Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2473–2477, March 1994
Biochemistry

Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold

(metalloprotease/protein structure/proteolytic enzyme/peptidase/homology)

J. F. BAZAN*, L. H. WEAVER†, S. L. RODERICK†‡, R. HUBER§, AND B. W. MATTHEWS†¶

†Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, OR 97403; *DNAX Research Institute of Molecular and Cellular Biology, 901 California Avenue, Palo Alto, CA 94304; and §Max-Planck-Institut für Biochemie, W-8033 Martinsreid bei München, Germany

Contributed by B. W. Matthews, November 18, 1993

ABSTRACT **Amino acid sequence comparison suggests that the structure of *Escherichia coli* methionine aminopeptidase (EC 3.4.11.18) and the C-terminal domain of *Pseudomonas putida* creatinase (EC 3.5.3.3) are related. A detailed comparison of the three-dimensional folds of the two enzymes confirms this homology: within an ≈260-residue chain segment, 218 C^{α} atoms of the structures superimpose within 2.5 Å; only 41 of these overlapping positions (i.e., 19%) feature identical amino acids in the two protein chains. Notwithstanding this striking correspondence in structure, methionine aminopeptidase binds and is stimulated by Co^{2+}, while creatinase is not a metal-dependent enzyme. Searches of protein data banks using sequence and structure-based profiles reveal other enzymes, including aminopeptidase P (EC 3.4.11.9), prolidase (EC 3.4.13.9), and agropine synthase, that likely share the same "pita-bread" fold common to creatinase and methionine aminopeptidase.**

A limited number of folding motifs can be used to describe the three-dimensional architecture of globular proteins (1, 2). As a consequence, structural families composed of related protein folds may include diverse amino acid sequences with faint or undiscernible similarity (3, 4). This manifold nature of protein structure is exemplified by enzymes whose biological function depends on the conservation of a precise constellation and geometry of catalytic residues in the midst of a divergent protein scaffold. Two examples are provided by the families of α/β-barrel enzymes (5) and α/β hydrolases (6). The present work defines the essential structural framework and variant catalytic functions of an enzyme fold that was first observed in the three-dimensional structure of creatine amidinohydrolase from *Pseudomonas putida* (7); this "creatinase" (EC 3.5.3.3) hydrolyzes creatine to urea plus sarcosine (Table 1), and a mechanism of action has been proposed (8). We show that the larger domain present in creatinase displays a fold that is strikingly similar to the recently determined structure of *Escherichia coli* methionine aminopeptidase (AMPM; EC 3.4.11.18), an enzyme that cleaves the amino-terminal methionine from protein chains (9). In addition, this structural family is predicted to encompass other enzymes with different specificities (summarized in Table 1) including aminopeptidase P (AMPP) (i.e., proline aminopeptidase; EC 3.4.11.9) and proline dipeptidase (or "prolidase"; EC 3.4.13.9) (10, 11).

METHODS AND RESULTS

Sequence Comparisons. Independent of knowledge of any structural relationship between creatinase and AMPM, protein and DNA data bases were first queried for sequence relatives of aminopeptidases by using BLAST (12). CLUSTAL (13) was then used to align the multiple related enzyme chains. Profile methods (14) including the modifications of Gibson *et al.* (15) were then used to rescreen the sequence data banks for more distant homologs of the assembled aminopeptidases. Fig. 1 confirms the previously reported relationship between the AMPMs of *E. coli, Salmonella typhimurium, Bacillis subtilis*, and *Saccharomyces cerevisiae* (16). In addition, the AMPM sequences are significantly related to two other proteolytic enzyme types: AMPP from *E. coli* and *Streptomyces lividans* and an AMPP-like open reading frame from *Mycobacterium tuberculosis* (15–19), and the prolidases of human and *E. coli* origin (20, 21). Closely grouped with the collection of aminopeptidases and prolidases was a sequence from rat labeled as a eukaryotic initiation factor 2 (eIF-2)-associated 67-kDa protein (or "p67") (22). Apparently p67 is not recognized as a proteolytic enzyme and instead is thought to promote eukaryotic protein synthesis by protecting the eIF-2 α-subunit from inhibitory phosphorylation by eIF-2 kinases (27, 28). A gene fragment cloned from the thermophilic archaebacterium *Methanothermus fervidus* (23) further resembles the rat p67 sequence.

