Biochem. J. (1988) 249, 789–793 (Printed in Great Britain)

The predicted secondary structures of class I fructose-bisphosphate aldolases

Lindsay SAWYER,*‡ Linda A. FOTHERGILL-GILMORE* and Paul S. FREEMONT†
*Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, U.K., and
†Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, U.S.A.

The results of several secondary-structure prediction programs were combined to produce an estimate of the regions of α-helix, β-sheet and reverse turns for fructose-bisphosphate aldolases from human and rat muscle and liver, from *Trypanosoma brucei* and from *Drosophila melanogaster*. All the aldolase sequences gave essentially the same pattern of secondary-structure predictions despite having sequences up to 50% different. One exception to this pattern was an additional strongly predicted helix in the rat liver and *Drosophila* enzymes. Regions of relatively high sequence variation generally were predicted as reverse turns, and probably occur as surface loops. Most of the positions corresponding to exon boundaries are located between regions predicted to have secondary-structural elements consistent with a compact structure. The predominantly alternating α/β structure predicted is consistent with the α/β-barrel structure indicated by preliminary high-resolution X-ray diffraction studies on rabbit muscle aldolase [Sygusch, Beaudry & Allaire (1986) Biophys. J. **49**, 287a].

INTRODUCTION

Amino acid sequences of fructose-bisphosphate aldolase from different mammalian tissues and from evolutionarily distant organisms are available, and are compared in the preceding paper (Freemont *et al.*, 1988). The rate of evolution of aldolase is quite low, with about 4% of the residues changing every 100 million years. Thus the mammalian and insect aldolases have about 65–70% of their residues identical, and the mammalian and protozoan aldolases are about 50% identical. The muscle and liver tissue-specific isoenzymes from mammals are about 65% identical. In the present paper we report the results of secondary-structure predictions of these aldolases based on the combination of eight separate prediction methods [Burgess *et al.* (1974); Chou & Fasman (1974), as modified by Lenstra *et al.* (1977) and with the updated indices of Geisow & Roberts (1980); Dufton & Hider (1977); Garnier *et al.* (1978); Lim (1974); McLachlan (1977); Nagano (1973); Kabat & Wu (1973)] as suggested by Eliopoulos *et al.* (1982) and as described in Sawyer *et al.* (1987).

