, K253R, and the triple replacement, K309R/K319R/K323R, are indeed shown to provide significant resistance to glucose-induced inactivation, with the K → R mutation at the buried interdimer interface position 253 displaying the largest effect. Single-site arginine mutants at positions 309 and 319 show very modest improvement in stability in the presence of glucose, while K323R behaves similarly to wild type. Combining the three mutants yields a larger stability gain probably resulting from the protection of the entire region.

In none of the cases analyzed here do the arginine substitutions generate any loss in stability. This implies that the grafted arginines are able to, at least, preserve the integrity of the electrostatic interactions made in the wild-type protein. In contrast, removing the lysine positive charge by replacing it with glutamine results in a loss of stability, except at position 323. A similar destabilizing effect is expected upon protein glycation which is accompanied by a significant decrease in the pK_a of the lysyl ϵ-amino group (Perrin et al., 1981).

A most plausible interpretation of our results is that the Lys → Arg substitutions interfere with glycation, either directly due to the arginine side chain being chemically unreactive or indirectly by impeding denaturation, or possibly through both mechanisms. This is further supported by evidence on the presence of sugar-modified peptides in digests of wild-type xylose isomerase after incubation with sugar derivatives.[6]

Although this study is the first to address the issue of enzyme denaturation by nonenzymatic glycation in industrial settings and to suggest an effective means to circumvent glucose-induced inactivation, its most significant contribution, and also its major emphasis, resides in the demonstration that the guanidinium-containing side chain is able to significantly enhance protein thermostability even in the *absence* of sugar substrates and hence under conditions where *glycation-dependent events are no longer involved.* Along with these findings, specific evidence is provided on the stabilizing role of electrostatic interactions.

Electrostatic Interactions at Intersubunit Interfaces Provide Important Contributions to Stability. Residue 253 in *A. missouriensis* XI lies at the boundary between the subunits of the tetramer, shielded from solvent. Crystal structure analysis suggests that the improved thermostability of the K253R mutant in the absence of glucose originates from strengthening existing electrostatic interactions at intersubunit contacts and/or by forming new ones. With this mutant, however, the stabilizing effect in the absence of sugar is only a fraction of that observed with other substitutions described here. This may be due to other unfavorable interactions formed by the longer Arg side chain with the amide group of Asn-225 in the neighboring subunit. In agreement with this interpretation, we find that the relative stabilizing effect of Arg versus Lys at position 253 is larger when the substitution is performed in a variant where Asn-225 has been replaced by Gly or Ala (N. T. Mrabet, unpublished observations).

When the X-ray structure of glyceraldehyde-3-phosphate dehydrogenase from *B. stearothermophilus* was compared to that of the lobster enzyme, it was suggested that thermostability results from ion pairs at the intersubunit interfaces that are absent in the mesophilic enzyme (Biesecker et al., 1977; Perutz, 1978; Walker et al., 1980). No definite conclusion could however be drawn, given the many other sequence changes. This issue was further complicated after it was found that in the even more stable *Thermus aquaticus* enzyme, along with many other amino acid differences, the surmised intersubunit ion pairs were absent (Harris et al., 1980; Walker et