Table 1. Specificities of hydrolytic enzymes

Enzyme	Substrate
Methionine aminopeptidase	H-Met↓-Xaa-Yaa- . . .
Aminopeptidase P	H-Xaa↓-Pro-Yaa- . . .
Prolidase	H-Xaa↓-Pro-OH
Creatinase	H_2N-C(=NH)↓-N(CH_3)-CH_2-COOH

Early in the alignment of the aminopeptidase/p67 sequence family, it was possible to determine, both by the patterns of conserved residues and the resemblance of half-chain profiles, that the common ≈260-amino acid catalytic domain in aminopeptidase enzymes and p67 was composed of a tandem set of related ≈130-residue modules (Fig. 1). We note that the resultant domain division of prolidase is not matched by an equivalent genetic division—introns (29) that map to either domain do not align (not shown). This permitted the assembly of two new smaller profiles (14, 15) centered on the N-terminal module sequence (K/R)PG and a more C-terminal motif [(K/R/Q/E)XG(D/M)*ffff*(D/E)(P/V/A)(G/M/T),

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

Abbreviations: AMPM, methionine aminopeptidase; AMPP, aminopeptidase P; eIF-2, eukaryotic initiation factor 2.

¶To whom reprint requests should be addressed.
‡Present address: Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461-1602.

