Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases

Hans Jörnvall, Hedvig von Bahr-Lindström, Klaus-Dieter Jany*, Wolfgang Ulmer* and Marion Fröschle*

Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden and **Institute for Organic Chemistry, Biochemistry and Isotope Research, Stuttgart University, D-7000 Stuttgart 80, FRG*

Received 17 November 1983

The recently determined primary structure of glucose dehydrogenase from *Bacillus megaterium* was scanned by computerized comparisons for similarities with known polyol and alcohol dehydrogenases. The results revealed a highly significant similarity between this glucose dehydrogenase and ribitol dehydrogenase from *Klebsiella aerogenes*. Sixty-one positions of the 262 in glucose dehydrogenase are identical between these two proteins (23% identity), fitting into a homology alignment for the complete polypeptide chains. The extent of similarity is equivalent to that between other highly divergent but clearly related dehydrogenases (two zinc-containing alcohol dehydrogenases, 25%; sorbitol and zinc-containing alcohol dehydrogenases, 25%; ribitol and non-zinc-containing alcohol dehydrogenases, 20%), and suggests an ancestral relationship between glucose and ribitol dehydrogenases from different bactera. The similarities fit into a previously suggested evolutionary scheme comprising short and long alcohol and polyol dehydrogenases, and greatly extend the former group to one composed of non-zinc-containing alcohol-polyol-glucose dehydrogenases.

Homology *Alignment* *Short dehydrogenase* *Superfamily* *Non-zinc alcohol dehydrogenase*

1. INTRODUCTION

Two different types of similarities have been established between various pyridine-nucleotide linked dehydrogenases. One is a folding similarity in the coenzyme-binding domains of different dehydrogenases, initially observed from crystallographically determined structures (review [1]). It is widespread and can be used to predict where this domain is located, as for example demonstrated for *Drosophila* alcohol dehydrogenase [2] and a mammalian glycerophosphate dehydrogenase [3].

The other concerns amino acid sequence similarities and folding similarities comprising more or less entire protein chains of different dehydrogenases, strongly suggesting evolutionary divergence of separate activities from a common ancestor. This type of relationship (independent of and beyond characterized isozyme variants) is apparent for liver sorbitol dehydrogenase and liver alcohol dehydrogenase [4], and separately also for lactate dehydrogenase and malate dehydrogenase (cytoplasmic/mitochondrial) [5,6], as well as presumably for a bacterial ribitol dehydrogenase and an insect alcohol dehydrogenase [7].

On the basis of the latter complete-molecule homologies, an evolutionary scheme has been suggested, comprising two groups of alcohol/polyol dehydrogenases [7]. One is composed of long polypeptides (~350 residues) containing ligands to catalytic zinc [8], the other of short polypeptides (~250 residues) apparently lacking zinc [7]. In addition, common building unit(s) between the two types appear likely [7] as well as additional enzyme members in each group (for example, possibly a 15-hydroxyprostaglandin dehydrogenase [9], beside the originally suggested alcohol/polyol dehydrogenases).

Published by Elsevier Science Publishers B.V.
001457593/84/$3.00

When the primary structure of a glucose dehydrogenase became available [10], it was screened against all these enzymes for further tests of any relationships. This report shows that a highly significant fit against ribitol dehydrogenase was found, greatly improving the similarities within the group of short dehydrogenases, and extending that group to include glucose dehydrogenase. This finding supports the general scheme of divergence into two types of alcohol dehydrogenases, and suggests the superfamily to contain alcohol-polyol-sugar dehydrogenases.

2. MATERIALS AND METHODS

The primary structures of glucose dehydrogenase from *Bacillus megaterium* [10], alcohol dehydrogenase from horse liver [11], *Saccharomyces cerevisiae* [12] and *Drosophila melanogaster* [13], sorbitol dehydrogenase from sheep liver [14], and ribitol dehydrogenase from *Klebsiella aerogenes* [15] were compared in segments of variable span [16] comprising the entire molecules as described for other polypeptides [17, 18]. Secondary structures were predicted [19] as in [20].

