Protein Science (1994), *3*:541–548. Cambridge University Press. Printed in the USA.

REVIEW

A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens

C.L. BORDERS, JR., JOHN A. BROADWATER, PAULA A. BEKENY, JOHANNA E. SALMON, ANN S. LEE, AIMEE M. ELDRIDGE, AND VIRGINIA B. PETT

Department of Chemistry, The College of Wooster, Wooster, Ohio 44691

(RECEIVED October 26, 1993; ACCEPTED January 19, 1994)

Abstract

We propose that arginine side chains often play a previously unappreciated general structural role in the maintenance of tertiary structure in proteins, wherein the positively charged guanidinium group forms multiple hydrogen bonds to backbone carbonyl oxygens. Using as a criterion for a "structural" arginine one that forms 4 or more hydrogen bonds to 3 or more backbone carbonyl oxygens, we have used molecular graphics to locate arginines of interest in 4 proteins: Arg 180 in *Thermus thermophilus* manganese superoxide dismutase, Arg 254 in human carbonic anhydrase II, Arg 31 in *Streptomyces rubiginosus* xylose isomerase, and Arg 313 in *Rhodospirillum rubrum* ribulose-1,5-bisphosphate carboxylase/oxygenase. Arg 180 helps to mold the active site channel of superoxide dismutase, whereas in each of the other enzymes the structural arginine is buried in the "mantle" (i.e., inside, but near the surface) of the protein interior well removed from the active site, where it makes 5 hydrogen bonds to 4 backbone carbonyl oxygens. Using a more relaxed criterion of 3 or more hydrogen bonds to 2 or more backbone carbonyl oxygens, arginines that play a potentially important structural role were found in yeast enolase, *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase, bacteriophage T4 and human lysozymes, *Enteromorpha prolifera* plastocyanin, HIV-1 protease, *Trypanosoma brucei brucei* and yeast triosephosphate isomerases, and *Escherichia coli* trp aporepressor (but not trp repressor or the trp repressor/operator complex). In addition to helping form the active site funnel in superoxide dismutase, the structural arginines found in this study play such diverse roles as stapling together 3 strands of backbone from different regions of the primary sequence, and tying α-helix to α-helix, β-turn to β-turn, and subunit to subunit.

Keywords: backbone carbonyl oxygens; buried arginine; carbonic anhydrase; enolase; glyceraldehyde-3-phosphate dehydrogenase; HIV-1 protease; lysozyme; manganese superoxide dismutase; multiple hydrogen bonds; plastocyanin; ribulose-1,5-bisphosphate carboxylase/oxygenase; structural arginine; triosephosphate isomerase; trp aporepressor; xylose isomerase

The guanidinium group of arginine is the most polar of all the common amino acid side chains found in proteins (Wolfenden, 1983) and is thus the residue most likely to be found on the surface in an aqueous environment. The side chain almost always has a $pK_a \geq 12$ and the positively charged planar guanidinium group can be a hydrogen donor in the formation of up to 5 hydrogen bonds (Fig. 1). Arginine plays a major role in the binding of negatively charged substrates, cofactors, and effectors to protein active sites (Riordan et al., 1977). In addition, it is often involved in the formation of singly or doubly hydrogen bonded salt bridges, usually with aspartate or glutamate side chains. These salt bridges stabilize the tertiary and quaternary structures of numerous proteins (Riordan et al., 1977) and are an important mode of interaction between different proteins (De Vos et al., 1992). Arginine is also reported frequently to hydrogen bond to other side-chain heteroatoms and to water molecules, but infrequently to backbone carbonyl oxygen atoms (Richardson & Richardson, 1989). No one has previously suggested that arginine may often play a general structural role in the maintenance of the native core structure of globular proteins.

Reprint requests to: Charles L. Borders, Jr., Department of Chemistry, The College of Wooster, Wooster, Ohio 44691; e-mail: borders@acs.wooster.edu.

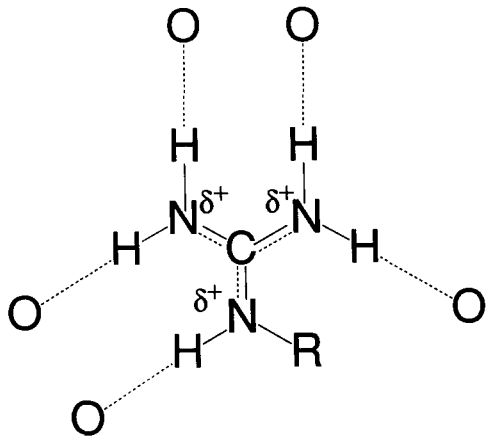

Fig. 1. Diagram of the planar protonated guanidinium group of an arginine showing 5 hydrogen bonds to 5 different oxygen atoms. The 3 nitrogen atoms are referred to as NE, NH1, and NH2 at various places in the paper. In this figure, NE is at the bottom, NH1 is at the top right above the R group, and NH2 is at the top left above the proton attached to NE.

We recently used data from chemical modification with group-specific reagents to determine that Arg 180 plays a key role in the catalytic function of most manganese-containing and iron-containing superoxide dismutases (Chan et al., 1990). We have used FRODO software on the Evans and Sutherland PS390 molecular graphics system to examine the local environment of Arg 180 in the 1.8-Å X-ray crystal structure of *Thermus thermophilus* manganese superoxide dismutase (Ludwig et al., 1991). Our initial report suggested that Arg 180 is in a highly unusual hydrogen bonding environment in which it forms 4 hydrogen bonds to 3 backbone carbonyl oxygen atoms, and we proposed that Arg 180 plays a combined structural and functional role at the active site of most manganese and iron superoxide dismutases (Chan et al., 1990).