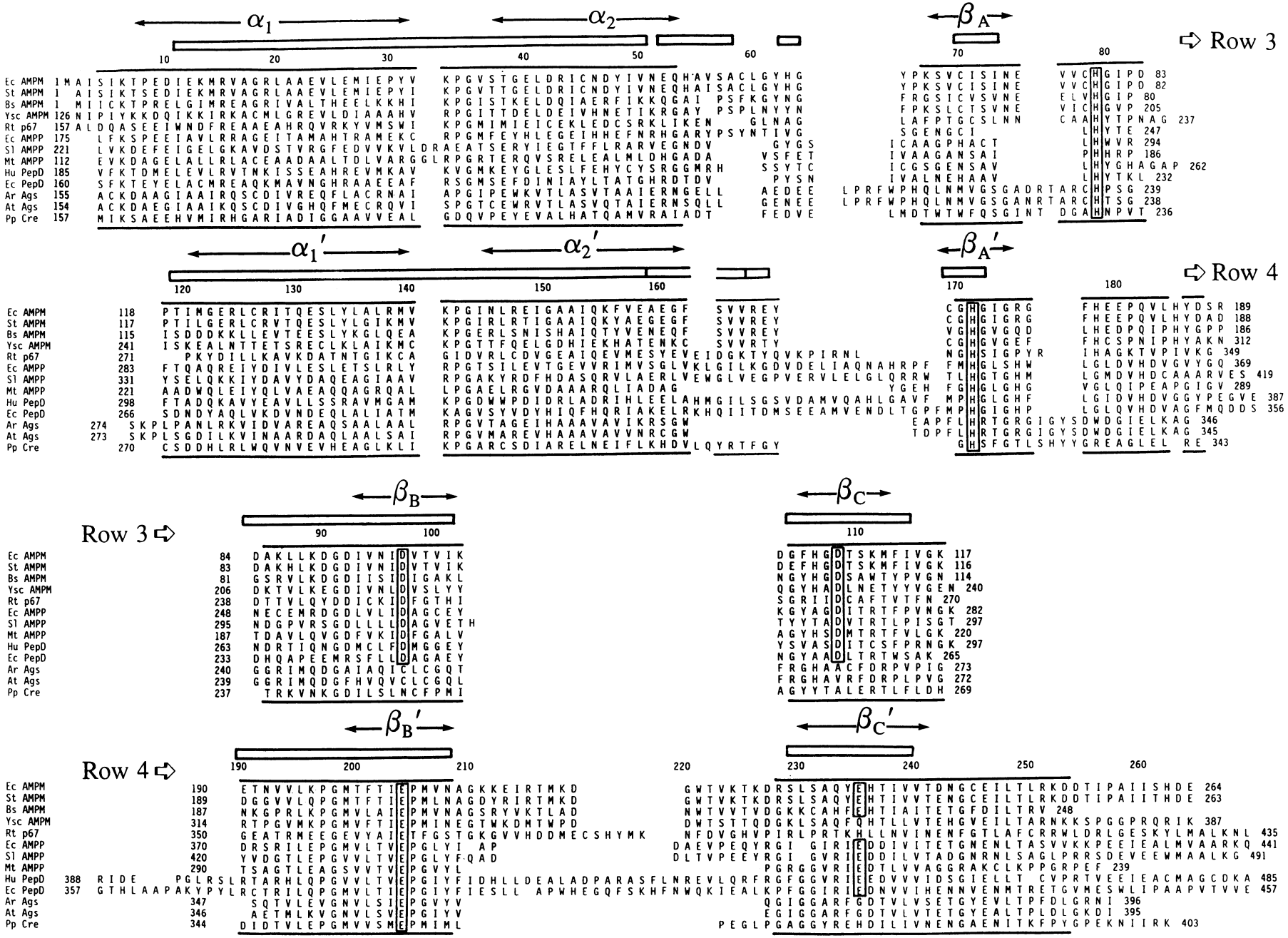

FIG. 1. (*Legend appears at the bottom of the opposite page.*)

where *f* denotes a hydrophobic residue]. Unexpectedly, these new profiles located a similar twofold modular pattern in the protein chains of creatinases from *P. putida* and *Flavobacterium* sp. (7, 26), along with distantly related agropine synthase sequences from *A. rhizogenes* and *A. tumefaciens* (24, 25), that could be integrated into the aminopeptidase/p67 fold (Fig. 1).

Intriguingly, neither creatinase nor agropine synthase functions as a peptide hydrolase; instead creatinase can be considered as a prototype of a class of amidinohydrolases that help catalyze the degradation of guanidino and ureido compounds in bacteria present in soil and water (30). On the other hand, agropine [(1′-deoxy-D-mannitol-1′-yl)-1-L-glutamine 1′,2′-lactone] is an unusual metabolite produced after infection by the crown gall tumor-inducing *Agrobacterium*; it can be metabolized by the parasite but not the host and thereby diverts nitrogen from the latter to the former (31, 32). Accordingly, agropine synthase controls the lactonization of the conjugate of deoxymannitol with glutamine to produce agropine (33).

Structural Resemblance. The relationship of creatinase to the larger family of aminopeptidases and p67 was illuminating because the three-dimensional x-ray structure of the dimeric *P. putida* enzyme had been determined to 1.9-Å resolution (7, 8) and likely could serve as the model for this nascent structural family. The $\alpha+\beta$ creatinase structure can be divided into a nonenzymatic N-terminal domain (residues 1–160) that contains a $\beta\beta\beta\alpha\beta\alpha\beta$ folding motif analogous to the component domains of actin/HSC70/hexokinase enzymes (34), and a C-terminal domain (residues 161–403) that contains a twice-repeated $\alpha\alpha\beta\beta\beta$ structure (7). Each of the latter folding units can be described as a helix hairpin that is packed against a three-strand sheet or β-meander (35). These protein modules associate sheet-to-sheet, with approximate dyad symmetry, to form a half-barrel pita-bread-like structure that features a deep groove at the open end where the inhibitor carbamoyl sarcosine is observed to bind (7, 8). By using the procedure of Rossmann and Argos (36, 37), 85 pairs of C^α atoms from each repeat module can be superimposed with a root-mean-square (rms) deviation of 2.7 Å (Figs. 1–3).

