Vol. 184, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 15, 1992 Pages 60-66

EVOLUTIONARY RELATIONSHIP OF NAD+-DEPENDENT D-LACTATE DEHYDROGENASE: COMPARISON OF PRIMARY STRUCTURE OF 2-HYDROXY ACID DEHYDROGENASES

Sunil Kochhar, Peter E. Hunziker*, Phaik Leong-Morgenthaler, and Herbert Hottinger

Nestlé Research Center, Vers-chez-les-Blanc, P.O. Box 44, CH-1000 Lausanne 26, Switzerland

*Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received December 19, 1991

A comparison of the primary structures of NAD+-dependent D-lactate dehydrogenase with L-lactate dehydrogenase and L-malate dehydrogenase failed to show any sequence similarity. However, D-2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*, glycerate dehydrogenase from cucumber, D-3-phosphoglycerate dehydrogenase and erythronate 4-phosphate dehydrogenase from *Escherichia coli* showed 38 %, 24 %, 24 % and 22 % amino acid identity, respectively. The profile analysis of the aligned sequences confirmed their relatedness. The hydropathy profiles of the aligned dehydrogenases were almost identical between residues 100-300 indicating largely preserved folding patterns of their polypeptide chains. The data suggest that L- and D-specific 2-hydroxy acid dehydrogenase genes evolved from two different ancestors and thus represent two different sets of enzyme families. © 1992 Academic Press, Inc.

Comparative studies of three-dimensional structures of enzymes utilizing NAD+ as the cofactor in hydride transfer showed that the protein moiety is made up of a highly conserved coenzyme-binding domain, and a variable catalytic domain (1). Nevertheless, this domain is similar when substrates are similar e.g., L-lactate and L-malate dehydrogenase both utilize 2-oxo and 2-hydroxy acid substrates (2). Both enzymes are structurally identical and show strict enantiomeric specificity for the L-isomer of the substrate. The three-dimensional structures of L-lactate dehydrogenase, e.g., from dogfish in unligated form (3-5), in complex with NADH (6,7) and in ternary complex with NADH and substrate analog (8) are known. In contrast, no structure-function information on D-lactate dehydrogenase is available.

Recently, we determined the primary structure of NAD+-dependent D-lactate dehydrogenase from *Lactobacillus bulgaricus* and identified the amino acid residues at or near the active site (9). The amino acid sequence comparison data indicate that similarities between D- and L-lactate dehydrogenase do not extend beyond their functional similarities and that both have independently evolved.

0006-291X/92 $1.50
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

EXPERIMENTAL PROCEDURES

The amino acid sequence of D-lactate dehydrogenase was compared with the protein sequence data base of the National Biomedical Research Foundation (NBRF) and of the SWISSPROT using the programme FASTA (10). Pair-wise comparisons were carried out by the programme GAP (11). The programme scores the similarity values from the scoring matrix PAM250 (12,13). The profile analysis was carried out using PILEUP, PROFILEMAKE and PROFILESEARCH programmes (14,15).

RESULTS

A computer search of the NBRF and the SWISSPROT protein sequence data banks with the FASTA alogrithm (10) using the D-lactate dehydrogenase as the query sequence indicated amino acid sequence similarity with four dehydrogenases, namely; glycerate dehydrogenase (also known as hydroxypyruvate reductase) from cucumber (16), D-3-phosphoglycerate dehydrogenase from *E. coli* (17), D-2-hydroxyisocaproate from *L. casei* (18) and erythronate 4-phosphate dehydrogenase from *E. coli* (19). The initial/optimized match score were 393/609, 177/158, 163/249 and 62/181 for D-2-hydroxyisocaproate, glycerate, D-3-phosphoglycerate and erythronate 4-phosphate dehydrogenase, respectively. The amino acid identity within the maximum overlap was >30 %. A rigorous pair-wise sequence comparison by GAP alogrithm (11) with optimized gap penalty of 10 revealed >20 % amino acid identity among these dehydrogenases whereas non-related proteins e.g., L-lactate dehydrogenase, L-malate dehydrogenase, glyceraldehyde 3-phosphate dehydrogenase or L-2-hydroxyisocaproate dehydrogenase showed ≤12 % amino acid identities (Table I). When the conservative mutations of amino acid

