BIOCHEMICAL LETTERS

An amino acid sequence motif observed among enzymes of the shikimate pathway

Amino acid sequence motifs involved in processes of molecular recognition have been identified in several classes of proteins, such as ATP-utilizing enzymes (Walker *et al.*, 1982), nucleotide-binding proteins (Rossmann *et al.*, 1974) and DNA-binding proteins (Klug & Rhodes, 1987; Struhl, 1989). Observation of motifs such as these usually implies a common mechanistic feature and/or binding of a common substrate or cofactor. Sequence similarity between enzymes of a biosynthetic pathway might therefore be expected if a common structural feature of the substrates is recognized by a common structural motif of the enzymes of the pathway.

Examples of pairwise sequence similarity between consecutive enzymes of biochemical pathways have been reported, reflecting the fact that the product of one enzyme is the substrate of the next (Yeh *et al.*, 1978; Belfaiza *et al.*, 1986; Kovacevic & Miller,

1991). Structural and sequence similarity has also been observed between two mechanistically distinct enzymes of the mandelate catabolic pathway (Neidhart *et al.*, 1990; Tsou *et al.*, 1990). We would like to report the observation of a sequence motif found in three enzymes of the shikimate biosynthetic pathway.

The shikimate pathway is the biosynthetic pathway to the aromatic amino acids phenylalanine, tyrosine, and tryptophan (Haslam, 1974). The intermediates on the pathway are structurally similar, being six-membered carbocyclic rings with a carboxylate group at C-1, and oxygenated functional groups at C-3, C-4, and C-5 which are found in most or all of the intermediates of the pathway. In particular, the C-4 hydroxyl group is found in all the metabolites of the pathway, and this functional group is thought to provide an important interaction with both shikimate dehydrogenase (Bugg *et al.*, 1988) and dehydroquinase (Bugg, 1989).

In the course of studies on the latter enzymes, an observation was made that a 20-amino-acid region of sequence in shikimate kinase, the fifth enzyme of the pathway, showed approximately

		Consensus sequence	-- A L -- A L -- -- F V D L D -- L E T L --		
(a)					Score
Shikimate kinase (Asp-32)	<i>N. crassa qa-1S</i>	I M A - S T A M K R K I V D L E S E F H H L T G			86
	<i>A. nidulans arom</i>	G N W V S K A L N R P F V D L D T E L E T V E G			111
	<i>S. cerevisiae arom</i>	S K W C A S A L G Y K L V D L D E L F E Q Q H N			101
	<i>E. coli aroL</i>	G M A L A D S L N R R F V D T D Q W L Q S Q L N			89
Dehydroquinase (Asp-114)	<i>N. crassa qa-1S</i>	N H A L R L A P D Y L T V D L G - - L D S G L L			87
	<i>A. nidulans arom</i>	R L A F R S G C E - - F V D L D I A F P E D M L			103
	<i>S. cerevisiae arom</i>	D I A L K N G V E - - F L D L E L T L P T D I Q			94
	<i>E. coli aroD</i>	- A A I D S G L V D - M I D L E - - L F T G D D			85
Dehydroquinase synthase (Asp-166)	<i>A. nidulans arom</i>	I G A I - W Q P T K I Y I D L E F - L E T L P V			93
	<i>S. cerevisiae arom</i>	I G A F - W Q P K F V L V D I K W - L E T L A K			82
	<i>E. coli aroB</i>	I G A F - Y Q P A S V V V D L D C - L K T L P P			104
(b)					
Anthranilate synthase I (Asp-367)	<i>E. coli trpE</i>	T D H K E L S E H L M L V D L A R N D L A R I C			
	<i>S. typhimurium</i>	T D H K E P C E H L M L V D P A R N D L A R I C			
	<i>B. lactofermenticum</i>	T D A K E I A D D T M L V D L A R N D L A R V S			
	<i>B. subtilis</i>	K D E K E K A E H Y M L V D L A R N D I G R V A			
	<i>S. cerevisiae</i>	G S L K D R A E H V M L V D L A R N D I N R I C			
<i>p</i> -Aminobenzoate synthase I (Asp-308)	<i>E. coli pabB</i>	N S A K D R A E N L M I V D L M R N D I G R V A			
Chorismate mutase/ prephenate dehydrogenase (Asp-175)	<i>E. coli tyrA</i>	I G K P P L P K D C I L V D L A S V K N G P L Q			