3. RESULTS

3.1. *Comparisons show significant structural relationships between glucose dehydrogenase and ribitol dehydrogenase*

The amino acid sequence of glucose dehydrogenase was compared with primary structures of horse liver, yeast and *Drosophila* alcohol dehydrogenases, sorbitol dehydrogenase, and ribitol dehydrogenase to test for any similarities. In each case, the comparison was carried out in spans of 30 residues, covering all possible segments of both proteins in each pair, and the results were related to chance coincidences in randomly generated proteins with identical compositions [17]. Similarly, glucose dehydrogenase was compared with itself to test for any repeats in the structure.

The best matches obtained between any 30-residue segments are shown in table 1 for each comparison with glucose dehydrogenase. As expected

Table 1

Values for maximal identities between glucose dehydrogenase and other alcohol/polyol dehydrogenases

GlcDH compared with	Maximal positional identities (per 30 residues	Number of such segment similarities in different alignments	Length of such similarities (in number of constituent 30-residue segments with positional identities >8/30)	
(1)	(2)	(3)	(4)	
LADH	8–10	10	1–2	
YADH	8	13	1	
DADH	8	4	1	
SDH	8–9	14*	1–3	
GlcDH	8	7	1–2	
RDH	13	1	7	Σ:18
	11	1	8	
	10	1	3	
	8–10	7	1–2	

GlcDH, glucose dehydrogenase; LADH, YADH, DADH, liver, yeast, *Drosophila* alcohol dehydrogenase; SDH, sorbitol dehydrogenase; RDH, ribitol dehydrogenase. In all cases, the lumped alignments with 8–10 identities per 30 residues are non-unique and cover single segments, while the top alignments with RDH are significantly different (uncorrected P value of 7.0×10^{-7} for random coincidence disregarding the number of comparisons; that is 3–4 orders of magnitude lower than the P values for the other similarities listed), unique and longer, covering many 30-residue spans

* Five of these with shifts close to 170 residues (cf. text)

from the results with the randomly generated proteins (when P, uncorrected for number of comparisons, is ~0.01 for 8 identities/30 residues), the real proteins generally show maximal similarities at the level of about 8–10 identities/30 residues (table 1, column 2). This level largely represents the top random matches, as revealed by many (table 1, column 3) and unordered such alignments, and by small sizes in each case (table 1, column 4).

However, ribitol dehydrogenase was found to be significantly similar to glucose dehydrogenase, giving much better values in all types of estimates than expected from chance. Thus, maximal identities in 30-residue segments were 13 (table 1, column 2), this alignment and two others with 10–11 identities/30 residues were unique (table 1, column 3), differing only by single-residue shifts compatible with the presence of gaps, and all alignments covered large parts of the polypeptide chains as shown by the fact that they together account for 18 different 30-residue segments (table 1, column 4), in these 250-odd residue polypeptides. Apart from the 3 ordered alignments, the general background level of random similarities was also detected in the ribitol dehydrogenase/glucose dehydrogenase pair (bottom row, table 1).

The 3 significant ribitol dehydrogenase/glucose dehydrogenase alignments detected (table 1) fit into one alignment by introduction of a few gaps at 3 positions, as shown in fig. 1. The single alignment thus obtained gives 61 identities between the whole molecules, thus covering 23% of all residues (262) in glucose dehydrogenase and 25% of all (246) in ribitol dehydrogenase. For long polypeptide chains, these values are extensive, resembling values obtained for proteins of established relationships, such as β_2-microglobulin and immunoglobulin domains [21], or yeast alcohol dehydrogenase/liver alcohol dehydrogenase also giving 23–25% [4]. Thus, it can be concluded that glucose