TABLE I
Percentage identity and similarity among certain L- and D-2-hydroxy acid dehydrogenases

	D-LDH	D-HicDH	GDH	PGDH	EryPDH	L-HicDH	*Lc*L-LDH	*Bst*L-LDH	PHL-LDH
D-LDH[a]	-	**38**	**24**	**24**	**22**	**11**	**12**	**13**	**10**
D-HicDH[b]	61	-	**21**	**25**	**22**	**11**	**11**	**11**	**8**
GDH[c]	46	40	-	**23**	**19**	**15**	**10**	**12**	**11**
PGDH[d]	44	43	44	-	**20**	**11**	**11**	**11**	**10**
EryPDH[e]	45	44	39	41	-	**13**	**12**	**12**	**10**
L-HicDH[f]	34	34	36	34	33	-	**31**	**31**	**25**
*Lc*L-LDH[g]	33	31	31	33	32	56	-	**53**	**37**
*Bst*L-LDH[h]	36	33	31	34	31	57	72	-	**35**
PHL-LDH[i]	29	27	33	30	31	51	61	60	-

Percent identity i.e. invariant residues (in bold letters) and similarity (conservative mutations) were calculated by alignment of the two sequences according to the programme GAP (11) which uses the mutational matrix PAM250 (12, 13). The gap weight values were set to 10.

[a]D-LDH, D-lactate dehydrogenase; [b]D-HicDH, D-2-hydroxyisocaproate dehydrogenase; [c]GDH, glycerate dehydrogenase but also known as hydroxypyruvate reductase; [d]PGDH, D-3-phosphoglycerate dehydrogenase; [e]EryPDH, erythronate 4-phosphate dehydrogenase; [f]L-HicDH, L-2-hydroxyisocaproate dehydrogenase; [g]*Lc*L-LDH, *L. casei* L-lactate dehydrogenase; [h]*Bst*L-LDH, *B. stearothermophilus* L-lactate dehydrogenase; [i]PHL-LDH, pig heart L-lactate dehydrogenase.