Fig. 1. The VDL motif (a) and its matches in chorismate-binding enzymes (b)

Residues identical to the consensus sequence are in bold type, and functionally related residues are boxed. Conservative replacements allowed are A/G, F/V/L/I, D/E and S/T. The position of the central aspartate residue of each sequence is indicated. (a) Sequence alignments are based on the published alignment of the *N. crassa qa-1S* protein and the *arom* proteins from *A. nidulans* and *S. cerevisiae* with the monofunctional enzymes from *E. coli* (Anton *et al.*, 1987; and references cited therein). Dashes indicate gaps that have been introduced to maximize the alignment. Alignment scores quoted in this Figure and in the text were calculated on the basis of correlation with a frequency table constructed from amino acids identical or functionally similar to the consensus sequence found in the 11 sequences shown in (a). (b) Matches for the VDL motif in chorismate-binding enzymes. Sequence alignments are based on published alignments of anthranilate synthase I sequences (Essar *et al.*, 1990; and references cited therein) with *p*-aminobenzoate synthase I from *E. coli* (Goncharoff & Nichols, 1984). Alignment of the *E. coli* chorismate mutase/prephenate dehydrogenase sequence (Hudson & Davidson, 1984) was carried out by hand. The heavy line indicates the region of the four amino acid repeat. Cys-377 is indicated by an asterisk (Tso & Zalkin, 1981).

50% sequence similarity with regions of sequence in dehydroquinase and dehydroquinase synthase, the third and second enzymes of the pathway respectively (Fig. 1a). These enzymes catalyse completely different types of transformation: phosphorylation, dehydration and carbon-carbon bond formation respectively. The sequences occur in different parts of the three enzymes, but in each case the sequences are strongly conserved between species. The sequence found in shikimate kinase occurs immediately after the presumed ATP-binding site (Millar *et al.*, 1986).

The sequence motif consists of a strongly conserved F/L-V-D-L-D/E pentapeptide sequence, flanked by three other partially conserved regions of sequence similarity. Insertion of gaps and alignment of functionally similar residues improved the alignment considerably, suggesting that any evolutionary relationship between these sequences is distant. In order to assess the significance of these sequence alignments, a consensus sequence A-L-X₂-A/G-L-X₃-F/L-V-D-L-D/E-X₂-L-E-T-L (where X is either any residue or no residue) was used to probe the PIR (Protein Identification Resource, National Biomedical Research Foundation) protein sequence database for any matching sequences, giving an alignment score for each match found (Fig. 1a).

All of the sequences in Fig. 1(a) receive an alignment score of at least 82, whereas fewer than 30 other sequences in the PIR database (release 23, 6550 entries) score 82 or more, consistent with the number of matching sequences expected to occur randomly in a database of this size. Only four sequences score more than 90; these are phosphorylase (score 94), UL38 protein Herpes simplex virus (102), *env* polyprotein avian reticuloendotheliosis virus (93), and gene 55.12 protein bacteriophage T4 (92). The fact that the alignment scores obtained for the sequences in Fig. 1(a) are higher than the distribution of scores from the database implies that the level of sequence similarity observed is above (although close to) the level that is statistically significant (Doolittle, 1986).