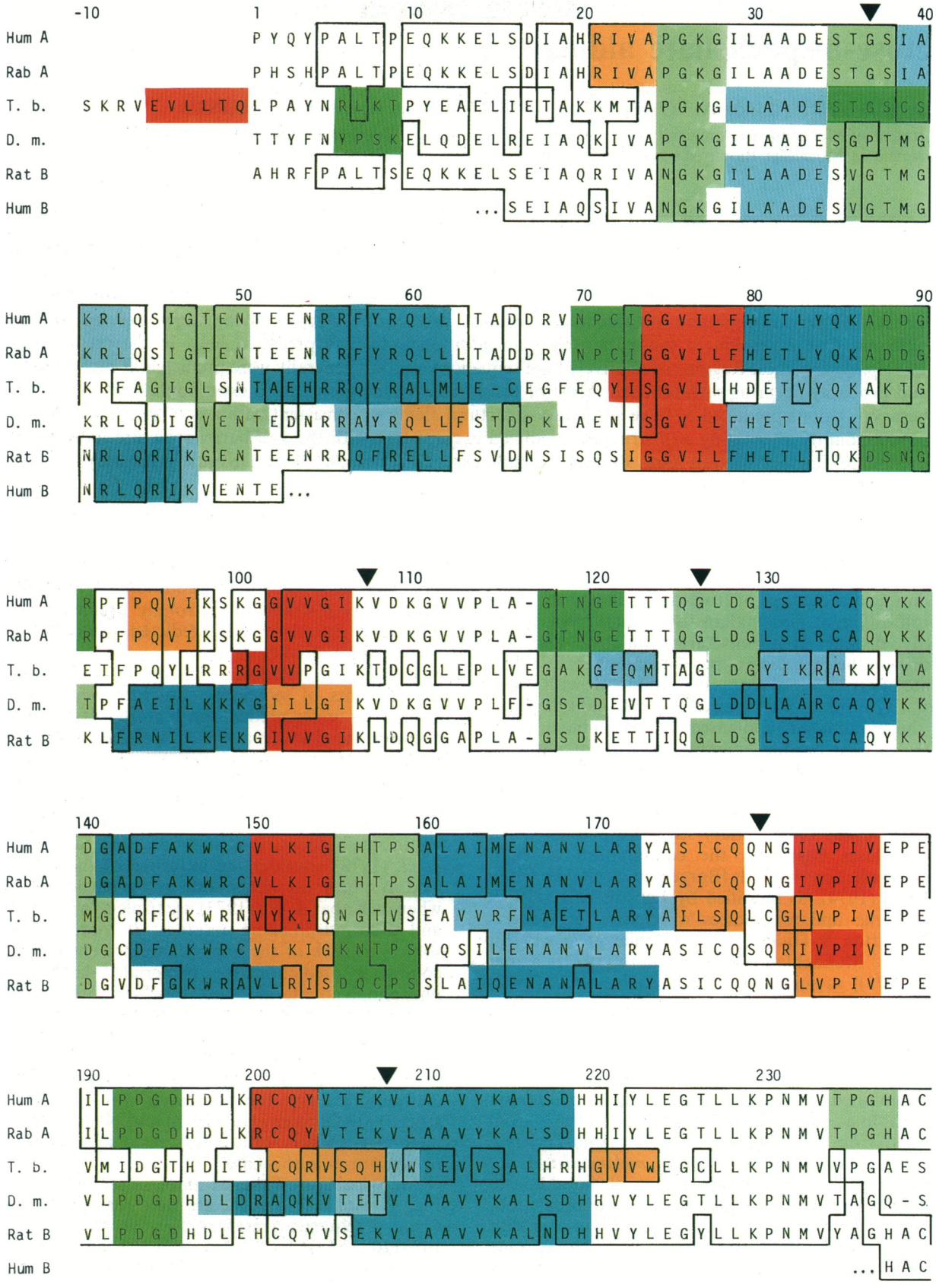
RESULTS AND DISCUSSION

The combined secondary-structure predictions for six different aldolases [from human muscle and liver (partial sequences), from rabbit muscle and liver, from *Trypanosoma brucei* and from *Drosophila melanogaster*] are compared in Fig. 1. It is striking that all the aldolases give essentially the same secondary-structure prediction, despite in some cases having amino acid sequences with as many as 50% of the residues being different. One notable exception to the apparent conservation of secondary structure is the strongly predicted helix at residues 90–100 in the rat liver and *Drosophila* enzymes. The pattern of secondary structure is predominantly alternating α/β, with reverse turns occurring between most of the α/β units. In total, there are 10 or 11 strongly predicted helices, with two additional weakly predicted helices. Six β-strands are strongly predicted, with two additional strands weakly predicted. There are up to 14 reverse turns predicted.

It should be noted that the criteria used for assigning the elements of secondary structure were deliberately chosen to under-predict the features. It is also worth pointing out that it is at the ends of the various elements that greatest uncertainty exists, as is exemplified, for example, by several of the helices in Fig. 1. Further, some identical sequences give different predictions. This results from the extent to which the neighbouring regions contribute in the various methods. For example, Garnier *et al.* (1978) consider the influence of the eight amino acid residues on either side of a particular residue, whereas Kabat & Wu (1973) use triplets to locate α-helix- and β-strand-breaking residues. In order to provide some yardstick of the method, we present in Fig. 2 the combined prediction for bacterial 2-keto-3-deoxy-6-phosphogluconate aldolase, together with its observed secondary structure determined by X-ray crystallography (Mavridis *et al.*, 1982). As can be seen, the helices are quite well predicted, whereas the β-strands and reverse turns are much more open to speculation. The single methods that give the best results for prediction of β-strands are Garnier *et al.* (1978) and Nagano (1973). No method gives very satisfactory results for prediction of reverse turns in this case.

