APPENDIX

Secondary-Structure Prediction from the Sequence of *Drosophila melanogaster* (Fruitfly) Alcohol Dehydrogenase

David R. THATCHER* and Lindsay SAWYER†

*Department of Molecular Biology, University of Edinburgh, Edinburgh, EH9 3JR, Scotland, U.K., and

†Department of Chemistry, Napier College, Edinburgh, EH10 5DT, Scotland, U.K.

(Received 27 November 1979)

NAD⁺-binding domains of all nucleotide-binding enzymes of known three-dimensional structure show striking similarities in their secondary-structural organization. These coenzyme-binding domains have an evolutionarily converged/conserved pattern of parallel β -pleated sheet, flanked by a number of α -helices (Rossman *et al.*, 1975). At the amino-acid-

sequence level this pattern takes the form of a series of alternating stretches of α -helix- and β -sheet-forming residues (the $\beta\alpha\beta$ unit of Levitt & Chothia, 1976). A recognizable pattern of $\beta\alpha\beta$ predictions in an amino-acid sequence may therefore be diagnostic for the presence of a nucleotide-binding fold. This rationale was successfully used by Wootton

Fig. 1. Secondary-structure prediction of *Drosophila* alcohol dehydrogenase

Stretches of α -helix are represented by (— α —) and β -sheet by (— β —). α/β means that both α -helix and β -sheet predictions were high and no decision could be made. Where both the α -helix and β -sheet predictions are low, no assignment has been made. Reverse turns are represented by (— $_{RT}$ —). Necessary details of the one-letter code for amino acids are given in the legend to Fig. 2 of the main paper (Thatcher, 1980).

(1974) to identify two such domains in the amino-acid sequence of glutamic dehydrogenase.

The method of Chou & Fasman (1974) was used to predict the secondary-structural propensities of the alcohol dehydrogenase sequence. All possible nucleation centres were examined, and sequences were extended until a 'breaker' residue was encountered. Where both α -helix and β -sheet were predicted, a decision was made by choosing the structure having the highest average potential for the stretch of sequence considered.

The nucleation centres once established were extended one residue at a time until the tetrapeptide potential (helix or sheet) fell below 1.0. In some cases residues at the termini were included or excluded according to the A3, A4, B3 and B4 conditions of Chou & Fasman (1974). The parameters used were those of Fasman *et al.* (1976), for α -helix and β -sheet and of Small *et al.* (1977) for reverse-turn prediction. The criteria of Chou & Fasman (1974) for reverse turns [i.e. the average frequency of occurrence of reverse turns (pt) = 0.5×10^{-4}] predicted an abnormally high proportion of the sequence in this conformation. A minimum of $pt = 2.0 \times 10^{-4}$ was therefore adopted, being three times the average reverse-turn potential

of the total sequence. Assignments of reverse turns were made where $pt \geq 2.0 \times 10^{-4}$, $\langle Pt \rangle \gg \langle P_\alpha \rangle$ and $\langle P_\beta \rangle$, and where $\langle P_\alpha \rangle$ and $\langle P_\beta \rangle < 1.0$. $\langle P_\alpha \rangle$ and $\langle P_\beta \rangle$ are the conformational potentials for a particular stretch of sequence and are the average of the P_α and P_β respectively of each particular residue in that stretch of sequence. $P_\alpha = f_\alpha / \langle f_\alpha \rangle$, and $P_\beta = f_\beta / \langle f_\beta \rangle$. f_α and f_β are the frequencies of a residue known to be in an α -helix or β -sheet conformation from X-ray-crystallographic data and $\langle f_\alpha \rangle$ and $\langle f_\beta \rangle$ are the average occurrence of these residues in the total structures. $pt = f_i f_{i+1} f_{i+2} f_{i+3}$ represents the relative probability that a particular peptide will form a reverse turn, where f_i , f_{i+1} etc. are the frequencies of occurrence of certain residues at the first, second etc. positions in the reverse turn. $\langle Pt \rangle$ is the average conformational potential of a reverse-turn tetrapeptide sequence. The individual conformational potentials of each residue (Pt) = $ft / \langle ft \rangle$ where ft is the number of residues in the reverse-turn conformation and $\langle ft \rangle$ the average frequency of those residues in the structures studied by Fasman *et al.* (1976).

The secondary-structure prediction obtained by this method is shown in Fig. 1. An alternating $\beta\alpha\beta$ pattern of high potential was located in the N-

Fig. 2. N-terminal 90 residues of *Drosophila* alcohol dehydrogenase aligned with the C-terminal nucleotide-binding domain of horse liver alcohol dehydrogenase

The sequences are aligned on the basis of a similarity in sequence near the first stretch of predicted β -sheet in the *Drosophila* alcohol dehydrogenase and the observed initial stretch of β -sheet in the horse liver alcohol dehydrogenase coenzyme-binding site. The Chou & Fasman (1974) secondary-structure prediction of Wootton (1974) is shown beneath the liver alcohol dehydrogenase sequence. The predictions are labelled as in Fig. 1 and the actual liver alcohol dehydrogenase secondary structures are shown. The sheet and helix nomenclature derives from Eklund *et al.* (1976). Abbreviations used: LADH, horse liver alcohol dehydrogenase; DADH, *Drosophila melanogaster* alcohol dehydrogenase. Details of the one-letter code for amino acids are given in the legend to Fig. 2 of the main paper (Thatcher, 1980).

terminal half of the molecule, consistent with the presumed structure of a nucleotide-binding domain in this region of the sequence. The sequence stretch from residues 128 to 150 is unusual in that the prediction of β -sheet is strong in this region but is accompanied by a number of intense reverse-turn potentials (e.g. $pt = 1.33 \times 10^{-3}$ at residue-128). The sheet prediction in this region has therefore been omitted from Fig. 1. The prediction for the secondary structure of the C-terminal half of the molecule was on the whole weaker than that of the N-terminal half, and α -helical segments interspersed with reverse turns appeared to predominate, with little β -structure and no $\beta\alpha\beta$ super-secondary-structural units. The C-terminus gave strong reverse-turn predictions and probably represents a distorted helix.

In Fig. 2 the hypothetical *Drosophila* alcohol dehydrogenase nucleotide-binding domain is aligned with both the known and predicted nucleotide-binding domains of horse liver alcohol dehydrogenase.

We thank Dr. J. Moulton for a computer program based on the Chou & Fasman (1974) method and the Physics Department, University of Edinburgh, Edinburgh, Scotland, U.K., for computing facilities.

References

- Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
- Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, B.-O., Tapia, O. & Branden, C.-I. (1976) *J. Mol. Biol.* **102**, 27–59
- Fasman, G. D., Chou, P. Y. & Adler, A. C. (1976) *Biophys. J.* **16**, 1201–1238
- Levitt, M. & Chothia, C. (1976) *Nature (London)* **250**, 194–199
- Rossmann, M., Liljas, A., Branden, C.-I. & Banaszak, L. J. (1975) *Enzymes 3rd Ed.* **11**, 61–102
- Small, D., Chou, P. T. & Fasman, G. D. (1977) *Biochem. Biophys. Res. Commun.* **79**, 341–346
- Thatcher, D. R. (1980) *Biochem. J.* **187**, 875–883
- Wootton, J. (1974) *Nature (London)* **252**, 542–546