The regions of sequence identified in Fig. 1(a) are strongly conserved between the monofunctional *Escherichia coli* enzymes and the multifunctional *arom* complexes of *Saccharomyces cerevisiae* and *Aspergillus nidulans*, and are also found in the sequence of the *qa-1S* repressor protein of *Neurospora crassa* (Geever *et al.*, 1989). This protein is involved in the regulation of the inducible quinate pathway, and shows 10–25% sequence identity with the dehydroquinase, shikimate dehydrogenase and shikimate kinase domains of the *arom* complex (Anton *et al.*, 1987). The repressor protein lacks catalytically important residues (e.g. in the dehydroquinase domain, the lysine involved in Schiff's base formation is replaced by an arginine), consistent with its role as a binding protein rather than an enzyme. Therefore the presence of the motif in the *qa-1S* protein is indicative that it has a role in substrate binding rather than in catalysis.

Searches through the amino acid sequences of other enzymes of the shikimate pathway revealed similar sequences centred on a L-V-D-L-A pentapeptide sequence in three enzymes involved in utilization of chorismate, the final intermediate of the pathway: anthranilate synthase I, *p*-aminobenzoate synthase I, and chorismate mutase/prephenate dehydrogenase (Fig. 1b). The region of sequence found in anthranilate synthase I lies in one of the most highly conserved regions of the protein sequence, and is situated close to Cys-377, which can be modified by bromopyruvate and is essential for activity, suggesting that this region lies at or near the active site (Tso & Zalkin, 1981). Interestingly, this region of the *E. coli* anthranilate synthase I sequence contains a four amino acid residue repeat (V-D-L-A-R-N-D-L-A-R-I) which is also repeated at the DNA level, with 75% conservation at the third 'wobble' position, suggesting that this sequence has arisen from an early duplication event (Ycas, 1976; Doolittle,

1985). It therefore appears that we have found an important region of the anthranilate synthase I sequence, also conserved in two other chorismate-binding proteins, which bears a striking similarity to the earlier sequences.

We have thus identified a motif common to, and specific for, enzymes of the shikimate pathway. In the absence of an X-ray crystal structure for one of these enzymes, it is not possible to say what the exact role of this motif is, although the fact that these sequences occur in highly-conserved regions of protein sequence in each enzyme is consistent with an active site role. Further work is required to elucidate its structural or functional significance, but one possible role that would be consistent with its widespread occurrence on the pathway would be binding of the C-4 hydroxyl group found in all the metabolites used by these enzymes.

The observation of this motif raises important questions about the relationships between enzymes on biochemical pathways. Have these enzymes found, convergently, a related solution to a shared problem, or are there indeed distant evolutionary links between the enzymes of the pathway? Do these sequences represent a common substrate binding domain, which was incorporated with other functional domains leading to the evolution of the diverse enzymes of the pathway? If this were the case, then one might expect the original sequence to have been mutated divergently in the individual enzymes, so that the motif reported in this paper is the remains, functional or vestigial, of the original domain.

We thank Drs. J. Coggins, R. F. Doolittle, M. Bradley, K. Duncan, C. Kleanthous, F. J. Leeper and J. Liu for helpful discussions about these observations.

Timothy D. H. BUGG*, Peter R. ALEFOUNDER† and Chris ABELL

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K. (*Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.

†Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.)