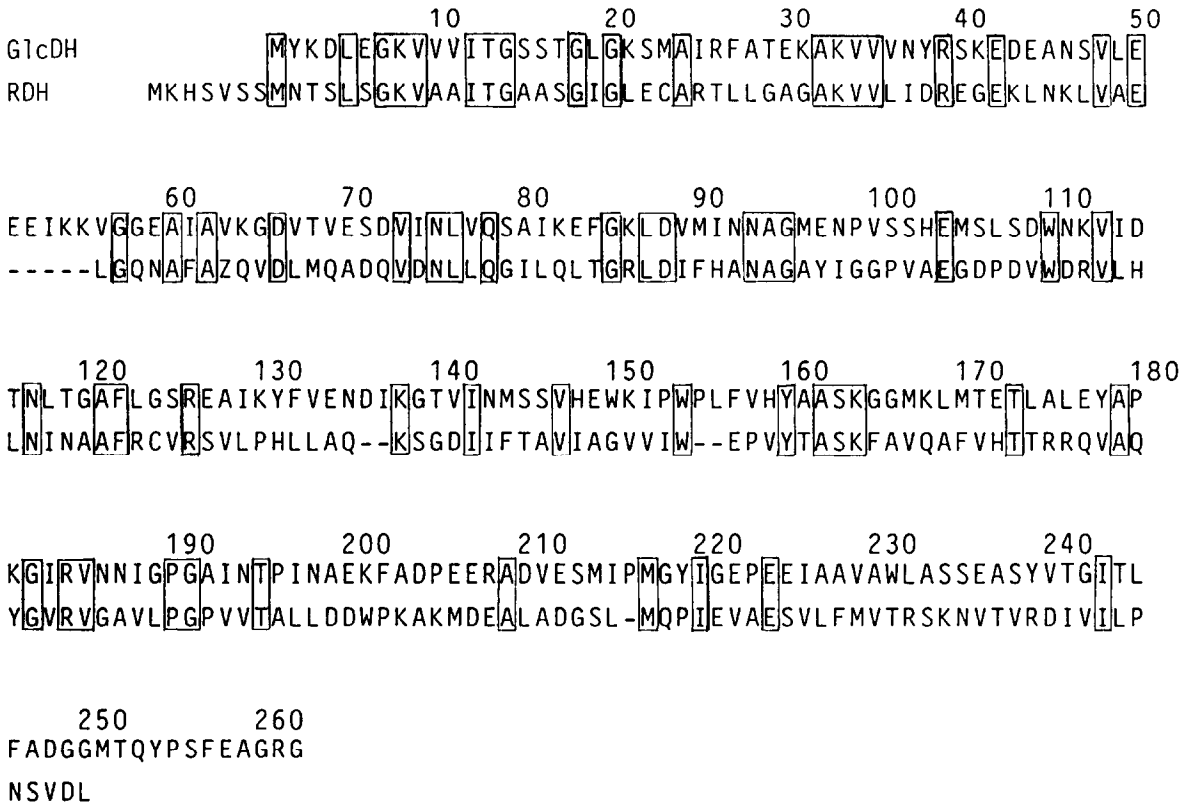

Fig. 1. Homology alignment between glucose dehydrogenase (GlcDH) from *B. megaterium* [10] and ribitol dehydrogenase (RDH) from *K. aerogenes* [15]. Positional identities are boxed.

Table 2

Residues in glucose dehydrogenase that are conserved in the alignment of fig. 1

Residue	Conserved	% of total
Cys	–	–
Asp	2	18
Asn	3	21
Thr	3	23
Ser	1	5
Glu	4	15
Gln	1	50
Pro	1	10
Gly	9	38
Ala	9	36
Val	8	33
Met	2	18
Ile	4	20
Leu	3	21
Tyr	1	13
Phe	1	13
Trp	2	50
Lys	4	21
His	–	–
Arg	3	50
Sum	61	23

dehydrogenase and ribitol dehydrogenase protein chains are structurally related and that this is visible, except possibly for the very C-terminal parts, over their entire lengths. Furthermore, as is typical for distantly related proteins [4], glycine is among the most conserved residues of those of common occurrence (table 2).

3.2. *Domain assignments in glucose dehydrogenase*

Coenzyme-binding domains of pyridine-nucleotide linked dehydrogenases have 6 β-strands in patterns of alternating α/β secondary structures [1]. Attempts at predictions of secondary structures for glucose dehydrogenase revealed several such alternating structures, most clearly in the N-terminal half (with possible β-strands at positions 8–14, 33–38, 64–70, 85–92, 112–116 and 128–133, although predictions may also suggest a few additional strands). The first two of the strands mentioned clearly fit the first two β-strands known or predicted for coenzyme-binding domains of other alcohol/polyol dehydrogenases, as shown in fig. 2. They also reveal that critical glycine residues in space-restricted coenzyme-adjacent positions [1] are conserved (fig. 2) provided a few gaps are accepted. These glycine residues correspond to positions 199, 201 and 204 in liver alcohol dehydrogenase [22] and to positions 14, 18 and 20 in glucose dehydrogenase. They have expected spacing and surroundings [1, 22, 23] typical for similar coenzyme-binding. Consequently, these residues, the β-predictions and the massive identities in this region among all characterized alcohol/polyol/sugar dehydrogenases suggest that the coenzyme-binding domain of glucose dehydrogenase can be identified as occurring in the N-terminal half of the protein chain, with βA starting at about position 8.