An inspection of the six aldolase sequences shown in Fig. 1 reveals that there are nine regions of relatively variable sequence (residues 1–9, 39–48, 63–72, 91–99, 108–120, 154–164, 234–245, 317–323 and 343–358). All but two of these regions (91–99 and 317–323) include a predicted reverse turn, and presumably correspond to surface loops with rather few functional constraints. Residues 91–99 are of particular interest because, as noted above, they are strongly predicted as α-helix in the

‡ To whom correspondence and reprint requests should be sent. Listings of the individual predictions may be obtained from the authors.


-10
1
10
20
30
40
Hum A
Rab A
T. b.
D. m.
Rat B
Hum B
50
60
70
80
90
100
110
120
130
140
150
160
170
190
200
210
220
230

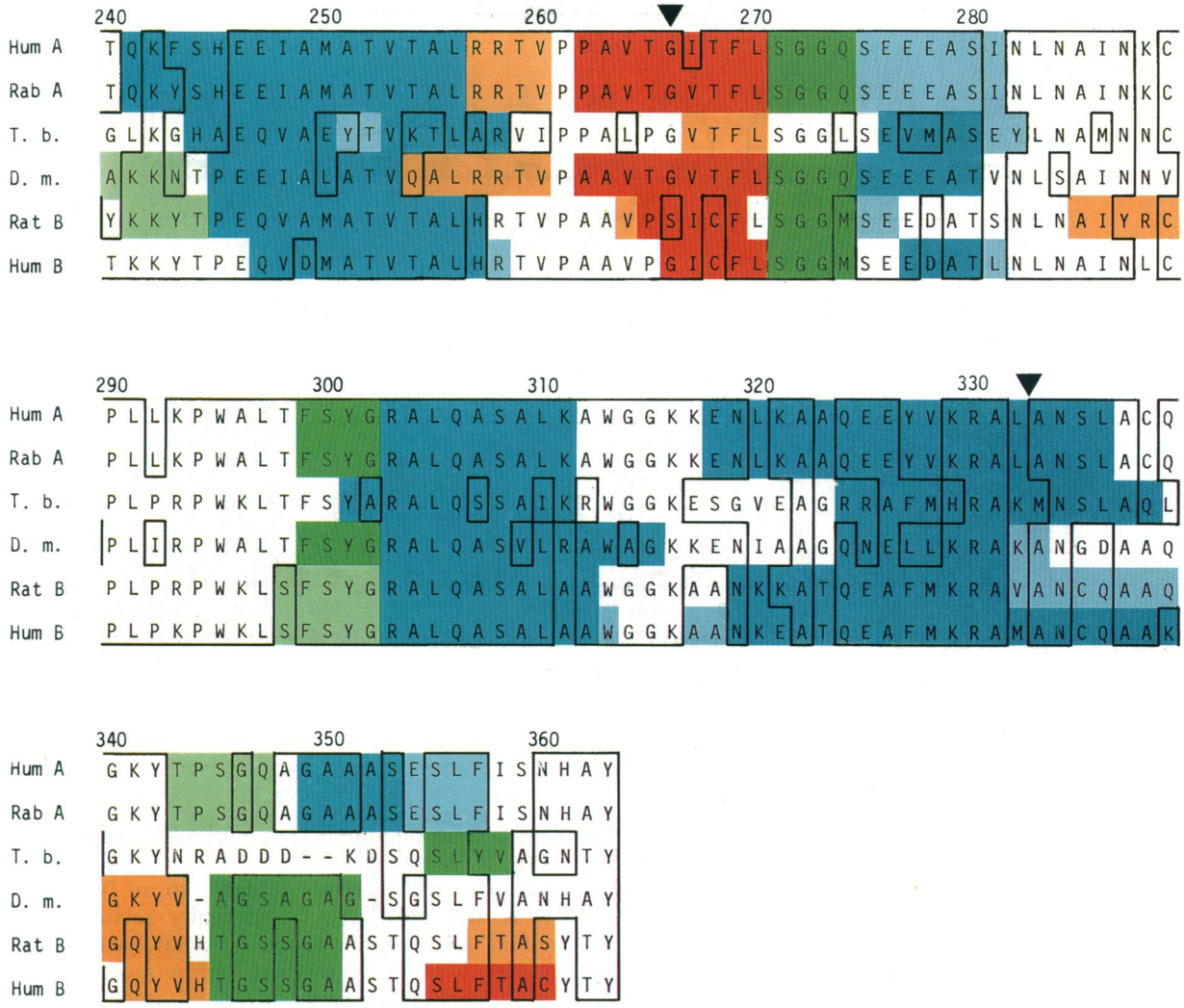


Fig. 1. Combined predictions for α-helix, β-sheet and reverse turn for class I fructose-bisphosphate aldolases