We have used molecular graphics to further examine Arg 180 in *T. thermophilus* MnSOD (see below), as well as the localized environments of all arginines in the X-ray crystal structures of 20 other proteins. Using a criterion of 4 or more hydrogen bonds to 3 or more backbone carbonyl oxygens, we have located 3 additional "structural" arginines. On relaxing the criteria to 3 or more hydrogen bonds to 2 or more backbone carbonyl oxygens, we found an additional 14 arginines. We propose that arginine side chains often play a general structural role in the maintenance of protein tertiary structure. A similar role has recently been proposed for arginine in the specific interaction of peptides and proteins with target RNA sequences (Calnan et al., 1991). Preliminary accounts of some of the work reported in our paper have previously appeared (Bekeny et al., 1991; Borders et al., 1991; Broadwater et al., 1993).

Results

Thermus thermophilus manganese superoxide dismutase

We originally proposed a structural role for arginine in proteins from our study of the inactivation of manganese (and iron) superoxide dismutases by phenylglyoxal, an arginine-specific reagent (Borders et al., 1989; Chan et al., 1990). Arg 180 is the only highly conserved arginine in all manganese and iron superoxide dismutases sequenced to date (Chan et al., 1990). In the single exception, this arginine is replaced by lysine in the yeast enzyme. A preliminary examination (Chan et al., 1990) of the local environment of Arg 180 in the 1.8-Å structure of *T. thermophilus* manganese superoxide dismutase (Ludwig et al., 1991) suggested to us that Arg 180 may play a combined structural and functional role in the active site channel of this enzyme.

Kinemage 1 shows the hydrogen bonding interactions of Arg 180 in *T. thermophilus* manganese superoxide dismutase. Arg 180 protrudes from the N-terminus of one helix and makes 4 hydrogen bonds to 3 carbonyl oxygens at the C-terminus of a second helix. These 4 hydrogen bonds, between NH1 and NH2 of Arg 180 and the carbonyl oxygens of Ala 124, Arg 127, and Gly 129, range from 2.51 to 3.38 Å, whereas the dihedral angles range from 5° to 34° (Table 1). In addition, NE of Arg 180 makes a hydrogen bond with the hydroxyl oxygen of Tyr 172.

The active site of manganese superoxide dismutase is composed of residues from 2 different subunits. In the active center around the Mn ion from subunit B, Arg 180 is contributed by subunit A. View 2 of Kinemage 1 shows a part of the channel, with Arg 180A at the top in the opening view. The catalytic Mn is at the bottom of a deep channel where it is ligated by the side chains of His 28A, His 83A, His 170A, and Asp 166A, plus a water molecule. The path to the Mn ion is quite open, with only the side chains of His 32A and Tyr 36A protruding into the channel (Ludwig et al., 1991). Arg 180, with its extensive hydrogen bonding network, provides a diffuse positive potential on one wall of the channel that likely serves as an orienting force for the proper approach of O_2^- to the catalytic center. The side of the arginine guanidinium plane facing the active site channel is filled with water molecules (Kinemage 1, View 2), providing an easy access route for superoxide. We thus believe that Arg 180 in manganese superoxide dismutase plays both a structural role, wherein it helps to mold the active site channel, and a functional role, in which it helps to guide the anionic substrate to the catalytic center.

Human carbonic anhydrase II

An examination of the 2.0-Å resolution structure (Eriksson et al., 1988) of human carbonic anhydrase II reveals that Arg 254 fits our criteria for a structural arginine. Kinemage 2, View 1, shows the position of Arg 254 relative to the zinc ion in active site channel of the monomeric protein. The guanidinium moiety of Arg 254, which is more than 15 Å from the Zn(II), is shielded from the active site channel by the hydrophobic core of the protein. Interestingly, the positively charged guanidinium group points inward toward the interior of the protein so that the α-carbon of Arg 254 is closer to the surface than any other side-chain atom.

Figure 2 and Kinemage 2, View 2, show close-up views of the hydrogen bonding interactions of Arg 254. The guanidinium nitrogens form 5 hydrogen bonds to 4 backbone carbonyl oxygens contributed by Gln 28, Pro 195, Pro 250, and Leu 251. All hydrogen bonding distances are <3.1 Å and the dihedral angles between each carbonyl oxygen and the guanidinium plane are <10° (Table 1). The guanidinium group of Arg 254 penetrates into the interior of the protein and effectively acts as a staple to anchor 3 different polypeptide chains, represented by 3 different colors in Kinemage 2, View 2. The polar guanidinium

Table 1. *Hydrogen bonding interactions of backbone carbonyl oxygens to Arg 180 in T. thermophilus manganese superoxide dismutase, Arg 254 in human carbonic anhydrase II, Arg 31 in S. rubiginosus xylose isomerase, and Arg 313 in R. rubrum ribulose-1,5-bisphosphate carboxylase/oxygenase*[a]

Protein	Arg	Atom	Residue	Distance[b]	Dihedral angle[b]
Superoxide dismutase	180	NH2	Gly 129	3.27	31
		NH2	Arg 127	3.38	5
		NH1	Arg 127	2.78	6
		NH1	Ala 124	2.51	34
Carbonic anhydrase	254	NE	Leu 251	2.80	7
		NH2	Pro 250	2.91	8
		NH2	Gln 28	2.86	10
		NH1	Gln 28	3.11	9
		NH1	Pro 195	2.87	10
Xylose isomerase	31	NE	Trp 20	3.33	17
		NH2	Trp 20	2.73	22
		NH2	Pro 290	2.97	15
		NH1	Pro 291	2.82	4
		NH1	Glu 294	2.99	1
Ribulose-1,5-bisphosphate carboxylase/oxygenase	313	NE	Lys 361	2.96	23
		NH2	Met 360	2.81	22
		NH2	Gln 345	3.09	5
		NH1	Gln 345	2.94	5
		NH1	Met 342	2.69	11

[a] Arg 180 in superoxide dismutase also makes a hydrogen bond to the hydroxyl oxygen of Tyr 172. All hydrogen bonds are intrasubunit. The dihedral angle is N_a–C_z–$N_b \cdots O_c$, where N_a is 1 of the 2 remote nitrogens of the guanidinium plane, C_z is the guanidinium C-zeta, N_b is the guanidinium nitrogen that donates the hydrogen bond of interest, and O_c is the backbone carbonyl oxygen that serves as the hydrogen bond acceptor.