The next example of this fold was revealed in the 2.4-Å structure of the monomeric *E. coli* AMPM (9). This AMPM consists of a single "pita-bread" domain (Fig. 2) that is quite distinct from the leucine aminopeptidase/carboxypeptidase fold (38). In this case, the structural similarity between the symmetric modules is striking: 85 C^α positions superimpose (36, 37) with an rms deviation of 1.8 Å. Nestled in the binding cavity are two cobalt ions that presumably mark the active site of the enzyme (9). The structural correspondence of the creatinase and AMPM catalytic domains, indicated in Fig. 1, is exactly as suggested from the initial sequence alignments. A rigid-body comparison of the two enzymes (36, 37) shows that there are 218 C^α atoms in the structures that superimpose within 2.5 Å. At 41 of these structurally equivalent sites, the amino acid in both protein chains is the same (i.e., 19% sequence identity).

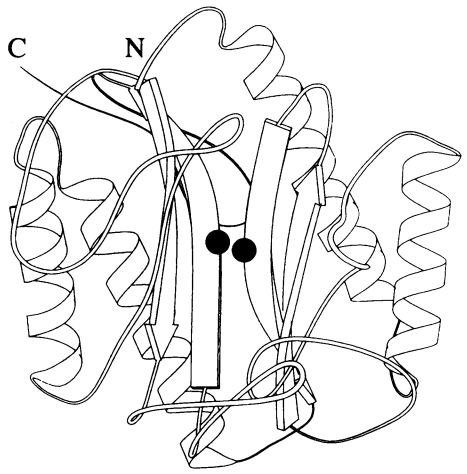

FIG. 2. Ribbon diagram showing the "pita-bread" fold of *E. coli* AMPM (9). The view direction is essentially parallel to the local twofold axis of symmetry and the active site is marked by the two cobalt ions shown as solid circles.

DISCUSSION

Perhaps the most striking result of this analysis is the remarkable similarity between the backbone of AMPM and the C-terminal domain of creatinase (Figs. 1–3). Both display the same type of internal twofold structural symmetry present in the "pita-bread" fold. The fact that 218 C^α atoms superimpose within 2.5 Å and 19% of these residues are identical in the two structures is compelling evidence that the two proteins evolved from a common precursor (cf. refs. 37 and 39). The reactions catalyzed by the two enzymes (Table 1) have elements in common but are not immediately related. Perhaps reflecting this dichotomy, the presumed active sites of these enzymes have some residue types in common, but they also have striking differences (Fig. 4*A*). In particular, AMPM binds two cobalt ions, whereas creatinase is not a metalloenzyme. In AMPM, the cobalt ions are coordinated by Asp-97, Asp-108, His-171, Glu-204, and Glu-235. In creatinase, the structurally equivalent residues, Asn-249, Ala-260, His-324, Glu-358, and His-376, are substantially different.

On the basis of the crystal structures of creatinase in complexes with various inhibitors and substrate analogs, His-232 is thought to be the key catalytic residue, acting as a general base and acid, and as a proton shuttle (8); AMPM also has a histidine (His-79) at the equivalent site (Figs. 1 and