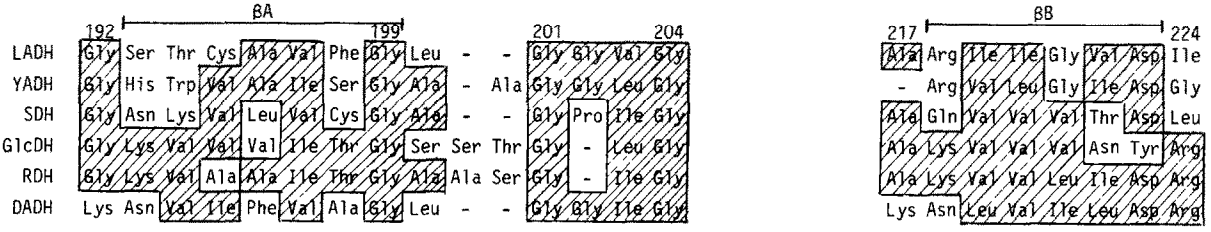

Fig. 2. Alignments between β-predicted secondary structures in glucose dehydrogenase (GlcDH) and the other dehydrogenases (abbreviated as in table 1). The predictions and homologies strongly suggest that βA and βB strands as well as Gly residues typical for coenzyme-binding domains [1,20] are conserved as shown. Outlines for these strands and numbers for the glycine residues given at the top are from the crystallographically determined [20] structure of LADH. Positions with identical residues (or only Val/Ile/Leu exchanges) are boxed. Positions shown are 192–204 and 217–224 for LADH, 170–183 and 196–202 for YADH, 170–182 and 194–202 for SDH, 7–20 and 32–39 for GlcDH, 14–27 and 39–46 for RDH, 7–19 and 32–39 for DADH.

4. DISCUSSION

4.1. *Homology and structure-function relationships*

The values in table 1 and identities in fig. 1 establish that glucose dehydrogenase (*B. megaterium*) and ribitol dehydrogenase (*K. aerogenes*) from two different bacteria are clearly homologous. The homology covers the whole protein chains and is compatible with ancestral connections in an evolutionary relationship (below).

Functionally, the coenzyme-binding domain of glucose dehydrogenase is suggested to occupy the N-terminal half of the polypeptide chain. Consequently, binding of NAD (or NADP [24]) may be ascribed to this part. This is suggested from 3 different facts:

(i) Direct prediction of secondary structures indicates a possible pattern of β-strands with alternating α-helices in this region. Two of these strands coincide with the extensive homologies (fig. 2) towards the first two such strands in other dehydrogenases, even fitting those dehydrogenases (alcohol/sorbitol dehydrogenases) that elsewhere hardly show any structural similarities with glucose dehydrogenase (cf. table 1).

(ii) Previous secondary structure predictions have suggested that ribitol dehydrogenase [7] and *Drosophila* alcohol dehydrogenase [2] also have their coenzyme-binding domains in the N-terminal halves, and these predictions coincide positionally with the region in glucose dehydrogenase that is anyway homology-related to ribitol dehydrogenase (fig. 1). Consequently, 3 separate secondary structure predictions ([2,7], and this work) and the present entire homology (fig. 1) fit in the assignment of the coenzyme-binding domain to the N-terminal part of glucose dehydrogenase.

The third support for this assignment is suggested by some weak sorbitol dehydrogenase/glucose dehydrogenase similarities. Thus, several of the background similarities for this dehydrogenase pair in table 1 are consistent with a weak partial similarity between the N-terminal region of glucose dehydrogenase and the region shifted to about position 170 in sorbitol dehydrogenase, also explaining the presence of some longer alignments (1–3 constituent 30-residue segments, table 1) in that comparison.