[b] Distance in Å; dihedral angles in degrees.

group is shielded on one face by van der Waals contacts with the side chains of Pro 195, Tyr 194, Val 31 and Leu 251, and Ile 256. The hydroxyl of Ser 29 seals off the bottom of the pocket, whereas the methylene carbons of Arg 27 and Lys 252 seal off the other. The only access to bulk solvent involves a small string of water molecules on one side of the guanidinium plane (Kinemage 2, View 2).

The crystal structure of human erythrocyte carbonic anhydrase I has also been reported (Kannan et al., 1984), and an examination of this structure shows that the Arg 254 in this isozyme makes the same hydrogen bonding interactions as found in carbonic anhydrase II. The Arg 254 side chains in the 2 structures are isostructural (data not shown). Significantly, Arg 254 is totally conserved in each of the 24 different carbonic anhydrases sequenced to date (Hewett-Emmett & Tashian, 1991). No role has previously been proposed for Arg 254 in carbonic anhydrase, but we believe that it plays a key role in maintaining the active structure of the native enzyme.

Streptomyces rubiginosus xylose isomerase

The X-ray structure of xylose isomerase from *S. rubiginosus* has been determined at 1.9 Å resolution (Carrell et al., 1989). Arg 31 in this enzyme easily meets the criteria we have established for a structural arginine, making 5 hydrogen bonds to 4 backbone carbonyl oxygens donated by Trp 20, Pro 290, Pro 291, and Glu 294 (Kinemage 3). The hydrogen bonding distances range from 2.82 to 3.33 Å, and all dihedral angles are ≤22° (Table 1). The guanidinium group is well removed from the active site, with its closest nitrogen atom 14.3 Å from one Mn ion and 15.9 Å away from the other (Kinemage 3). As with Arg 254 in carbonic anhydrase, the α-carbon of Arg 31 in xylose isomerase is the side-chain atom closest to the surface of the protein and the guanidinium group penetrates into the core (data not shown). One face of the guanidinium plane of Arg 31 makes van der Waals contacts with the side chains of Trp 20 and Thr 30, whereas the other face is shielded from solvent by carbon atoms donated by Pro 290, Glu 294, and Val 299. In addition, Phe 296 makes van der Waals contacts with the methylene groups of Arg 31 (Kinemage 3). As a consequence of all these close interactions, the side chain of Arg 31 is completely buried and inaccessible to solvent.

The 1.65-Å X-ray crystal structure of *S. albans* xylose isomerase has also been reported (Dauter et al., 1990), and an examination of this structure reveals that the equivalent arginine makes identical contacts (data not shown). This is not surprising, because the 2 enzymes have almost identical amino acid sequences (Dauter et al., 1990). More significantly, Arg 31 is conserved in all xylose isomerases sequenced to date (Lee et al., 1990; Dekker et al., 1991). These data collectively suggest that Arg 31 plays a key structural role in xylose isomerase.

Rhodospirillum rubrum ribulose-1,5-bisphosphate carboxylase/oxygenase

An examination of the 1.7-Å X-ray structure of *R. rubrum* ribulose-1,5-bisphosphate carboxylase/oxygenase (Schneider

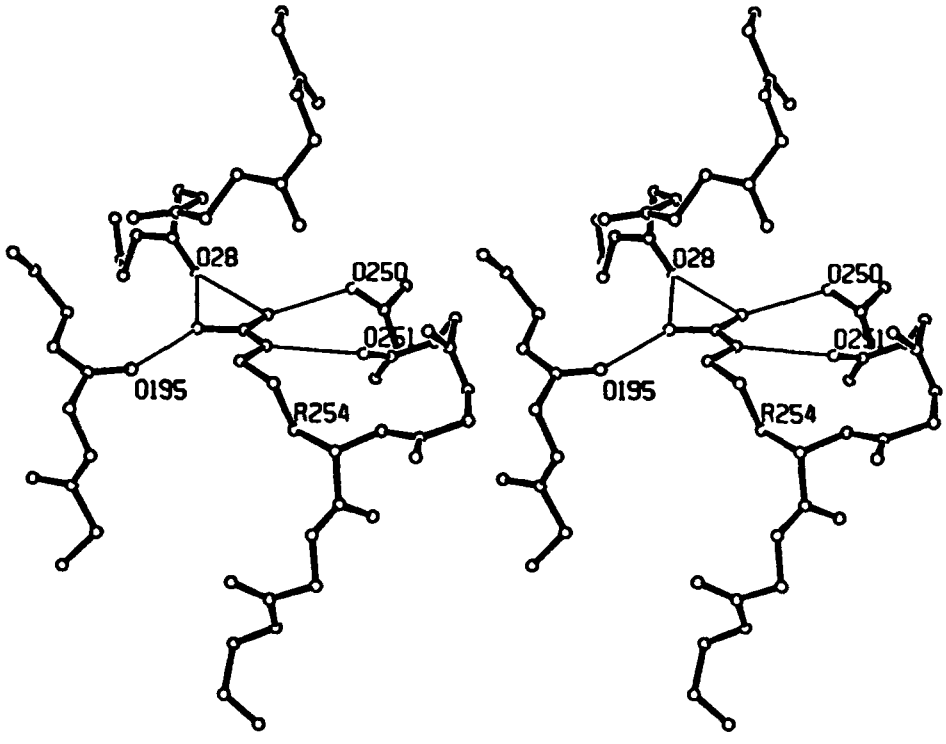

Fig. 2. Stereo view of the environment of the guanidinium group of Arg 254 in human carbonic anhydrase II. The backbone carbonyl oxygen atoms involved in hydrogen bonding are labeled, and the hydrogen bonds are shown as thin lines. The C_α of Arg 254 is closest to the surface of the protein, and the side chain is directed into the interior. The positively charged guanidinium group is almost completely buried.