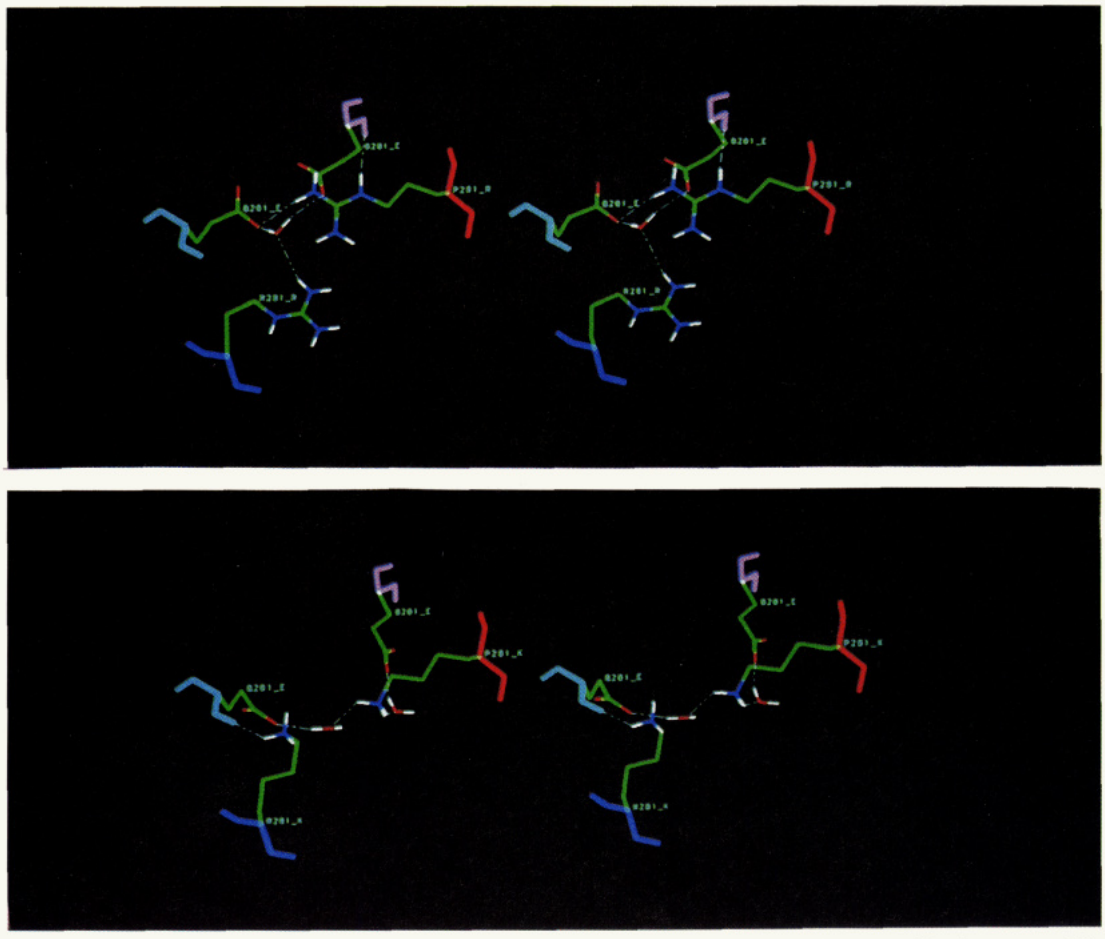

FIGURE 10: Modeled structures of mutants G281R and G281K of *B. subtilis* D-glyceraldehyde-3-phosphate dehydrogenase. The interactions involving arginine (panel A, top) or lysine (panel B, bottom) at position 281 are shown on stereoviews of structures modeled using *B. stearothermophilus* GAPDH as a template. For clarity, only the side chains of the residues of interest and the intervening water molecules are shown with polar hydrogens (in white); the carbon atoms are in green and the nitrogen atoms in dark blue, while the oxygen atoms are in red. Fragments of the backbone are shown with heavy lines and in magenta for subunit O, in red for subunit P, in light blue for subunit Q, and in dark blue for subunit R. The white dashed lines represent the hydrogen bonds. In panel A (Arg-281), atomic coordinates are from the crystal structure deposited in the Brookhaven Protein Data Bank (Berenstein et al., 1977). The conformations of the Arg-281 side chains in subunits P and R are different, corresponding to a well-documented departure from 2-fold symmetry (Skarzynski et al., 1987). Panel B shows the modeled orientation of the Lys-281 side chain (see Methods).

al., 1980). The experiments reported here on the closely related *B. subtilis* enzyme significantly clarify this point. In the absence of a positively charged residue at position 281, *B. subtilis* GAPDH has low stability. Introducing a lysine substantially enhances stability, and an arginine makes it even more stable. The shift in thermostability resulting from the presence of a positive charge in place of Gly-281 is large and comparable in size to the stability differences observed between proteins from mesophiles and thermophiles.

Thus, our findings clearly demonstrate the direct implication of subunit interactions in protein stability, and they further emphasize the roles of electrostatics and of arginine in this respect. Since only inactivation rates are measured here, it is, however, not possible to determine whether stabilization results from altered kinetic barriers of denaturation or from thermodynamic effects. Yet, it is reasonable to assume (Friedman & Beychok, 1979; Mrabet et al., 1986a; also our unpublished observations[10,11]) that subunit dissociation occurs as an early step in the denaturation of multimeric proteins and, thus, that improved electrostatic interactions between subunits may also delay denaturation.