- Anton, I. A., Duncan, K. & Coggins, J. R. (1987) *J. Mol. Biol.* **197**, 367–371
- Belfaiza, J., Parsot, C., Martel, A., Bouthier de la Tour, C., Margarita, D., Cohen, G. N. & Saint-Girons, I. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 867–871
- Bugg, T. D. H. (1989) Ph.D. Thesis, University of Cambridge
- Bugg, T. D. H., Abell, C. & Coggins, J. R. (1988) *Tetrahedron Lett.* **29**, 6779–6782
- Doolittle, R. F. (1985) *Trends Biochem. Sci.* **10**, 233–237
- Doolittle, R. F. (1986) *Of Urfs and Orfs: a Primer on how to Analyse Derived Amino Acid Sequences*, University Science Books, Mill Valley, CA
- Essar, D. W., Eberly, L., Han, C. & Crawford, I. P. (1990) *J. Bacteriol.* **172**, 853–866
- Geever, R. F., Huiet, L., Baum, J. A., Tyler, B. M., Patel, V. B., Rutledge, B. J., Case, M. E. & Giles, N. H. (1989) *J. Mol. Biol.* **207**, 15–34
- Goncharoff, P. & Nichols, B. P. (1984) *J. Bacteriol.* **159**, 57–62
- Haslam, E. (1974) *The Shikimate Pathway*, Butterworths, London
- Hudson, G. S. & Davidson, B. E. (1984) *J. Mol. Biol.* **180**, 1023–1051
- Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469
- Kovacevic, S. & Miller, J. R. (1991) *J. Bacteriol.* **173**, 398–400
- Millar, G., Lewendon, A., Hunter, M. G. & Coggins, J. R. (1986) *Biochem. J.* **237**, 427–437
- Neidhart, D. J., Kenyon, G. L., Gerlt, J. A. & Petsko, G. A. (1990) *Nature (London)* **347**, 692–694
- Rossmann, M. G., Moras, D. & Olsen, K. M. (1974) *Nature (London)* **250**, 194–199
- Struhl, K. (1989) *Trends Biochem. Sci.* **14**, 137–140
- Tso, J. Y. & Zalkin, H. (1981) *J. Biol. Chem.* **256**, 9901–9908

- Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Buechter, D. D., Babbitt, P. C. & Kenyon, G. L. (1990) *Biochemistry* **29**, 9856–9862
- Walker, J. E., Saraste, M., Runswick, M. J. & Gray, M. J. (1982) *EMBO J.* **1**, 945–951
- Ycas, M. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2139–2140
- Yeh, W. K., Davis, G., Fletcher, P. & Ornston, L. N. (1978) *J. Biol. Chem.* **253**, 4920–4923

Received 23 January 1991

What determines the increase in liver cell volume in the fasted-to-fed transition: glycogen or insulin?

Liver size decreases during fasting and increases during refeeding in parallel with changes in glycogen content. It is traditionally assumed that an increase in cell volume occurs in association with, and as a consequence of, glycogen deposition. The recent evidence that changes in liver cell volume regulate glycogen metabolism [1–3] together with the well-established role of insulin in the regulation of potassium homeostasis [4–6] suggests an alternative hypothesis: that insulin regulates cell volume by stimulation of K⁺ uptake and that the increase in cell volume is the *cause* rather than the consequence of glycogen deposition.

Recent studies from four independent laboratories have established that changes in liver cell volume induced either by aniso-osmotic media or by Na⁺-driven substrate uptake cause co-ordinate changes in carbohydrate [1–3], amino acid [7–10] and protein [11] metabolism in the perfused liver and in isolated hepatocyte suspensions. This raises several questions on fundamental concepts of the control of metabolism. Häussinger, Lang and colleagues showed that glycogen degradation in the perfused rat liver is stimulated when liver cell volume is decreased

with hyper-osmotic media and inhibited when cell volume is increased [1,2]. Hue, Meijer and colleagues showed that glycogen deposition in hepatocyte suspensions correlates directly with cell volume irrespective of whether cell volume is increased by amino acids or by hypo-osmotic media [3]. Similar effects of changes in cell volume on the rates of glycogen synthesis are observed in hepatocyte cultures (M. Al-Habori, M. Peak & L. Agius, unpublished work).

The changes in glycogen metabolism induced by cell swelling cannot be explained by changes in cyclic AMP or cytosolic Ca²⁺ concentration [12] and it is currently not known what the mechanisms are that link changes in cell volume with glycogen deposition. Since glycogen deposition in the post-absorptive state involves an increase in the number of glycogen particles [13] it is possible that a cytoskeletal mechanism triggered by an increase in cell volume may be involved in the development of new glycogen particles. Regardless of the mechanism, the remarkable correlation *in vitro* between the rate of glycogen synthesis and cell volume, when the latter is altered experimentally [3], raises the issue whether the well-known increase in liver size in the fasted-to-fed transition is the cause rather than consequence of increased glycogen storage.