et al., 1990) reveals that it also has a structural arginine that makes 5 hydrogen bonds to 4 backbone carbonyl oxygens. As shown in Kinemage 4, Arg 313 protrudes from helix $\alpha 5$ (Schneider et al., 1990) in the C-terminal domain of *R. rubrum* ribulose-1,5-bisphosphate carboxylase/oxygenase and interacts with the carbonyl oxygens of Met 342 and Gln 345 at the end of helix $\alpha 6$, as well those of Met 360 and Lys 361 in a β-turn. The hydrogen bond distances range from 2.69 to 3.09 Å, and all dihedral angles are $\leq 23°$ (Table 1). In addition, a fifth carbonyl oxygen, that of Leu 343, lies 3.28 Å from NH1 of Arg 313, but its dihedral angle of 68° makes it unlikely that a hydrogen bond is formed. Unlike the structural arginines in carbonic anhydrase and xylose isomerase (see above), the side chain of Arg 313 protrudes from inside the globular protein and its guanidinium moiety is directed toward, but barely reaches, the surface. The peptide backbone, which donates the hydrogen bond acceptors, plus the indole ring of Trp 357, which stacks ideally on one face of the planar guanidinium moiety, effectively shield most of the Arg 313 side chain from solvent (Kinemage 3). Only NH2 of Arg 313 extends into the exterior solvent, where it makes a hydrogen bond with a water molecule.

Although ribulose-1,5-bisphosphate carboxylase/oxygenase from most species is an L_8S_8 hexadecamer made up of 8 large (52-kDa) catalytic subunits and 8 small (14-kDa) subunits of unknown function (Miziorko & Lorimer, 1983), *R. rubrum* ribulose-1,5-bisphosphate carboxylase/oxygenase is a 104-kDa homodimer. However, the sequence of the *R. rubrum* 52-kDa polypeptide chain (Hartman et al., 1984) is approximately 25% homologous with the sequences of the large catalytic subunit of all L_8S_8 enzymes (Schneider et al., 1990). Significantly, Arg 313 and Trp 357 are conserved in all ribulose-1,5-bisphosphate carboxylases/oxygenases from bacteria, cyanobacteria, green and nongreen algae, and plants (Hudson et al., 1990; Gibson et al., 1991; Hwang & Tabita, 1991). Arg 313 appears to play a critical role in all of these enzymes, and the data reported herein suggest that the role is a structural one.

Other examples

We also found numerous other instances (Table 2) where a guanidinium moiety of an arginine forms 3 or more hydrogen bonds with 2 or more backbone carbonyls, but always well outside the active site of the protein. Two of the enzymes discussed above have additional arginines that make multiple hydrogen bonds to backbone carbonyl oxygens; *R. rubrum* ribulose-1,5-bisphosphate carboxylase/oxygenase has 2 other arginines that make 3 hydrogen bonds to 2 backbone carbonyls, whereas *S. rubiginosus* xylose isomerase has 1 additional structural arginine. In addition, we found that yeast triosephosphate isomerase has 3 arginines of interest and human lysozyme has 2, whereas yeast enolase, *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase, HIV-1 protease, bacteriophage T4 lysozyme, *Enteromorpha prolifera* plastocyanin, trypanosomal triosephosphate isomerase, and *Escherichia coli* trp aporepressor have 1 (Table 2). We found no evidence for a structural arginine in adenylate kinase, calmodulin, crambin, cytochrome c_3, dihydrofolate reductase, myoglobin, or pancreatic trypsin inhibitor.

Discussion

Arginine is well known to form hydrogen bonding salt bridges to anionic substrates, cofactors, and effectors at protein active sites (Riordan et al., 1977), as well as to form salt bridges with carboxylate side chains in the maintenance of protein tertiary structure. This latter role has been accepted for decades, but no

Table 2. *Hydrogen bonding interactions to backbone carbonyl oxygens of "structural" arginines in other proteins*[a]

Protein	Arg	Atom	Residue	Distance[b]	Dihedral angle[b]
Enolase (yeast)	391	NE	143	3.31	6
		NH2	143	3.12	6
		NH1	434	2.89	34
Glyceraldehyde-3-phosphate dehydrogenase (*B. stearothermophilus*)	183	NE	188	3.03	11
		NH2	188	3.32	9
		NH1	182	3.37	16
HIV-1 protease	87B	NE	5A	3.32	19
		NH2	5A	3.08	22
		NH2	6A	3.04	2
Lysozyme (human)	62	NE	71	2.84	11
		NH2	71	3.04	10
		NH2	49	2.34	18
		NH1	49	3.09	12
	5	NH2	124	3.16	3
		NH2	126	2.67	4
		NH1	126	2.89	4
Lysozyme (bacteriophage T4)	95	NE	153	2.69	15
		NH2	153	3.43	11
		NH2	154	3.30	24
Plastocyanin (*E. prolifera*)	72	NE	55	2.73	3
		NH2	56	2.89	6
		NH1	71	3.07	10
Ribulose-1,5-bisphosphate carboxylase/ oxygenase (*R. rubrum*)	92	NE	90	3.18	4
		NH2	90	2.64	4
		NH2	87	2.96	0
	152	NH2	159	2.95	8
		NH1	159	2.69	8
		NH1	157	2.99	17
Triosephosphate isomerase (*T. brucei brucei*)	99	NE	94	2.97	18
		NE	126	3.26	14
		NH2	126	2.64	20
Triosephosphate isomerase (yeast)	3	NE	202	2.35	3
		NH2	202	3.15	2
		NH1	224	3.12	4
	99	NE	94	2.88	35
		NE	126	3.34	14
		NH2	126	2.53	18
	247	NH2	220	2.56	17
		NH2	226	3.12	10
		NH1	226	3.45	9
Trp aporepressor (*E. coli*)	56A	NE	52A	2.93	6
		NH2	52A	2.58	6
		NH2	18B	3.13	18
Xylose isomerase (*S. rubiginosus*)	284	NH2	280	3.14	19
		NH1	280	3.01	20
		NH1	282	2.91	33