Electrostatic Interactions on the Protein Surface May Also Provide Stabilizing Contributions. We find that arginine substitutions at solvent-accessible positions 309 and 319 in XI lead to enhanced stabilization and that replacing the surface-exposed lysine 9 in human CuZnSOD by an arginine also results in a significant increase in heat stability. We therefore conclude that electrostatic effects can contribute to protein stability even when they occur on the solvent-exposed protein surface, as was first suggested more than a decade ago by Perutz and Raidt (1975) and further confirmed by more recent studies (Anderson et al., 1990; Serrano et al., 1990).

Charges Poorly Solvated by the Protein Matrix Have a Destabilizing Effect. Model building suggests that the enhanced stability of Bsu-GAPDH mutants is due to the formation of ion pairs between the engineered positive charges and carboxylates in neighboring subunits. One can hence deduce that in the wild-type *B. subtilis* protein, which lacks the positive counter charge, the carboxylates are less efficiently solvated by the protein with a corresponding loss in stability. In the very stable GAPDH from the extreme thermophile *T. aquaticus*, both members of the ion pair are absent. Since favorable electrostatic interactions with the surrounding water

[10] Tetramer–dimer dissociation of XI mutant K253Q, but not that of wild-type XI or K253R, is induced by treatment with 5 M urea, pH 8.0, for up to 9 h, as shown by gel-permeation chromatography on a Superose-12 column run in 50 mM Tris-HCl, pH 8.0, 0.15 M NaCl, and 0.02% NaN_3.

[11] Gel-permeation chromatography of wild-type and mutant Bsu-GAPDHs prepared in 0.1 M TEA, pH 7.8, 1 mM DTT, 1 mM EDTA, and 1 mM NAD^+, was performed using a Superose-12 column run in 50 mM Tris-HCl, pH 8.0, 0.15 M NaCl, and 0.02% NaN_3.

are now also missing in the dissociated state, stability loss due to poor charge stabilization in the tetramer is avoided. Consistent with this interpretation, we find that wild-type *B. subtilis* GAPDH elutes as a dimer on gel-permeation chromatography, while the G281K and G281R mutants behave as tetramers.[11]

The reduced heat stability of the Lys → Gln mutants of xylose isomerase may have more complex origins. In the K253Q and K319Q mutants, the absence of the positive charge is likely to offset the electrostatic interactions with nearby Asp-190 and Asp-9, respectively, as well as with surrounding dipoles, particularly in the K253Q mutants where lysine in the wild type makes several H-bonds with main-chain and side-chain polar groups of residues Asn-185 and Gly-189 from neighboring subunits. Moreover, to explain the very drastic effect of replacing lysine at position 319 by glutamine, which leads to the production of the corresponding recombinant protein in inclusion bodies, interference with the folding process may be invoked, which is consistent with this lysine being strictly conserved in all isomerases sequenced so far. The smaller, yet detectable, destabilizing effect of glutamine at position 309 is consistent with weaker, albeit present, electrostatic interactions made in the wild type by the disordered lysine side chain and strengthened in the arginine mutant. On the other hand, the absence of any influence on protein stability upon mutation of Lys-323 to Arg or Gln is in accord with the observation that both Lys and Arg at that position are completely surrounded by solvent which effectively shields their charge.

On the Stabilizing Role of Arginine Relative to Lysine. The stabilizing role of arginine has been suspected for quite some time on the basis of the observation that the Arg/Lys ratio is larger in proteins from thermophilic organisms than in those from mesophiles (Argos et al., 1979; Menèndez-Arias & Argos, 1989) but so far not explicitly proven. Interfaces between subunits in oligomeric proteins as well as recognition sites for protein ligands also appear to favor the occurrence of arginines over lysines (Janin et al., 1988; Janin & Chothia, 1990). Our analysis provides a direct proof for the stabilizing influence of arginine and gives insights into why it occurs.