Fig. 1. Summary of overall correspondence of sequence and structure. The AMPM sequences include enzymes from *E. coli* (Ec), *S. typhimurium* (St), *B. subtilis* (Bs), and yeast (Ysc) (16). AMPP sequences are available from *E. coli* HB101 (Ec) (17) and *Streptomyces lividans* (Sl) (18) and are inferred from *Mycobacterium tuberculosis* (Mt) (19). Prolidase sequences (PepD) are obtained from human (Hu) (20) and *E. coli* (Ec) (21). The p67 sequence is from rat (Rt p67) (22) and a related open reading frame sequence fragment from *Methanothermus fervidus* (23). Agropine synthase (Ags) sequences are from the closely related *Agrobacterium tumefaciens* (At) and *Agrobacterium rhizogenes* (Ar) (24, 25). The sequence of creatinase from *P. putida* (Pp Cre) (7) is substantially identical to the *Flavobacterium* sequence (26), which is not shown. The numbering at the top of the alignment corresponds to the *E. coli* AMPM sequence (9). The locations of the α-helices and β-strands in the N-terminal module of AMPM are denoted, respectively, α_1 and α_2 and β_A, β_B, and β_C; the corresponding C-terminal elements are labeled α_1' and α_2' and β_A', β_B', and β_C'. The locations of the secondary structure in creatinase are similar although not necessarily identical. The overall alignment of the sequences is as described in the text except for some adjustments in the vicinity of residues 55–85 of AMPM to bring the sequence alignment into register with the structural superposition (see following). The lines above the AMPM sequence and below the creatinase sequence indicate residues that are structurally equivalent (see text). The open bars above the sequences indicate residues in the N-terminal and the C-terminal halves of AMPM that are structurally related by an internal twofold axis of symmetry (e.g., residues 11–50 match with 119–158; see text). (A very similar relationship, not shown, applies to creatinase.) Residues that are cobalt ligands in the structure of *E. coli* AMPM, and are conserved in other sequences, are boxed. His-232, the critical histidine in *P. putida* creatinase (7, 8) and conserved histidines at the same position in the other sequences are also boxed.

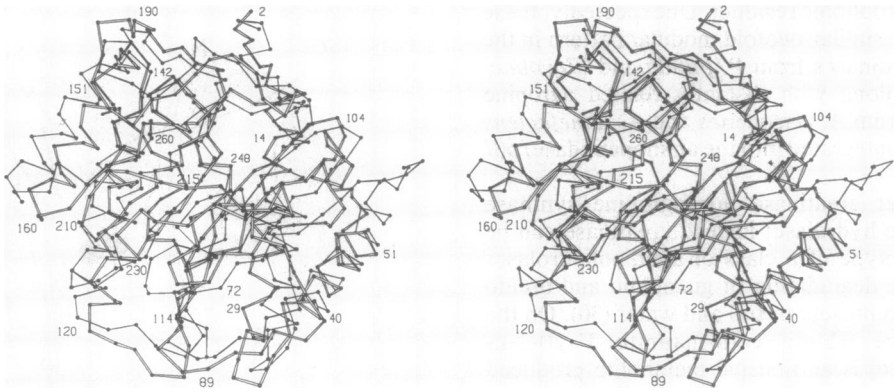

FIG. 3. Stereo drawing showing the superposition of common catalytic domains of creatinase (thin bonds) (7, 8) and AMPM (thick bonds) (9). The numbering is for the *E. coli* AMPM, and the corresponding numbering for creatinase can be obtained from Fig. 1.

4*A*). To date, no structural information is available on the binding of substrate or substrate analogs to AMPM. In the absence of such information, Fig. 4*B* shows the structure of carbamoylsarcosine, as seen in its complex with creatinase (7, 8), geometrically transposed into the active site of AMPM, based on the overall structural correspondence between the protein molecules. The position occupied by the carbamoyl group (analogous to the guanidinium group of creatine) is in the vicinity of Asp-97, Asp-108, and Glu-204, suggesting that this may correspond to the site occupied by the terminal amino group of a bound polypeptide chain. The *N*-methyl group is directed toward Phe-177, which is relatively solvent exposed, suggesting that this might form the binding site of the side chain of the N-terminal methionine (although there is no obvious hydrophobic pocket). When bound to creatinase, the carboxylate group of creatine is liganded to Arg-64,