[a] The residue numbers of the arginine and the residue that donates the carbonyl oxygen are given, and interactions are intrasubunit except those where different letters are shown with the residue numbers. There are often extensive interactions with other protein or solvent heteroatoms that are not shown in this compilation. The dihedral angle is N_a–C_z–$N_b \cdots O_c$, where N_a is 1 of the 2 remote nitrogens of the guanidinium plane, C_z is the guanidinium C-zeta, N_b is the guanidinium nitrogen that donates the hydrogen bond of interest, and O_c is the backbone carbonyl oxygen that serves as the hydrogen bond acceptor.

[b] Distance in Å; dihedral angles in degrees.

one has systematically called attention to the interaction of arginine guanidinium groups with backbone carbonyl oxygen atoms as an important general interaction. Our study suggests that the frequency of occurrence of hydrogen bonds to backbone carbonyl oxygens is nearly equal to the frequency of hydrogen bonding salt bridges to carboxylates. For example, an examination of the structures listed in Table 3 (1CA2, 7ENL, 2RUS, 1TRO, and 6XIA were not examined in this detail) suggests that the 209 arginines in the proteins studied make 143 hydrogen bonds to carboxylates and 138 hydrogen bonds to backbone carbonyl oxygen atoms. Using our relaxed criterion of 3 or more hydrogen bonds to 2 or more backbone carbonyl oxygens, 18 of 209, or 8.6% of all arginines examined, may play an important structural role.

The best examples of structural arginines found in this work, Arg 254 in carbonic anhydrase, Arg 31 in xylose isomerase, and Arg 313 in ribulose-1,5-bisphosphate carboxylase/oxygenase, form their full complement of 5 hydrogen bonds to backbone carbonyl oxygens. All are more or less buried in the "mantle" (i.e., inside, but near the surface) of the protein interior, two with the side chain directed toward the core of the globular protein and one with the reverse orientation. In none of these cases is the arginine at the active site. The fact that each arginine of interest is fully conserved in numerous examples of each protein sequenced to date suggests that a structural arginine that displays this level of hydrogen bonding to backbone carbonyl oxygens may play an essential role in protein function.

Arg 180 in manganese superoxide dismutase is the only example uncovered in our study of a structural arginine in an active site channel. This arginine is conserved in all but one of the manganese and iron superoxide dismutase sequences reported to date; it is replaced by lysine in the yeast manganese enzyme (Ditlow et al., 1982). The fact that yeast superoxide dismutase has comparable activity and stability to those of other superoxide dismutases (Bjerrum, 1987) suggests that lysine, with or without some additional contributing interactions, may also be capable of playing a similar structural role. In support of this possibility, we found that Arg 3 in yeast triosephosphate isomerase is conserved in all other triosephosphate isomerases except that from *Trypanosoma brucei brucei*, in which it is replaced by lysine (Fothergill-Gilmore & Michels, 1993). We examined the local environment of Lys 3 in the *T. brucei brucei* isomerase and found that it indeed makes 3 hydrogen bonds to 3 backbone carbonyl oxygens, although its partners are oxygens other than those that form hydrogen bonds to Arg 3 in the yeast enzyme (J.A. Broadwater & C.L. Borders, Jr., unpubl. results). We found several examples of buried lysines in other proteins that make 3 or more hydrogen bonds to 2 or more backbone