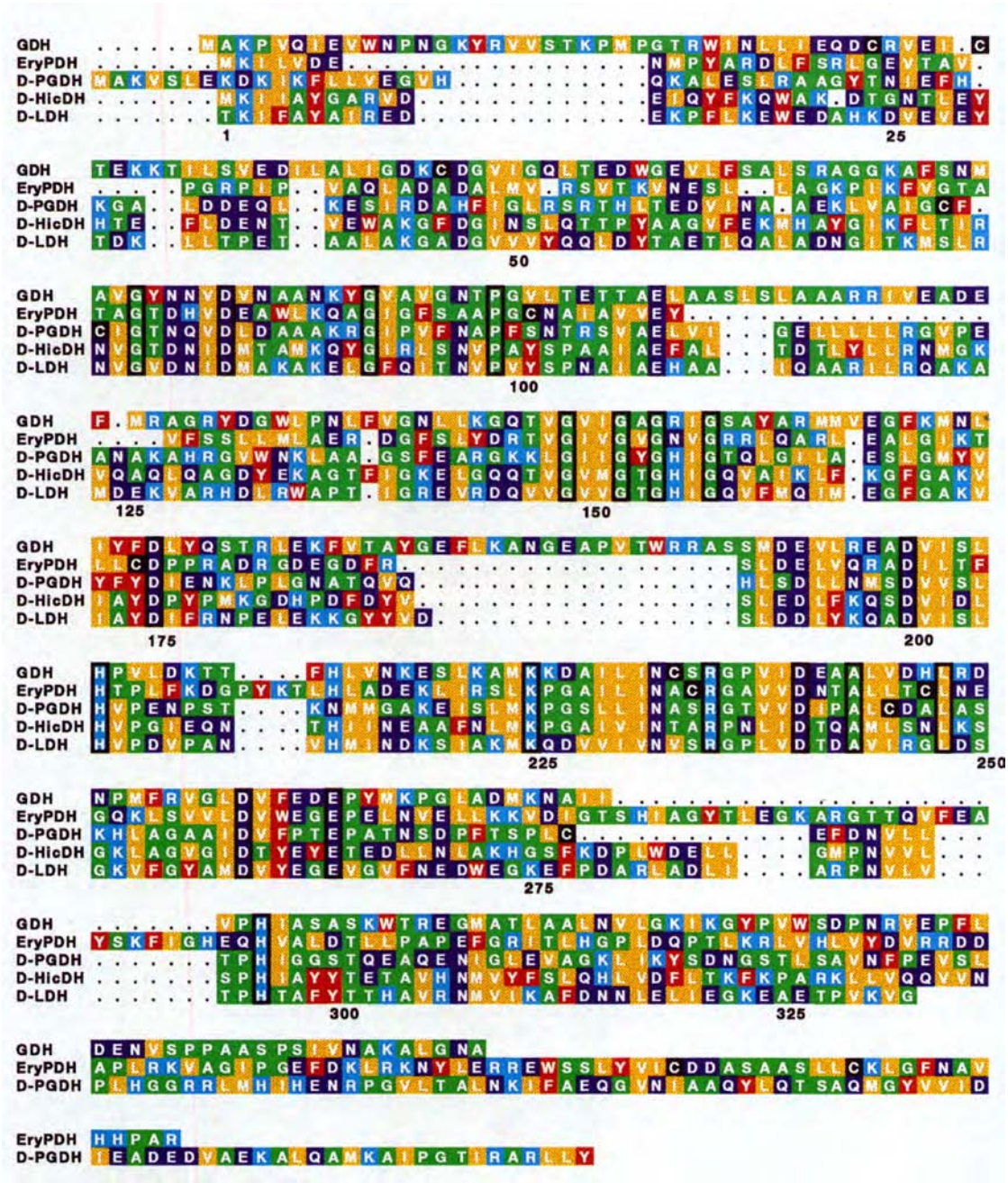

FIG. 1. **Aligned amino acid sequences of D-lactate, glycerate, D-3-phosphoglycerate, D-2-hydroxyisocaproate and Erythronate 4-phosphate dehydrogenases.** The residues are numbered according to the sequence of D-lactate dehydrogenase. The invariant residues are boxed. The conservative mutations according to PAM250 scoring matrix (12) are presented in colored background: nonaromatic nonpolar (M,I,L,V), yellow; aromatic nonpolar (F,Y,W), red; small with near neutral polarity (S,T,P,A,G), green; acid and uncharged polar (N,D,E,Q), dark blue and basic polar (H,R,K), light blue. The cysteine residues are shown with black background and gaps are represented as dots. The abbreviations of the names of the proteins are given in Table I.