Inspection of the high-resolution X-ray structures of wild-type and mutant xylose isomerase enzymes shows that in nearly all cases with improved stability the arginine side chain forms at least one additional H-bond with another polar group in the protein. Model building suggests that this should also be the case in the arginine mutants of GAPDH and CuZnSOD. The additional H-bonding interactions probably occur as a consequence of the increased hydrogen-bonding potential of the guanidinium group and/or of the capacity of the longer arginine side chain to reach out further in space. The observed increase in thermostability should, however, result from the difference between interactions made by the guanidinium group with other polar groups in the protein native state and those made with the competing water molecules in the denatured state. Experimental measures of solvation free energies of arginine and lysine, available only for the *uncharged* species (Wolfenden et al., 1981), show arginine to be more soluble in water than lysine. This is consistent with some hydropathy scales but not with others [see, e.g., Table 6 of Makhatadze and Privalov (1990)]. It is therefore hard to assess the contribution from solubility properties to the gain in protein stability observed with arginine relative to lysine.

As a consequence of arginine being longer yet not more flexible than lysine, stability may also be gained by replacing water-mediated hydrogen bonds made by the lysine side chain with direct H-bonds with the guanidinium group, since reduced contact with solvent is expected to strengthen the corresponding electrostatic interactions. This is observed in the K253R and K319R mutants of xylose isomerase, whose stability is improved. Moreover, if the newly formed interactions in the mutant involve, over all, fewer crystallographically determined water molecules than in the wild type, this may result in entropy gain and provide additional stabilizing contributions.

All these considerations taken together provide an improved basis for understanding the origins of the stabilizing effect of arginine relative to lysine observed in several instances in this study and suggest novel approaches to engineering more stable proteins.

The fact that arginine is not more abundant than lysine in proteins known to date suggests that factors other than thermostability may have a more important role in evolution (Malcom et al., 1990).

Finally, measurements of the free energy of oligomer dissociation and of protein denaturation (Privalov, 1979, 1982; Kellis et al., 1989), done under conditions of reversibility not achieved in our experiments, along with theoretical analyses (Gao et al., 1989), would be required to shed further light on the nature of the physical–chemical effects of Lys → Arg mutations.

Added in Proof

After our manuscript was submitted for publication, the amino acid sequence of D-xylose isomerase from *T. thermophilus* was reported (Dekker et al., 1991). This highly thermostable enzyme contains an arginine at both position 253 and position 309.

Acknowledgments

We thank Marijke Beyaert and Adri Van Vliet for expert technical assistance and Philippe Delhaise and Michel Bardiaux for help with the BRUGEL package. We are deeply thankful to Dr. Patrick Dhaese for communicating results before publication and for the gift of the cloned *B. subtilis* GAPDH gene, to Dr. Stefan Marklund, University of Umeå School of Medicine, for advice in the determination of Cu-ZnSOD enzymatic activity, to Dr. Michel Honig, Research Institute of the Belgian Ministry of Agriculture, for metal analysis in xylose isomerase preparations, and to Dr. Franklin Bunn, Harvard Medical School, for insightful comments on an early version of the manuscript.

Registry No. Arg, 74-79-3; Lys, 56-87-1; XI, 9023-82-9; SOD, 9054-89-1; GAPDH, 9001-50-7.

References

Ahren, T. J., & Klibanov, A. M. (1985) *Science 228*, 1280–1284.

Alber, T. (1989) *Annu. Rev. Biochem. 58*, 765–798.

Alber, T., Dao-pin, S., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987a) *Biochemistry 26*, 3754–3758.

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P., & Matthews, B. W. (1987b) *Nature 330*, 41–46.

Amelunxen, R. E., & Carr, D. O. (1982) *Methods Enzymol. 89*, 264–267.

Amore, A., & Hollenberg, C. P. (1989) *Nucleic Acids Res. 17*, 7515.

Anderson, D. E., Becktel, W. J., & Dahlquist, F. W. (1990) *Biochemistry 29*, 2403–2408.

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G., & Tratschin, J. D. (1979) *Biochemistry 18*, 5698–5703.

Beauchamp, C., & Fridovich, I. (1971) *Anal. Biochem. 44*, 276–287.

Berenstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanoushi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Biesecker, G., Harris, J. I., Thierry, J., Walker, J. E., & Wonacott, A. J. (1977) *Nature 266*, 328–333.
Bookchin, R. M., & Gallop, P. M. (1968) *Biochem. Biophys. Res. Commun. 32*, 86–93.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.
Branlant, C., Oster, T., & Branlant, G. (1989) *Gene 75*, 145–155.
Brot, N., & Weissenbach, H. (1983) *Arch. Biochem. Biophys. 223*, 271–281.
Brownlee, M., Vlassara, H., & Cerami, A. (1984) *Ann. Intern. Med. 101*, 527–537.
Bunn, H. F., Gabbay, K. H., & Gallop, P. M. (1978) *Science 200*, 21–27.
Callens, M., Kerters-Hilderson, H., Van Opstal, O., & De Bruyne, C. K. (1986) *Enzyme Microb. Technol. 8*, 696–700.
Carlioz, A., & Touati, D. (1986) *EMBO J. 5*, 623–630.
Carrell, H. L., Rubin, B. H., Hurley, T. J., & Glusker, J. P. (1984) *J. Biol. Chem. 259*, 3230–3236.
Casal, J. I., Ahren, T. J., Davenport, R. C., Petsko, G. A., & Klibanov, A. M. (1987) *Biochemistry 26*, 1258–1264.
Creighton, T. E. (1988) *BioEssays 8*, 57–63.
Dao-pin, S., Sauer, U., Nicholson, H., & Matthews, B. W. (1991) *Biochemistry 30*, 7142–7153.
Dauter, Z., Dauter, M., Hemker, J., Witzel, H., & Wilson, K. (1989) *FEBS Lett. 247*, 1–8.
De Boer, H. A., Comstock, L. J., & Vasser, M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 21–25.
Dekker, K., Yamagata, H., Sakaguchi, K., & Udaka, S. (1991) *J. Bacteriol. 173*, 3078–3083.
Delhaise, P., Van Belle, D., Bardiaux, M., Alard, P., Hamers, P., Van Cutsem, E., & Wodak, S. J. (1985) *J. Mol. Graphics 3*, 116–119.
Estell, D. A., Graycar, T. P., & Wells, J. A. (1985) *J. Biol. Chem. 260*, 6518–6521.
Forman, H. J., & Fridovich, I. (1973) *J. Biol. Chem. 248*, 2645–2649.
Fridovich, I. (1975) *Annu. Rev. Biochem. 44*, 147–159.
Fridovich, I. (1989) *J. Biol. Chem. 264*, 7761–7764.
Friedman, F. K., & Beychok, S. (1979) *Annu. Rev. Biochem. 48*, 217–250.
Furth, A. (1988) *New Sci. 117* (1602), 58–62.
Gao, J., Kuczera, K., Tidor, B., & Karplus, M. (1989) *Science 224*, 1069–1072.
Getzoff, E. D., Tainer, J. A., Stempien, M. M., Bell, G. I., & Hallewell, R. A. (1989) *Proteins: Struct., Funct., Genet. 5*, 322–336.
Hallewell, R. A., Masiarz, F. R., Najarian, R. C., Puma, J. P., Quiroga, M. R., Randolph, A., Sanchez-Pescador, R., Scandella, C. J., Smith, B., Steimer, K. S., & Mullenbach, G. T. (1985) *Nucleic Acids Res. 13*, 2017–2034.
Harding, J. J. (1985) *Adv. Protein Chem. 37*, 247–334.
Harris, J. I., Hocking, J. D., Runswick, M. J., Suzuki, K., & Walker, J. E. (1980) *Eur. J. Biochem. 108*, 535–547.
Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.
Henrick, K., Blow, D. M., Carell, H. L., & Glusker, J. P. (1987) *Protein Eng. 1*, 467–469.
Howard-Flanders, P., Boyce, R. P., & Theriot, L. (1966) *Genetics 53*, 1119–1136.
Jabusch, J. R., Farb, D. L., Kerschensteiner, D., & Deutsch, H. F. (1980) *Biochemistry 19*, 2310–2316.
Janin, J., & Chothia, C. (1990) *J. Biol. Chem. 265*, 16027–16030.
Janin, J., Miller, S., & Chothia, C. (1988) *J. Mol. Biol. 204*, 155–164.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Kellis, J. T., Jr., Nyberg, K., Sali, D., & Fersht, A. R. (1988) *Nature 334*, 784–786.
Kellis, J. T., Jr., Nyberg, K., & Fersht, A. R. (1989) *Biochemistry 28*, 4914–4922.
Kersters-Hilderson, H., Callens, M., Van Opstal, O., Vangrysperre, W., & De Bruyne, C. K. (1987) *Enzyme Microb. Technol. 9*, 145–148.
Lepock, J. R., Arnold, L. D., Torrie, B. H., Andrews, B., & Kruuv, J. (1985) *Arch. Biochem. Biophys. 241*, 243–251.
Leslie, A. G. W. (1987) *Acta Crystallogr. A43*, 134–136.
MacCord, J. M., & Fridovich, I. (1969) *J. Biol. Chem. 244*, 6049–6055.
Makhatadze, G. I., & Privalov, P. L. (1990) *J. Mol. Biol. 213*, 375–384.
Malcom, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F., & Wilson, A. C. (1990) *Nature 345*, 86–89.
Malinowsky, D. P., & Fridovich, I. (1979) *Biochemistry 18*, 5055–5060.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marklund, S., & Marklund, G. (1974) *Eur. J. Biochem. 47*, 469–479.
Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.
Matthews, B. W., Nicholson, H., & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6663–6667.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–559.
Menẽndez-Arias, L., & Argos, P. (1989) *J. Mol. Biol. 206*, 397–406.
Miller, S., Janin, J., Lesk, A. M., & Chothia, S. (1987) *J. Mol. Biol. 196*, 641–656.
Misset, O., Bos, O. J. M., & Opperdos, F. R. (1986) *Eur. J. Biochem. 157*, 441–453.
Mitchinson, C., & Baldwin, R. L. (1986) *Proteins: Struct., Funct., Genet. 1*, 23–33.
Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C., & Rossmann, M. G. (1975) *J. Biol. Chem. 250*, 9137–9162.
Mrabet, N. T., Shaeffer, J. R., McDonald, M. J., & Bunn, H. F. (1986a) *J. Biol. Chem. 261*, 1111–1115.
Mrabet, N. T., McDonald, M. J., Turci, S., Sarkar, R., Szabo, A., & Bunn, H. F. (1986b) *J. Biol. Chem. 261*, 5222–5228.
Nicholson, H., Becktel, W. J., & Matthews, B. W. (1988) *Nature 336*, 651–656.
Perrin, D. D., Dempsey, B., & Serjeant, E. P. (1981) in pK_a *Prediction for Organic Acids and Bases*, Chapman and Hall, New York.
Perutz, M. F. (1978) *Science 201*, 1187–1191.
Perutz, M. F., & Raidt, H. (1975) *Nature 255*, 256–258.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.
Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1–104.
Rey, F., Jenkins, J., Janin, J., Lasters, I., Alard, P., Claessens, M., Matthyssens, G., & Wodak, S. J. (1988) *Proteins: Struct., Funct., Genet. 4*, 165–172.
Riordan, J. F. (1979) *Mol. Cell. Biochem. 26*, 71–92.
Rocha, H. A., Bannister, W. H., & Bannister, J. V. (1984) *Eur. J. Biochem. 145*, 477–484.
Roe, J. A., Butler, A., Scholler, D. M., Valentine, J. S., Marky, L., & Breslauer, K. J. (1988) *Biochemistry 27*, 950–958.