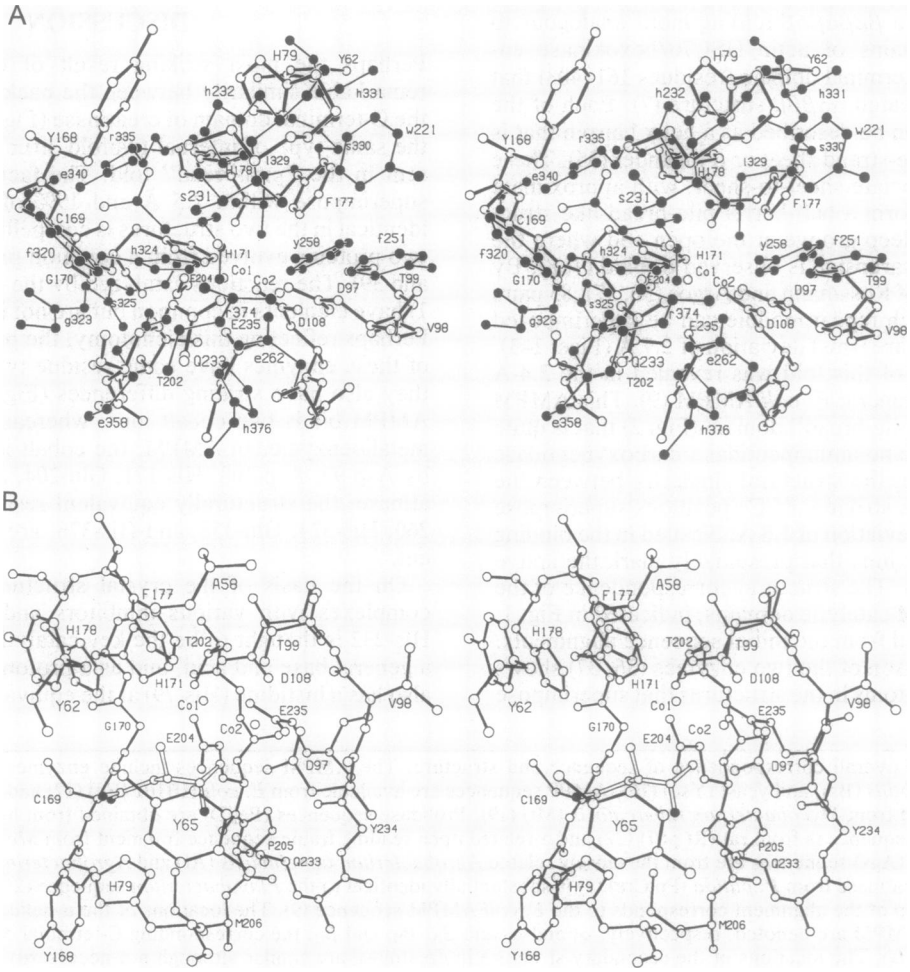

FIG. 4. (*A*) Superposition of the active-site region of creatinase (solid atoms) on that of AMPM (open atoms). Carbamoylsarcosine, bound to creatinase, is also included, with the oxygen atoms shown as open circles and the remaining atoms solid. Amino acids are identified using the one-letter code—uppercase for AMPM and lowercase for creatinase. The alignment of the active-site regions is based on the overall alignment of the backbones of the two structures (Fig. 3). (*B*) Expanded view, similar to *A*, showing carbamoylsarcosine (open bonds with oxygen atoms drawn solid) geometrically transformed into the active-site of AMPM (thin solid bonds). The active site of AMPM opens toward the "left" of the figure—i.e., the carboxylate group of the carbamoylsarcosine is directed toward solvent.

which extends into the active site from the dimer-related N-terminal domain (7). Because the chain-economical AMPM has no counterpart to the N-terminal domain of creatinase (although the N-terminal chain extensions of AMPP and prolidase enzymes do appear to correspond to the first domain of creatinase; data not shown) and is also monomeric instead of dimeric, the active site is more accessible and better able to accommodate an extended polypeptide substrate. It is to be stressed that these extrapolations are based entirely on structural homology and are not intended as a substitute for experimental data.