Combined, the comparisons suggest a clear relationship between glucose dehydrogenase and ribitol dehydrogenase, plus an assignment of coenzyme-binding in glucose dehydrogenase to the N-terminal part, as revealed by secondary structure predictions and some partial sequence similarities to additional alcohol/polyol dehydrogenases.

4.2. *Evolutionary scheme*

The demonstration of a glucose dehydrogenase/ribitol dehydrogenase relationship extends and considerably strengthens a previously suggested evolutionary scheme. Thus, alcohol/polyol dehydrogenases have been suggested to be divided into two evolutionary branches, one containing small (~250 residues) protein chains without zinc, and one containing long (~350 residues) protein chains with zinc [7]. The small protein chains were previously confined to *Drosophila* alcohol dehydrogenase and *Klebsiella* ribitol dehydrogenase, which only have just-discernible similarities [7], whereas the long protein chains were previously represented by extensively similar protein chains from sorbitol and zinc-containing alcohol dehydrogenases [7]. Now, this imbalance between the two branches has been removed, and the ribitol dehydrogenase/glucose dehydrogenase pair among the small protein chains is as related (23–25%) as the sorbitol dehydrogenase/alcohol dehydrogenase pairs among the long protein chains [4]. Consequently, support for the whole scheme is considerably stronger. At the same time, the previous two types of enzymes, alcohol dehydrogenases and polyol dehydrogenases, are now extended by a sugar dehydrogenase as shown in fig. 3. Although these activities formally all concern a hydroxyl group, the substrates are distinguishable and the enzymes were previously not expected to be related in this manner. Superficial properties may also suggest that still further hydroxyl-attacking enzymes may belong to the branch of small dehydrogenases in this scheme (fig. 3), for example possibly 15-hydroxyprostaglandin dehydrogenase with similar composition and small size [8]. It may be concluded that alcohol/polyol/sugar dehydrogenases form a complex superfamily of related proteins. As shown in fig. 3, this superfamily contains two probably

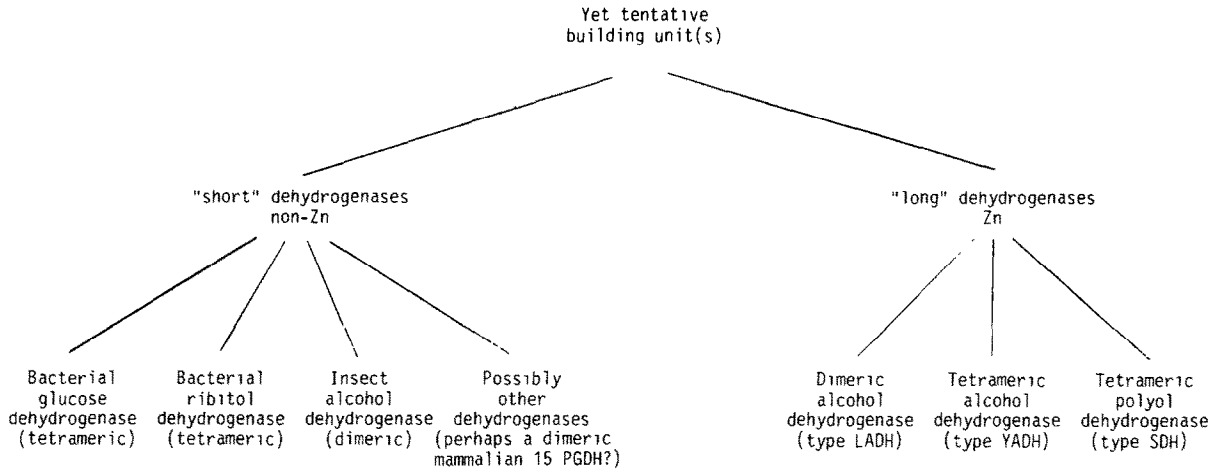

Fig. 3. Scheme of suggested relationships within alcohol-polyol-sugar dehydrogenases. Early separation into two types and inclusion of polyol dehydrogenases are from [7]. Early binding units are not characterized and entire chains need not be repeats or rearrangements of just one unit. Addition of glucose dehydrogenase is from fig. 1. The possible inclusion of 15-hydroxyprostaglandin dehydrogenase (15-PGDH) is as yet completely tentative, from total composition only [9]. Liver and yeast alcohol dehydrogenase (LADH) and YADH, respectively); sorbitol dehydrogenase (SDH).