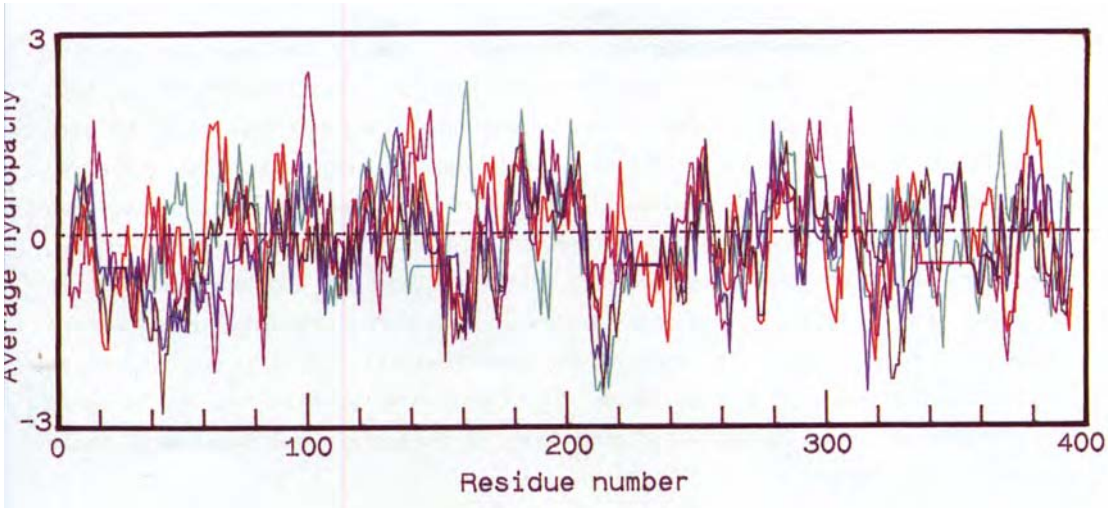

FIG. 2. Hydropathy profiles of D-lactate, glycerate, D-3-phosphoglycerate, D-2-hydroxyisocaproate and Erythronate 4-phosphate dehydrogenases. Consecutive hydropathy averages are plotted for a 9-residue window. Gaps for the optimal alignment were same as shown in Fig. 1. The hydropathy profile of D-lactate dehydrogenase is presented in black; D-2-hydroxyisocaproate dehydrogenase in blue; D-3-phosphoglycerate dehydrogenase in red; erythronate 4-phosphate dehydrogenase in magenta and of glycerate dehydrogenase in green.

residues with similar property, i.e. size, charge or hydrophobicity (12) were taken into consideration, an average score of >50 % similarity was observed (Table I). The aligned sequences of the similar dehydrogenases are presented in Fig. 1. The NAD^+-dependent dehydrogenases inherently show very poor sequence similarity (1,20). However, aligned sequences of homologous proteins intimately relate to the structure and the function of the proteins (15). In order to identify distant relatives of the D-2-hydroxy acid dehydrogenase family, a profile sequence was constructed by using the multiple sequence alignment programmes PILEUP (21) and PROFILEMAKE (14). A search through the SWISSPROT data bank using PROFILESEARCH (15) showed high scores (ZScore) of >30 for the aligned sequences whereas the other sequences in the data bank scored <4. A more rigorous search in a data bank constructed from the 32 L-lactate dehydrogenase, 24 L-malate dehydrogenase, 56 alcohol dehydrogenase, 52 glyceraldehyde 3-phosphate dehydrogenase, 18 alanine dehydrogenase and 26 phosphoglycerate kinase sequences also produced the same results. All lactate dehydrogenase sequences scored <1, thus confirming no sequence relatedness between L- and D-specific dehydrogenases. The N-1 profile sequence constructed by all but one of the aligned sequences (22) always showed a score of >29, thus confirming its unique relationship with other members of D-2-hydroxy acid dehydrogenases.

A total of 19 residues were found to be invariant among the aligned dehydrogenases (the number increases to 43 when a conservative mutation at one

of the five positions is accepted). Six out of the 19 residues are glycine residues (G79, G92, G149, G152, G154 and G157); five aspartate residues (D84, D175, D200, D240 and D259), two glutamate residues (D108 and D264) and one each of proline (P99), asparagine (N232), lysine (K225), arginine (235), alanine (A243) and histidine (H296) residues. Five out of six invariant glycine residues (G92, G149, G152, G154 and G157) are located at the NADH binding domain. Secondary structure predictions showed that invariant G79 and P99 residues are situated in a highly conserved β-sheet, apparently satisfying important structural requirements. The rest of the invariant residues are expected to be of substantial catalytic importance. The hydropathy profiles of the aligned dehydrogenases (Fig. 2) indicated that many of the structural domains of these proteins are similar.

DISCUSSION

The overall sequence identity score of 22 % to 38 % with glycerate, D-3-phosphoglycerate, erythronate 4-phosphate and D-2-hydroxyisocaproate dehydrogenases together with the average sequence similarity of about 50 % when conservative substitutions are also considered (12), clearly classifies them to be homologous proteins. The high degree of sequence similarity in D-2-hydroxy acid dehydrogenases indicate that the replacements have occurred subject to common structural and functional constraints. This is corroborated by their near identical hydropathy profiles and profile analysis of sequences constructed from the aligned dehydrogenases. The comparatively longer length of the polypeptides chains of D-3-phosphoglycerate dehydrogenase, glycerate dehydrogenase and erythronate 4-phosphate dehydrogenase mainly corresponded to the extensions of the polypeptide chain at the carboxyl terminus. Considering the fact that the average degree of identity among L-lactate dehydrogenase from many different species is around 38 % (similarity 60 %), a 38 % identity (61 % similarity) between D-lactate dehydrogenase and D-2-hydroxyisocaproate dehydrogenase with only one and two single amino acid deletions in the respective proteins together with their near identical secondary structure predictions suggests that both enzymes may interact with each other's substrate. The absence of any detectable reaction or inhibition of D-lactate dehydrogenase activity with 2-oxoisocaproate (9) indicates a different active site architecture in these two proteins.