Given the structural correspondence between AMPM and creatinase, the sequence alignment shown in Fig. 1 suggests that the "pita-bread" fold extends to several other hydrolases. Particularly striking is the preservation of the cobalt-ligating residues. As noted above, these are not generally conserved in creatinase. Nevertheless, all five residues, as seen in AMPM, are, with two exceptions (site 235, Gln in *S. cerevisiae* AMPM and His in p67) strictly conserved in all four AMPMs, in the three AMPPs, in the three prolidases, and in the p67 homologs. This not only supports the suggestion of a common "pita-bread" fold for these proteins but also suggests that they bind metals. Yeast AMPM is known to be a cobalt-dependent enzyme (16). All forms of AMPP examined to date appear to be metalloproteins and are activated by, or are sensitive to, metals such as Mn^{2+}, Co^{2+}, and Zn^{2+} (11, 40–42). Human prolidase is in turn activated by Mn^{2+} (43, 44). Prolidase in humans is thought to be required for the metabolism of degradation products of collagen. Deficiency in the enzyme leads to abnormalities of the skin and other collagenous tissues, as well as to mental retardation (11, 29, 45, 46). It has been shown in two affected patients that Asp-276 of prolidase is replaced by Asn. Strikingly, this site corresponds to Asp-97 in *E. coli* AMPM, a cobalt ligand (Fig. 1).

A surprising finding of this work is the clear prediction, given the conservation of putative active site residues (Fig. 1), that the eIF-2-associated p67 factors function as metal-dependent (amino)peptidases, perhaps modifying and inactivating the attacking eIF-2 kinases. Agropine synthases, more closely related to creatinases than to the aminopeptidase/p67 group, do not conserve either of the cobalt ligands of AMPM but may conserve the "critical" active site His-232 of creatinase in the guise of His-236 (in the *Agrobacterium rhizogenes* sequence; Fig. 1). However, it is still premature to speculate on the catalytic nature of this latter enzyme. In summary, the structural family of "pita-bread"-fold enzymes represents a protein scaffold that can support a diverse set of catalytic functions.

We are grateful to Dr. Robert DuBose for helpful comments on the manuscript. This project was supported in part by the National Institutes of Health (Grant GM20066 to B.W.M.). Work at DNAX Research Institute (J.F.B.) was supported by Schering-Plough.