divergent lines, presently distinguished by protein chains of different sizes and by the presence or absence of catalytic zinc. They furthermore demonstrate evolution of similar enzyme activities (alcohol dehydrogenases) separately in these long and short lines, thus demonstrating functional convergence but structural divergence. Finally, the inclusion of glucose dehydrogenase in this scheme further strengthens the conclusion that the quaternary structure is an apparently late and variable property in evolution since members of both lines suggest the presence of both dimeric and tetrameric molecules (fig. 3).

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Medical Research Council (project 13x-3532) and the Deutsche Forschungsgemeinschaft (Ja 308/5).

REFERENCES

[1] Rossmann, M.G., Liljas, A., Brändèn, C.-I. and Banaszak, L.J. (1975) in: The Enzymes, 3rd edn (Boyer, P.D., ed.) pp. 61–102, Academic Press, New York.
[2] Thatcher, D.R. and Sawyer, L. (1980) Biochem. J. 187, 884–886.
[3] Otto, J., Argos, P. and Rossmann, M.G. (1980) Eur. J. Biochem. 109, 325–330.
[4] Jörnvall, H., Von Bahr-Lindström, H. and Jeffery, J. (1983) Eur. J. Biochem., in press.
[5] Hill, E., Tsernoglou, D., Webb, L. and Banaszak, L.J. (1972) J. Mol. Biol. 72, 577–589.
[6] Birktoft, J.J., Fernley, R.T., Bradshaw, R.A. and Banaszak, L.J. (1982) Proc. Natl. Acad. Sci. USA 79, 6166–6170.
[7] Jörnvall, J., Persson, M. and Jeffery, J. (1981) Proc. Natl. Acad. Sci. USA, 78, 4226–4230.
[8] Jeffery, J., Chester, J., Mills, C., Sadler, P. and Jörnvall, H. (1983) EMBO J., submitted.
[9] Mak, O.T., Jörnvall, H. and Jeffery, J. (1982) Biosci. Rep. 2, 503–508.
[10] Jany, K.-D., Ulmer, W., Fröschle, M. and Pfleiderer, G. (1983) FEBS Lett., in press.
[11] Jörnvall, H. (1970) Eur. J. Biochem. 16, 25–40.
[12] Jörnvall, H. (1977) Eur. J. Biochem. 72, 425–442.
[13] Thatcher, D.R. (1980) Biochem. J. 187, 875–883.
[14] Jeffery, J., Cederlund, E. and Jörnvall, H. (1983) Eur. J. Biochem., in press.
[15] Moore, C.H., Taylor, S.S., Smith, M.J. and Hartley, B.S. (1974) and Hartley, B.S. (1977), cited in Atlas of Protein Sequence and Structure, 1978 (Dayhoff, M.D. ed.) vol. 5, suppl. 3, p. 68, National Biomedical Research Foundation, Washington, DC.
[16] McLachlan, A.D. (1971) J. Mol. Biol. 61, 409–424.
[17] Jörnvall, H., Mutt, V. and Persson, M. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 475–483.

[18] Jörnvall, H., Carlström, A., Pettersson, T., Jacobsson, B., Persson, M. and Mutt, V. (1981) Nature 291, 261–263.
[19] Von Bahr-Lindström, H., Galante, Y.M., Persson, M. and Jörnvall, H. (1983) Eur. J. Biochem. 134, 145–150.
[20] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222–245.
[21] Peterson, P.A., Cunningham, B.A., Berggård, I. and Edelman, G.E. (1972) Proc. Natl. Acad. Sci. USA, 69, 1697–1701.
[22] Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg., B.-O., Tapia, O., Brändén, C.-I. and Åkeson, Å. (1976) J. Mol. Biol. 102, 27–59.
[23] Wierenga, R.K. and Hol, W.J.G. (1983) Nature 302, 842–844.
[24] Pauly, H.E. and Pfleiderer, G. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1613–1623.