Rogers, N. K. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 359–389, Plenum Press, New York.
Rosenberg, S., Barr, P. J., Najarian, R. C., & Hallewell, R. A. (1984) *Nature 312*, 77–80.
Sali, D., Bycroft, M., & Fersht, A. R. (1988) *Nature 335*, 741–743.
Sandberg, W. S., & Terwilliger, T. C. (1989) *Science 245*, 54–57.
Sanger, F., Nicklen, S., & Coulson, A. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 296–299.
Serrano, L., Horovitz, A., Avron, B., Bycroft, M., & Fersht, A. R. (1990) *Biochemistry 29*, 9343–9352.
Skarzynski, T., Moody, P. C. E., & Wonacott, A. J. (1987) *J. Mol. Biol. 193*, 171–187.
Stanssens, P., Opsomer, C., McKeown, Y. M., Kramer, W., Zabeau, M., & Fritz, H.-J. (1989) *Nucleic Acids Res. 17*, 4441–4454.
Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S., & Richardson, D. C. (1982) *J. Mol. Biol. 160*, 181–217.
Tartof, K. D., & Hobbs, C. A. (1988) *Gene 67*, 169–182.
Viaene, A., & Dhaese, P. (1989) *Nucleic Acids Res. 17*, 1251.
Walker, J. E., Wonacott, A. J., & Harris, J. I. (1980) *Eur. J. Biochem. 108*, 581–586.
Warshel, A., & Levitt, M. (1976) *J. Mol. Biol. 103*, 227–249.
Wetzel, R. (1987) *Trends Biochem. Sci. 12*, 478–482.
Wilkinson, A. J., Fersht, A. R., Blow, D. M., & Winter, G. (1983) *Biochemistry 22*, 3581–3586.
Wodak, S. J., DeCoen, J. L., Edelstein, S. J., Demarne, H., & Beuzard, Y. (1986) *J. Biol. Chem. 261*, 14717–14724.
Wolfenden, R., Andersson, L., Cullis, P. M., & Southgate, C. C. B. (1981) *Biochemistry 20*, 849–855.
Yannisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.
Young, F. E., Smith, D., & Reilly, B. E. (1969) *J. Bacteriol. 98*, 1087–1097.
Zell, R., & Fritz, H.-J. (1987) *EMBO J. 6*, 1809–1815.