Liver size decreases during fasting and increases during refeeding (Table 1) [14]. In the rat the decrease in liver weight during a 48 h fast (44%) is associated with a proportional decrease in the total protein content ([14]; Table 1). However, during refeeding with carbohydrate, the initial increase in liver weight is proportionally greater than the small increase in protein content and consequently the protein concentration (per unit wet weight of liver) is diluted by about 36% relative to the fasted state (Table 1). During refeeding with carbohydrate and casein, the increase in liver weight (and in liver weight/body weight ratio) is greater than during refeeding with carbohydrate alone and consequently the protein concentration is diluted further (by about 44%; Table 1). The greater increase in liver weight during refeeding with carbohydrate and protein compared with carbohydrate alone can now be explained, at least in part, by amino-

Table 1. Changes in the total hepatic content and concentrations of glycogen, water and K⁺ during fasting and refeeding in the rat

Data are from [14]. Normal rats were fasted for 48 h or they were fasted for 48 h and then allowed to refeed *ad libitum* for 20 h either on carbohydrate (sucrose and starch) alone (Refed-1) or on carbohydrate and casein (Refed-2). All animals had access to water, but the refed rats also had access to a salts solution (10 × Ringer's salts solution). The water content represents intracellular and extracellular water. The values for total liver weight and content of water, protein, glycogen and K⁺ were determined from the concentrations (per unit wet wt.) and liver weight/body weight ratios in Table 1 in [14]. Values are means ± S.E.M. for the number of animals indicated, except for Refed-2 which are individual values of two rats. Statistics were by the unpaired *t*-test: **P* < 0.05, ***P* < 0.005 relative to values of fasted rats.

State of rats ... n ...	Fed (5)	48 h-fasted (4)	Refed-1 (6)	Refed-2 (2)
Body wt. (g)	205 ± 11*	163 ± 10	181 ± 8	165, 165
Liver wt. (g)	7.5 ± 0.4**	4.2 ± 0.2	7.3 ± 0.4**	10.6, 9.7
Liver wt./body wt. (%)	3.7 ± 0.1**	2.6 ± 0.1	4.0 ± 0.2**	6.4, 5.9
Total liver content (per liver)				
Water content (g)	5.2 ± 0.3**	3.0 ± 0.1	5.0 ± 0.3**	7.4, 6.7
Dry weight (g)	2.3 ± 0.1**	1.2 ± 0.02	2.3 ± 0.3**	3.2, 3.0
Protein content (g)	1.65 ± 0.12*	0.96 ± 0.03	1.05 ± 0.08	1.3, 1.3
Glycogen content (g)	0.29 ± 0.03**	0.01 ± 0.01	0.83 ± 0.07**	1.49, 1.42
K ⁺ (mmol)	0.73 ± 0.04**	0.43 ± 0.02	0.70 ± 0.04**	1.0, 0.91
Concentrations (per 100 g wet wt. of liver)				
Water (% wet wt.)	69.1 ± 0.1	70.4 ± 0.7	69.0 ± 0.3	69.7, 68.8
Protein (% wet wt.)	21.8 ± 0.5	22.9 ± 0.4	14.6 ± 0.7**	12.1, 13.4
Glycogen (% wet wt.)	3.9 ± 0.4**	0.23 ± 0.10	11.3 ± 0.7**	14.1, 14.6
K ⁺ (mmol/100 g)	9.7 ± 0.1*	10.2 ± 0.2	9.6 ± 0.1*	9.4, 9.3
K ⁺ /protein	0.446 ± 0.013	0.444 ± 0.010	0.661 ± 0.033**	0.78, 0.70