Table 3. *Information on protein structures examined in this study*

Protein	PDB code	Resolution, Å (*R*-factor)	Reference
Adenylate kinase (porcine)	3ADK	2.1 (0.193)	Dreusicke et al., 1988
Calmodulin (*D. melanogaster*)	4CLN	2.2 (0.197)	Taylor et al., 1991
Carbonic anhydrase I (human)	1CAB	2.0 (0.193)	Kannan et al., 1984
Carbonic anhydrase II (human)	1CA2	2.0 (0.176)	Eriksson et al., 1988
Crambin	1CRN	1.5 (0.114)	Teeter and Whitlow, 1986
Dihydrofolate reductase (chicken liver)	8DFR	1.7 (0.188)	Matthews et al., 1985
Enolase (yeast)	6ENL	2.2 (0.157)	Lebioda et al., 1991
Enolase (yeast)	7ENL	2.2 (0.169)	Lebioda and Stec, 1991
Glyceraldehyde-3-phosphate dehydrogenase (*B. stearothermophilus*)	1GD1	1.8 (0.177)	Skarzynski et al., 1987
HIV-1 protease	4HVP	2.3 (0.176)	Miller et al., 1989
Lysozyme (human)	1LZ1	1.5 (0.177)	Artymiuk and Blake, 1981
Lysozyme (bacteriophage T4)	3LZM	1.7 (0.157)	Weaver and Matthews, 1987
Manganese superoxide dismutase (*T. thermophilus*)	—[a]	1.8 (0.176)	Ludwig et al., 1991
Myoglobin (sperm whale)	1MBD	1.4 (N.A.)	Phillips and Schoenborn, 1981
Plastocyanin (*E. prolifera*)	7PCY	1.8 (0.117)	Collyer et al., 1990
Ribulose-1,5-bisphosphate carboxylase/oxygenase (*R. rubrum*)	5RUB	1.7 (0.180)	Schneider et al., 1990
Ribulose-1,5-bisphosphate carboxylase/oxygenase (*R. rubrum*)	2RUS	2.3 (0.193)	Lundqvist and Schneider, 1991
Triosephosphate isomerase (yeast)	1YPI	1.9 (0.210)	Lolis et al., 1990
Triosephosphate isomerase (*T. brucei brucei*)	5TIM	1.83 (0.183)	Wierenga et al., 1991
Trp aporepressor (*E. coli*)	3WRP	1.8 (0.204)	Zhang et al., 1987
Trp repressor (*E. coli*)	2WRP	1.65 (0.180)	Schevitz et al., 1985
Trp repressor/operator complex (*E. coli*)	1TRO	1.9 (0.167)	Otwinowski et al., 1988
Trypsin inhibitor (bovine pancreas)	5PTI	1.0 (0.200)	Wlodawer et al., 1987
Xylose isomerase (*S. albans*)	6XIA	1.65 (0.141)	Dauter et al., 1990
Xylose isomerase (*S. rubiginosus*)	8XIA	1.9 (0.140)	Carrell et al., 1989

[a] Coordinates kindly provided by Prof. M.L. Ludwig.

carbonyl oxygens, i.e., they appear to be "structural" lysines. Our best example is Lys 263 in yeast enolase, which makes 5 hydrogen bonds to 4 backbone carbonyl oxygens. Interestingly, Lys 263 is fully conserved in all enolase sequences (Fothergill-Gilmore & Michels, 1993), suggesting that its structural role may be critical.

We realize that the exact criteria for what constitutes a "structural" arginine are debatable. Using our initial criterion of 3 or more hydrogen bonds to 2 or more backbone carbonyl oxygens, we found 18 arginines that might play a structural role (Tables 1, 2). However, if the role was critical, it would seem that the structural arginine would be conserved through evolutionary time. We note that indeed many of the arginines of interest that make only 3 hydrogen bonds to 2 backbone carbonyl oxygens are not conserved, i.e., Arg 391 in yeast enolase (Fothergill-Gilmore & Michels, 1993), Arg 183 in *B. stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase (Fothergill-Gilmore & Michels, 1993), and Arg 72 in *E. prolifera* plastocyanin. However, other arginines that have the same hydrogen bonding pattern are highly conserved, i.e., Arg 99 in triosephosphate isomerase (Fothergill-Gilmore & Michels, 1993). When the criterion is narrowed to 4 or more hydrogen bonds to 3 or more backbone carbonyl oxygens, every example located in this study involves a highly (Arg 180 in manganese superoxide dismutase) or fully conserved arginine.

The work reported herein may offer some explanation of the report by Mrabet and coworkers (1992) that Lys to Arg mutations in a number of enzymes lead to enhanced thermal stability. Arginine may be better able to interact with backbone carbonyl oxygens, providing added structural rigidity to the protein. Mrabet et al. (1992) report X-ray crystallographic and model-building data on several of their mutants that suggest that several of the engineered arginines indeed interact with backbone carbonyl oxygens. We also note along these lines that Arg 284 (Table 2) is conserved in all thermostable xylose isomerases, but not in isomerases of lower thermal stability.

The variety of interactions with hydrogen bonding partners suggests that there is no single structural motif for a structural arginine. We found examples where the structural arginine ties α-helix to α-helix, β-turn to β-turn, and subunit to subunit. In the 3 cases where the structural arginine forms hydrogen bonds to backbone carbonyl oxygens at the C-terminus of an α-helix (Arg 180 in manganese superoxide dismutase, Arg 313 in ribulose-1,5-bisphosphate carboxylase/oxygenase, and Arg 95 in bacteriophage T4 lysozyme), it hydrogen bonds to the first oxygen not involved in the helix hydrogen bonding network. The positively charged guanidinium group thus strongly interacts with the negative end of the helix dipole. Arg 254 in carbonic anhydrase plays an interesting role of stapling together 3 strands of backbone from different regions of the primary sequence. Such shallow buried interactions of polar partners are likely to influence the overall electrostatic potential of the protein surface.

One simple model for the arginine/backbone carbonyl oxygen interactions that are the focus of our work is the well-known "crown ether" complex formed between macrocyclic oligooxy ligands and inorganic cations. However, instead of the hard spherical cation of inorganic metal ions, the cation of interest in our model is the planar guanidinium moiety of an arginine, which is tethered to the polypeptide backbone. This guanidinium group has a delocalized positive charge and is capable of donating up to 5 hydrogen bonds. The "macrocycle" is made up of several peptide carbonyl oxygen atoms that may originate from various parts of the polypeptide backbone, but which come together when the protein nears its native conformation and are locked in place by their interaction with the guanidinium moiety. These interactions may lend added stability to any protein.

Materials and methods

The atomic coordinates for the proteins examined were obtained from the Protein Data Bank (Bernstein et al., 1977; Abola et al., 1987) at Brookhaven National Laboratory. The criteria for this random sampling of proteins were high resolution ($\leq$2.3 Å with an R-factor $\leq$0.21) and a known primary sequence. Coordinates for *T. thermophilus* manganese superoxide dismutase at 1.8 Å resolution were kindly provided by Dr. Martha L. Ludwig before publication (Ludwig et al., 1991). Pertinent data on protein structures examined in this study are given in Table 3. Structures were viewed with an Evans and Sutherland PS390 graphics terminal using FRODO software (Jones, 1978).