The catalytic domains of D-specific 2-hydroxy acid dehydrogenases are expected to be divergent due to different substrate specificities of each protein. Nevertheless, an interesting mechanistic feature found in the family of L-2-hydroxy acid dehydrogenases is the presence of a similarly oriented Asp-His pair linked by a hydrogen bond and functioning as a proton relay system during catalysis (2). This type of catalytic arrangement is also found in the

family of serine proteases (23), thermolysin (24) and phospholipase A2 (25). Assuming that D-lactate dehydrogenase also possesses such an arrangement, rigorous analysis of the primary structure revealed two possible candidates for Asp-His pairs, namely, D175 - H205 and D259 - H296. In the former pair, the spatial neighbor of Asp, the substrate binding arginine residue, is also present at the expected position of D-X-X-R. However, in the present case D175 occupies the position of the invariant aspartate residue of the consensus sequence of the NADH binding domain, G-X-(G or A)-X-X-G-20X-D (1,20). This makes D175 equivalent to D53 of L-lactate dehydrogenase rather than its active site D146. Furthermore, the spatial neighbor of D175, R178 is not conserved among other D-2-hydroxy acid dehydrogenases. In the case of, D259 - H296 pair, no substrate binding arginine or lysine residues are present at the expected position. Chemical modification of D-lactate dehydrogenase with diethylpyrocarbonate and subsequent isolation of the modified peptide showed that modification of H303 was responsible for the loss of enzyme activity (9). A conserved glutamate residue, E264 is also present at the expected position. However, the histidine residue is not conserved, thus making it difficult to speculate which one of the His-Asp pairs is catalytically important.

In conclusion, the sequence comparison data suggest that L- and D- specific dehydrogenases have independent pedigrees. The possibility that L- and D-specific dehydrogenases have diverged so early that any similarity between their sequences would be too weak for detection seems improbable. It is very unlikely that stereospecificity with respect to substrate would drift whereas specificity with respect to cofactor would be retained. The sequence similarities as well as the similar hydropathy profiles of glycerate dehydrogenase, D-3-phosphoglycerate dehydrogenase, erythronate 4-phosphate dehydrogenase, D-lactate dehydrogenase and D-2-hydroxyisocaproate dehydrogenase, all from diverse species and different with respect to their substrate specificity but identical in their enantioselectivity, strongly suggests that D-specific 2-hydroxy acid dehydrogenases constitute a separate family of proteins. A high degree of sequence similarity between L-2-hydroxyisocaproate dehydrogenase with L-lactate dehydrogenase (26) and between D-2-hydroxyisocaproate dehydrogenase with D-lactate dehydrogenase on the one hand and a lack of any sequence similarity among L- and D-lactate dehydrogenases as well as L- and D-2-hydroxyisocaproate dehydrogenases on the other hand, strongly supports the conclusion that L- and D-specific 2-hydroxy acid dehydrogenases may not be evolutionarily related and their functional similarity is merely coincidental. Nevertheless, pair-wise identities of 10 % to 13 % between D- and L-lactate dehydrogenase are typical for the so-called "twilight zone" in which usually automatic alignment procedures fail (27,28), but background information argues in favor of their similarity. Thus, conclusive evidence for non-relatedness of L- and D-lactate dehydrogenase must await the X-ray crystallographic structure of D-lactate dehydrogenase.

ACKNOWLEDGMENTS

We thank Prof. Ph. Christen (Biochemisches Institut der universität Zürich, Zürich) for invaluable discussions during the course of our studies and critical review of the manuscript. We thank Dr. G.-A. Gross for helping to make the colored amino acid sequence alignment figure.