Histidine Residues at the N- and C-Termini of α-Helices: Perturbed pK_as and Protein Stability

Javier Sancho, Luis Serrano, and Alan R. Fersht*
MRC Unit for Protein Function and Design, Cambridge IRC for Protein Engineering, Cambridge University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.
Received August 14, 1991; Revised Manuscript Received November 13, 1991

ABSTRACT: A single histidine residue has been placed at either the N-terminus or the C-terminus of each of the two α-helices of barnase. The pK_a of that histidine residue in each of the four mutants has been determined by ^{1}H NMR. The pK_as of the two residues at the C-terminus are, on average, 0.5 unit higher, and those of the residues at the N-terminus are 0.8 unit lower, than the pK_a of histidines in unfolded barnase at low ionic strength. The conformational stability of the mutant proteins at different values of pH has been measured by urea denaturation. C-Terminal histidine mutants are $\sim$0.6 kcal mol^{-1} more stable when the introduced histidine is protonated, both at low and high ionic strength. N-Terminal mutants with a protonated histidine residue are $\sim$1.1 kcal mol^{-1} less stable at low ionic strength and 0.5 kcal mol^{-1} less stable at high ionic strength (1 M NaCl). The low-field ^{1}H NMR spectra of the mutant proteins at low pH suggest that the C-terminal histidines form hydrogen bonds with the protein while the N-terminal histidines do not form the same. The perturbations of pK_a and stability result from a combination of different electrostatic environments and hydrogen-bonding patterns at either ends of helices. The value of 0.6 kcal mol^{-1} represents a lower limit to the favorable electrostatic interaction between the α-helix dipole and a protonated histidine residue at the C-terminal end of the helix. Part of this electrostatic interaction should be attributed to the difference in stability between the charged and the noncharged hydrogen bond of C-terminal histidines.

The properties of amino acid side chains in a protein depend on their local and global environments. Charged residues, in particular, are affected by long-range as well as local electrostatic effects. These can perturb the ionic properties of the residue, and conversely, the mutual interactions affect the stability of the protein. It has been suggested, for example, that α-helices have a macroscopic dipolar character arising from the parallel alignment of the dipolar peptide bonds of the helix (Wada, 1976). The effect of the α-helix dipole has been suggested to be equivalent to that of a −0.5 elemental charge at the C-terminus plus a +0.5 elemental charge at the N-terminus of the helix (Hol et al., 1978). The dipolar nature of α-helices has been invoked to explain the structure of ligand binding sites (Hol et al., 1978), the relative disposition of α-helices within proteins (Sheridan et al., 1982; Presnell & Cohen, 1989), and the clustering of positive and negative charges toward the C- and N-termini of helices (Richardson & Richardson, 1988). The interaction between α-helix dipoles and charged residues in small peptides (Shoemaker et al., 1985; Fairman et al., 1989) as well as in proteins (Perutz et al., 1985; Sali et al., 1988; Nicholson et al., 1988) has been considered to contribute to stability, but little quantitative data are available (Perutz et al., 1985; Sali et al., 1988; Nicholson et al., 1988).

One of the most detailed measurements on the effects of charge–helix dipole interactions has come from the raising of the pK_a of His18, a residue at the C-terminus of the first helix in barnase (Sali et al., 1988). The imidazole side chain of His18 resides at the negatively charged terminus of the helix and makes a hydrogen bond with the backbone carboxamide

0006-2960/92/0431-2253$03.00/0