Eight separate secondary-structure predictions were combined for each of the six aldolase sequences. The regions denoted as being strongly predicted for α-helix (dark blue) or β-sheet (dark orange) are those in which at least four consecutive residues are predicted by at least five methods. Weakly predicted α-helix (light blue) or β-sheet (light orange) are those regions with at least four consecutive residues predicted by at least four methods. For reverse turns (strong, dark green; weak, light green), only three consecutive residues are required. The sequences are available in the following publications: Hum A, human muscle (Freemont *et al.*, 1988); Rab A, rabbit muscle (Tolan *et al.*, 1984); T. b., *T. brucei* (Clayton, 1985); D. m., *D. melanogaster* (Malek *et al.*, 1985); Rat B, rat liver (Tsutsumi *et al.*, 1983); Hum B, human liver (partial sequences) (Costanzo *et al.*, 1983; Besmond *et al.*, 1983). The residues are numbered in accordance with the human muscle sequence. The symbol – indicates that a gap has been introduced to maximize homology, and the symbol ▼ shows the positions of exon boundaries found in the rat liver gene. Residues that are identical in three or more sequences are boxed in.

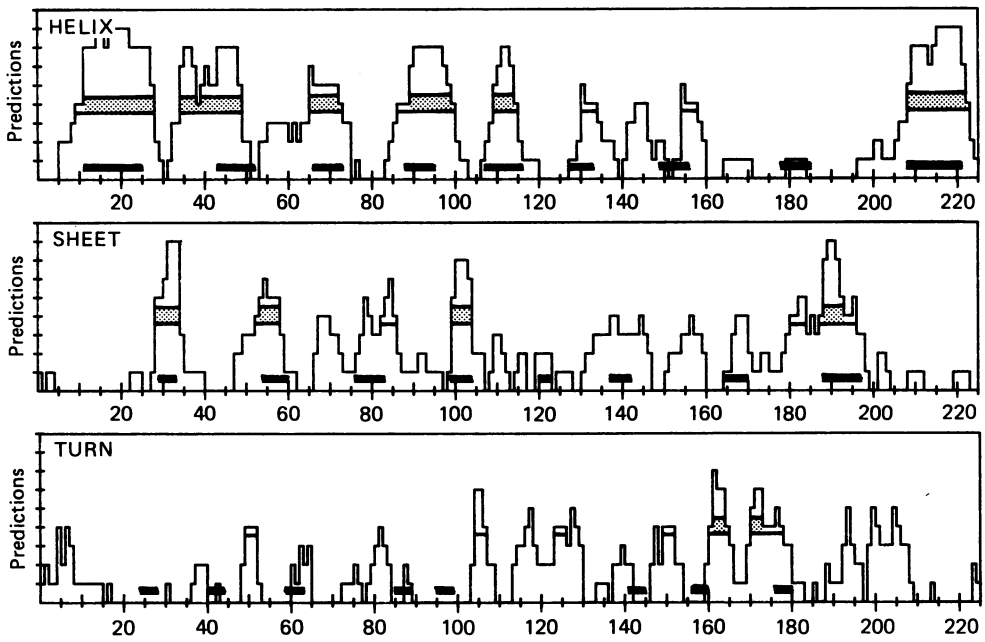


Fig. 2. Combined predictions for 2-keto-3-deoxy-6-phosphogluconate aldolase together with the observed secondary-structure elements

The vertical axis represents the number of separate predictions for each structural element, and the horizontal axis gives the residue number. The horizontal bars half way up each panel show the predicted secondary structure, and the observed secondary structure is indicated by the bars near the foot of each panel.

Drosophila and rat liver sequences, but not in the others.