REFERENCES

1. Rossmann, M.G., Liljas, A., Brändén, C.-I., and Banaszak, L.J. (1975) in *The Enzymes*, 3rd Ed. (P. D. Boyer, Ed.), pp. 61-102, Academic Press, New York.
2. Birktoft, J.J., and Banaszak, L.J. (1983) *J. Biol. Chem.* **258**, 472-482.
3. Abad-Zpatero, C., Griffith, J.P., Sussman, J.L. and Rossmann, M.G. (1987) *J. Mol. Biol.* **198**, 445-467.
4. Adams, M.J., Haas, D.J., Jeffery, B.A., McPherson, A. jr, Mermall, H.L., Rossmann, M.G., Schevitz, R.W. and Wonacott, A.J. (1969) *J. Mol. Biol.* **41**, 159-188.
5. Adams, M.J., Ford, G.C., Koekoek, R., Lentz, P.J. jr, McPherson, A. jr, Rossmann, M.G., Smiley, I.E., Schevitz, R.W. and Wonacott, A.J. (1970) *Nature (London)* **227**, 1098-1103.
6. Adams, M.J., Liljas, A., and Rossmann, M.G. (1973) *J. Mol. Biol.* **76**, 519-531
7. Chandrasekra, K., McPherson, A. jr, Adams, M.J., and Rossmann, M.G. (1973) *J. Mol. Biol.* **76**, 503-518.
8. White, J.L., Hackert, M.L., Buehner, M., Adams, M.J., Ford, G.C., Lentz, P. J. jr, Smiley, I.E., Steindel, S.J., and Rossmann, M.G. (1976) *J. Mol. Biol.* **102**, 759-779.
9. Kochhar, S., Hunziker, P.E., Leong-Morgenthaler, P., and Hottinger, H. (1992)
10. Lipman, D.J., and Pearson, W.R. (1985) *Science* **227**, 1435-1441.
11. Needleman, S.B., and Wunsch, C.D. (1970) *J. Mol. Biol.* **48**, 443-453.
12. Dayhoff, M. Schwartz, R.M., and Orcutt, B.C. (1978) in *Atlas of Protein Sequence and Structure* (M. Dayhoff, Ed.), vol. 5 suppl. 3, pp. 345-352, National Biomedical Research Foundation, Washington D. C..
13. Gribskov, M., and Burgess, R.R. (1986) *Nucleic Acid Res.* **14**, 6745-6763
14. Gribskov, M., McLachlan, A.D., and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4355-4358.
15. Gribskov, M., Lüthy, R., and Eisenberg, D. (1990)*Methods Enzymol.* **183**, 146-159.
16. Greenler, J.M., Sloan, J.S., Schwartz, B.W., and Becker, W.M. (1989) *Plant Mol. Biol.* **13**, 139-150.
17. Tobey, K.L., and Grant, G.A. (1986) *J. Biol. Chem.* **261**, 12179-12183.
18. Lerch, H.P., Blöcker, H., Kallwass, H., Hoppe, J., Tsai, H., and Collins, J. (1989) *Gene* **78**, 47-57.
19. Schoenlein, P.V., Roa, B.B., and Winkler, M.E. (1989) *J. Bacteriol.* **171**, 6084-6092.
20. Birktoft, J.J., and Banaszak, L.J. (1984) in *Peptide and Protein Reviews* (H.T.W. Hearn, Ed.) Vol. 4, pp. 1-46, Marcel Dekker, New York.
21. Feng, D.-F., and Doolittle, R.F. (1987) *J. Mol. Evol.* **35**, 351-360.
22. Kraut, J. (1977) *Annu. Rev. Biochem.* **46**, 331-358.
23. Mehta, P.K., Hale, T.I., and Christen, P. (1991) in *Vitamin B6 and Carbonyl Catalysis* (T. Fukui, H. Kagamiyama, K. Soda, and H. Wada, Eds.) pp. 35-45, Pergamon Press, Oxford.
24. Weaver, L.H., Kester, W.R., and Mathews, B.W. (1977) *J. Mol. Biol.* **114**, 119-132.
25. Dijkstra, B.W., Drenth, J., and Kalk, K.H. (1981) *Nature (London)* **289**, 604-606.
26. Lerch, H.P., Frank, R., and Collins, J. (1989) *Gene* **83**, 263-270.
27. Doolittle, R.F. (1990) *Methods Enzymol.* **183**, 99-110.
28. Taylor, W.R. (1988) *Prot. Eng.* 2, 77-86.