A diagram for the 5 possible hydrogen bonding interactions for the positively charged guanidinium group of arginine is shown in Figure 1. Oxygen atoms within 3.5 Å of one of the guanidinium nitrogen atoms and with a dihedral angle (N_a-C_z-$N_b\cdots O_c$, where N_a is one of the 2 remote nitrogens of the guanidinium plane, C_z is the guanidinium C-zeta, N_b is the guanidinium nitrogen that donates the hydrogen bond of interest, and O_c is the backbone carbonyl oxygen that serves as the hydrogen bond acceptor) of 35° or less (Artymiuk & Blake, 1981) were considered possible hydrogen bond acceptors. In deducing the likely hydrogen bonding scheme, it was assumed (Baker & Hubbard, 1984) that the C_z-$N_b\cdots O_c$ angle would average 120 $\pm$ 30°.

Acknowledgments

This work was taken in part from the Senior Independent Study theses of P.A.B (1991) and A.M.E. (1993), The College of Wooster. J.E.S. (1991) and A.S.L. (1993) received summer support from Bristol-Myers Company grant C2890 of the Research Corporation to C.L.B. J.A.B. worked on this project during the summer of 1992 with support from an NSF Research Experiences for Undergraduates grant. V.B.P. was supported by NSF Career Advancement Award MCB-9109676. We acknowledge a Howard Hughes Medical Institute grant to The College of Wooster that provided funding for the Evans and Sutherland molecular graphics system. This research was also supported in part by Pittsburgh Supercomputing Center grant 1-P41-RR06009 from the NIH National Center for Research Resources. We thank Dr. William C. Stallings for helpful suggestions on the manuscript.

References

Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. 1987. Protein Data Bank. In: Allen FH, Bergerhoff G, Sievers R, eds. *Crystallographic databases—Information content, software systems, scientific applications*. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography. pp 107–132.

Artymiuk PJ, Blake CCF. 1981. Refinement of human lysozyme at 1.5 Å resolution: Analysis of non-bonded and hydrogen-bonded interactions. *J Mol Biol 152*:737–762.

Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Progr Biophys Mol Biol 44*:97–179.

Bekeny PA, Borders CL Jr, Pett VB. 1991. An unusual arginine in manganese superoxide dismutase: A model for structural arginine residues in proteins. *EUREKA Fifth National Conference on Undergraduate Research, Pasadena, California, June 1991*. Abstr BCM.1.3B.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535-542.

Bjerrum MJ. 1987. Structural and spectroscopic comparison of manganese-containing superoxide dismutases. *Biochim Biophys Acta 915*:225-237.

Borders CL Jr, Bekeny PA, Pett VB. 1991. A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Fifth Symposium of The Protein Society, Baltimore, Maryland, June 1991*. Abstr S161.

Borders CL Jr, Horton PJ, Beyer WF Jr. 1989. Chemical modification of iron- and manganese-containing superoxide dismutases from *Escherichia coli*. *Arch Biochem Biophys 268*:74-80.

Broadwater JA, Borders CL Jr, Pett VB. 1993. A structural role for arginine in proteins. *EUREKA Seventh National Conference on Undergraduate Research, Salt Lake City, Utah, March 1993*. Abstr 6-03-1.

Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD. 1991. Arginine-mediated RNA recognition: The arginine fork. *Science 252*:1167-1171.

Carrell HL, Glusker JP, Burger V, Manfre F, Tritsch D, Biellmann JF. 1989. X-ray analysis of D-xylose isomerase at 1.9 Å: Native enzyme in complex with substrate and with a mechanism-designed inactivator. *Proc Natl Acad Sci USA 86*:4440-4444.

Chan VWF, Bjerrum MJ, Borders CL Jr. 1990. Evidence that chemical modification of a positively charged residue at position 189 causes the loss of catalytic activity of iron-containing and manganese-containing superoxide dismutases. *Arch Biochem Biophys 279*:195-201.

Collyer CA, Guss JM, Sugimura Y, Yoshizaki F, Freeman HC. 1990. The crystal structure of plastocyanin from a green alga, *Enteromorpha prolifera*. *J Mol Biol 211*:617-632.

Dauter Z, Terry H, Witzel H, Wilson KS. 1990. Refinement of glucose isomerase from *Streptomyces albus* at 1.65 Å with data from an imaging plate. *Acta Crystallogr B 46*:833-841.

Dekker K, Yamagata H, Sakuguchi K, Udaka S. 1991. Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus*: Cloning, sequencing, and comparison with other thermostable xylose isomerases. *J Bacteriol 173*:3078-3083.

De Vos AM, Ultsch M, Kosiakoff AA. 1992. Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science 255*:306-312.

Ditlow C, Johansen JT, Martin BM, Svendsen I. 1982. The complete amino acid sequence of manganese-superoxide dismutase from *Saccharomyces cerevisiae*. *Carlsberg Res Commun 47*:81-91.

Dreusicke D, Karplus PA, Schulz GE. 1988. Refined structure of porcine cytosolic adenylate kinase at 2.1 Å resolution. *J Mol Biol 199*:359-371.

Eriksson AE, Jones TA, Liljas A. 1988. Refined structure of human carbonic anhydrase II at 2.0 Å resolution. *Proteins Struct Funct Genet 4*:274-282.

Fothergill-Gilmore LA, Michels PAM. 1993. Evolution of glycolysis. *Progr Biophys Mol Biol 59*:105-235.

Gibson JL, Falcone DL, Tabita FR. 1991. Nucleotide sequence, transcriptional analysis, and expression of genes encoded within the form I CO_2 fixation operon of *Rhodobacter sphaeroides*. *J Biol Chem 266*:14646-14653.