1. Chothia, C. (1992) *Nature (London)* **357**, 543–544.
2. Blundell, T. L. & Johnson, M. S. (1993) *Protein Sci.* **2**, 877–883.
3. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138.
4. Orengo, C. A., Flores, T. P., Taylor, W. R. & Thornton, J. M. (1993) *Protein Eng.* **6**, 485–500.
5. Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991) *Biochemistry* **30**, 9161–9169.
6. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I. & Schrag, J. (1992) *Protein Eng.* **5**, 197–211.
7. Hoeffken, H. W., Knof, S. H., Bartlett, P. A., Huber, R., Moellering, H. & Schumacher, G. (1988) *J. Mol. Biol.* **204**, 417–433.
8. Coll, M., Knof, S. H., Ohga, Y., Messerschmidt, A., Huber, R., Moellering, H., Russmann, L. & Schumacher, G. (1990) *J. Mol. Biol.* **214**, 597–610.
9. Roderick, S. L. & Matthews, B. W. (1993) *Biochemistry* **32**, 3907–3912.
10. Taylor, A. (1993) *Trends Biochem. Sci.* **18**, 167–172.
11. Yaron, A. & Naider, F. (1993) *Crit. Rev. Biochem. Mol. Biol.* **28**, 31–81.
12. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
13. Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) *Comput. Appl. Biosci.* **8**, 189–191.
14. Gribskov, M., McLachlan, A. D. & Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4355–4358.
15. Gibson, T. J., Rice, P. M., Thompson, J. D. & Heringa, J. (1993) *Trends Biochem. Sci.* **18**, 331–333.
16. Chang, Y. H., Teichert, U. & Smith, J. A. (1992) *J. Biol. Chem.* **267**, 8007–8011.
17. Henrich, B., Monnerjahn, U. & Plapp, R. (1990) *J. Bacteriol.* **172**, 4641–4651.
18. Butler, M. J., Bergeron, A., Soostmeyer, G., Zimny, T. & Malek, L. T. (1993) *Gene* **123**, 115–119.
19. Garbe, T., Servos, S., Hawkins, A., Dimitriadis, G., Young, D., Dougan, G. & Charles, I. (1991) *Mol. Gen. Genet.* **228**, 385–392.
20. Endo, F., Tanoue, A., Nakai, H., Hata, A., Indo, Y., Titani, K. & Matsuda, I. (1989) *J. Biol. Chem.* **264**, 4476–4481.
21. Yoshimoto, T., Tone, H., Honda, T., Osatomi, K., Kobayashi, R. & Tsuru, D. (1989) *J. Biochem.* **105**, 412–416.
22. Wu, S., Gupta, S., Chatterjee, N., Hileman, R. E., Kinzy, T. G., Danslow, N. D., Merrick, W. C., Chakrabarti, D., Osterman, J. C. & Gupta, N. K. (1993) *J. Biol. Chem.* **268**, 10796–10801.
23. Haas, E. S., Daniels, C. J. & Reeve, J. N. (1989) *Gene* **77**, 253–263.
24. Bouchez, D. & Tourneur, J. (1991) *Plasmid* **25**, 27–39.
25. Barker, R. F., Idler, K. B., Thompson, D. V. & Kemp, J. D. (1983) *Plant Mol. Biol.* **2**, 335–350.
26. Koyama, Y., Kitao, S., Yamamoto-Otake, H., Suzuki, M., Nakano, E. (1990) *Agric. Biol. Chem.* **54**, 1453–1457.
27. Ray, M. K., Datta, B., Chakraborty, A., Chattopadhyay, A., Meza-Keuthen, S. & Gupta, N. K. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 539–543.
28. Ray, M. K., Chakraborty, A., Datta, B., Chattopadhyay, A., Saha, D., Bose, A., Kinzy, T. G., Wu, S., Hileman, R. E., Merrick, W. C. & Gupta, N. K. (1993) *Biochemistry* **32**, 5151–5159.
29. Tanoue, A., Endo, F. & Matsuda, I. (1990) *J. Biol. Chem.* **265**, 11306–11311.
30. Tricot, C., Pierard, A. & Stalon, V. (1990) *J. Gen. Microb.* **136**, 2307–2317.
31. Firmin, J. L. & Fenwick, G. R. (1978) *Nature (London)* **276**, 842–844.
32. Tate, M. E., Ellis, J. G., Kerr, A., Tempe, J., Murray, K. E. & Shaw, K. J. (1982) *Carbohydr. Res.* **104**, 105–120.
33. Ellis, J. G., Ryder, M. H. & Tate, M. E. (1984) *Mol. Gen. Genet.* **195**, 466–473.
34. Bork, P., Sander, C. & Valencia, A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7290–7294.
35. Orengo, C. A. & Thornton, J. M. (1993) *Structure* **1**, 105–120.
36. Rossmann, M. G. & Argos, P. (1976) *J. Mol. Biol.* **105**, 75–96.
37. Matthews, B. W. & Rossmann, M. G. (1985) *Methods Enzymol.* **115**, 397–420.
38. Artymiuk, P. J., Grindley, H. M., Park, J. E., Rice, D. W. & Willett, P. (1992) *FEBS Lett.* **303**, 48–52.
39. Chothia, C. & Lesk, A. M. (1986) *EMBO J.* **5**, 823–826.
40. Hooper, N. M., Hryszko, J. & Turner, A. J. (1990) *Biochem. J.* **267**, 509–515.
41. Fleminger, G. & Yaron, A. (1984) *Biochim. Biophys. Acta* **789**, 245–256.
42. Rusu, I. & Yaron, A. (1992) *Eur. J. Biochem.* **210**, 93–100.
43. Davis, N. C. & Smith, E. L. (1957) *J. Biol. Chem.* **224**, 261–275.
44. King, G. F., Crossley, M. J. & Kuchel, P. W. (1989) *Eur. J. Biochem.* **180**, 377–384.
45. Mock, W. L. & Green, P. C. (1990) *J. Biol. Chem.* **256**, 19606–19610.
46. Endo, F., Tanoue, A., Kitano, A., Arata, J., Danks, D. M., Lapiere, C. M., Sei, Y., Wadman, S. K. & Matsuda, L. (1990) *J. Clin. Invest.* **85**, 162–169.