Hartman FC, Stringer CD, Lee EA. 1984. Complete primary structure of ribulosebisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum*. *Arch Biochem Biophys 232*:280-295.

Hewett-Emmett D, Tashian RE. 1991. Structure and evolutionary origins of the carbonic anhydrase multigene family. In: Dodgson SJ, Tashian RE, Gros G, Carter ND, eds. *The carbonic anhydrases: Cellular physiology and molecular genetics*. New York: Plenum Press. pp 15-32.

Hudson GS, Mahon JD, Anderson PA, Gibbs MJ, Badger MR, Andrews TJ, Whitfeld PR. 1990. Comparisons of *rbcL* genes for the large subunit of ribulosebisphosphate carboxylase from closely related C_3 and C_4 plant species. *J Biol Chem 265*:808-814.

Hwang SR, Tabita FR. 1991. Cotranscription, deduced primary structure, and expression of the chloroplast-encoded *rcbL* and *rcbS* genes of the marine diatom *Cylindrotheca* sp. strain N1. *J Biol Chem 266*:6271-6279.

Jones TA. 1978. A graphics model building and refinement system for macromolecules. *J Appl Crystallogr 11*:268-272.

Kannan KK, Ramanadham M, Jones TA. 1984. Structure, refinement, and function of carbonic anhydrase isozymes: Refinement of human carbonic anhydrase I. *Ann NY Acad Sci 429*:49-60.

Lee C, Bagdasarian M, Meng M, Zeikus JG. 1990. Catalytic mechanism of xylose (glucose) isomerase from *Clostridium thermosulfurogenes*: Characterization of the structural gene and function of active site histidine. *J Biol Chem 265*:19082-19090.

Lebioda L, Stec B. 1991. Mechanism of enolase: The crystal structure of enolase-Mg^{2+}-2-phosphoglycerate/phosphoenolpyruvate complex at 2.2-Å resolution. *Biochemistry 30*:2817-2822.

Lebioda L, Stec B, Brewer JM, Tykarska E. 1991. Inhibition of enolase: The crystal structures of enolase-Ca^{2+}-2-phosphoglycerate and enolase-Zn^{2+}-phosphoglycolate complexes at 2.2-Å resolution. *Biochemistry 30*:2823-2827.

Lolis E, Alber T, Davenport RC, Rose D, Hartman FC, Petsko GA. 1990. Structure of yeast triosephosphate isomerase at 1.9-Å resolution. *Biochemistry 29*:6609-6618.

Ludwig ML, Metzger AL, Pattridge KA, Stallings WC. 1991. Manganese superoxide dismutase from *Thermus thermophilus*: A structural model refined at 1.8 Å resolution. *J Mol Biol 219*:335-358.

Lundqvist T, Schneider G. 1991. Crystal structure of the ternary complex of ribulose-1,5-bisphosphate carboxylase, Mg(II), and activator CO_2 at 2.3 Å-resolution. *Biochemistry 30*:904-908.

Matthews DA, Bolin JT, Burridge JM, Filman DJ, Volz KW, Kaufman BT, Beddell CR, Champness JN, Stammers DK, Kraut J. 1985. Refined crystal structure of *Escherichia coli* and chicken liver dihydrofolate reductase containing bound trimethoprim. *J Biol Chem 260*:381-391.

Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SBH, Wlodawer A. 1989. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science 246*:1149-1152.

Miziorko HM, Lorimer GH. 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu Rev Biochem 52*:507-535.

Mrabet NT, Van den Broeck A, Van den Brande I, Stanssens P, Laroche Y, Lambeir AM, Matthijssens G, Jenkins J, Chiadmi M, van Tilbeurgh H, Rey F, Janin J, Quax WJ, Lasters I, De Maeyer M, Wodak SJ. 1992. Arginine residues as stabilizing elements in proteins. *Biochemistry 31*:2239-2253.

Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi BF, Sigler PB. 1988. Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature 335*:321-329.

Phillips SEV, Schoenborn BP. 1981. Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. *Nature 292*:81-82.

Richardson JS, Richardson DC. 1989. Principles and patterns of protein conformation. In: Fasman GD, ed. *Prediction of protein structure and the principles of protein conformation*. New York: Plenum Press. pp 1-98.

Riordan JF, McElvany KD, Borders CL Jr. 1977. Arginyl residues: Anion recognition sites in enzymes. *Science 195*:884-886.

Schevitz RW, Zdzislaw O, Joachimiak A, Lawson CL, Sigler PB. 1985. The three-dimensional structure of trp repressor. *Nature 317*:782-786.

Schneider G, Lindqvist Y, Lundqvist T. 1990. Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from *Rhodospirillum rubrum* at 1.7 Å resolution. *J Mol Biol 211*:989-1008.

Skarzynski T, Moody PCE, Wonacott AJ. 1987. Structure of hologlyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J Mol Biol 193*:171-187.

Taylor DA, Sack JS, Maune JF, Beckingham K, Quiocho FA. 1991. Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution. *J Biol Chem 266*:21375-21380.

Teeter MM, Whitlow MD. 1986. Hydrogen bonding in the high resolution structure of the protein crambin. *Trans Am Crystallogr Assoc 22*:75-88.

Weaver LH, Matthews BW. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J Mol Biol 193*:189-199.

Wierenga RK, Noble MEM, Vriend G, Nauche S, Hol WGJ. 1991. Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulfate: A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. *J Mol Biol 220*:995-1015.

Wlodawer A, Diesenhofer J, Huber R. 1987. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J Mol Biol 193*: 145-156.

Wolfenden R. 1983. Waterlogged molecules. *Science 222*:1087-1093.

Zhang RG, Joachimiak A, Lawson CL, Schevitz RW, Otwinowski Z, Sigler PB. 1987. The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature 327*:591-597.