The frequently observed binding of the phospho moieties of substrate molecules at the *N*-termini of α-helices is considered to be favoured by the partial positive charge from the helix dipole (Hol *et al.*, 1978). Is it possible to identify any of the predicted helices as playing this role in aldolase? A number of residues have been implicated from chemical modification studies as being important in the catalytic mechanism as discussed in the preceding paper (Freemont *et al.*, 1988). One of these residues, Arg-55, is located at the *N*-terminus of a strongly predicted helix. Substrate protection studies have indicated that Arg-55 binds the phospho group on C-1 of fructose bisphosphate and dihydroxyacetone phosphate (Lubini & Christen, 1979; Pathy *et al.*, 1979), and it is likely that the positive helix dipole will also contribute to the binding of this phospho group. Most of the other residues implicated as being important for activity occur in regions devoid of predicted secondary structure (for example, Lys-229, His-361 and Tyr-363).

The positions of the seven exon boundaries that occur in the gene encoding rat liver aldolase (Tsutsumi *et al.*, 1983) are indicated in Fig. 1. With one exception, the positions are located between regions predicted to have secondary structural elements that are consistent with the formation of a compact structure. The third exon (residues 108–126), however, is unusually small, and has no predicted α or β structure. Thus the predicted secondary structures for the aldolases generally conform to the proposal of Go (1981) that exons correspond to compact structural units of a protein.

A previous secondary-structure prediction for aldolase (Stellwagen, 1976) was based on one sequence [rabbit muscle aldolase, now known to have 16 mistakes in the *C*-terminal half of the molecule (Lai, 1975; Tolan *et al.*, 1984) and on one prediction method (Chou & Fasman, 1974)]. It was proposed that residues 147–299 have an alternating α/β structure folded into an NAD domain that is structurally and genetically related to NAD domains such as that found in lactate dehydrogenase. Residues 1–147 were predicted to be mostly β-strands. Overall, nine helices were predicted, and 14 β-strands. Five of the helices correspond to ones strongly predicted in this study: residues 160–172, 241–256, 303–311, 318–336 and 349–353. Four of the β-strands correspond: residues 74–79, 102–106, 200–203 and 262–270.

The overall percentages of residues occurring in secondary structure predicted by this study are 37% in α-helix, 14% in β-sheet and 17% in reverse turns. Previous o.r.d. and c.d. measurements have given rather disparate values for the α-helix content of rabbit muscle aldolase, ranging from 24% to around 40% [Hsu & Neet (1975) (c.d., 29%); Magar (1967) (o.r.d., 38%); Fasman *et al.* (1970) (c.d. from thin films, 40%); Leonard & Walker (1967) (o.r.d., 24% or 40%, depending on the calculation used)]. Our predicted values appear to favour the upper range of the physically measured values. Two of the above publications include c.d. curves, and we have calculated percentages of residues occurring in helix, sheet and random coil by the method of Saxena & Wetlaufer (1971). The c.d. curve measured at pH 7.2 by Hsu & Neet (1975) gives values of 29% α-helix, 24% β-sheet and 47% random coil, whereas the values from the curve determined at pH 8.5 by Fasman *et al.* (1970) are 36%, 24% and 40% respectively.

X-ray-diffraction studies on crystals of aldolase from human muscle, rabbit muscle and *Drosophila* have been undertaken, and low-resolution structures have been published (Millar *et al.*, 1981; Sygusch *et al.*, 1985; Brenner-Holzach & Smit, 1982). A preliminary report of

the high-resolution structure of rabbit muscle aldolase (Sygusch *et al.*, 1986) indicates that most of the polypeptide backbone folds into an alternating α/β structure arranged into an eight-stranded α/β-barrel, such as is found in triose-phosphate isomerase (Banner *et al.*, 1975), in pyruvate kinase (Stuart *et al.*, 1979) and in 2-keto-3-deoxy-6-phosphogluconate aldolase (Mavridis *et al.*, 1982). The secondary-structure predictions are certainly consistent with this type of protein fold, although two β-strands have been missed with the relatively stringent criteria used. We predict that these two strands will be located within residues 220–240 and 281–297, with residues 220–223 and 285–289 being the most likely positions as they are weakly predicted as β-strands in one sequence. It is relevant to note that these putative β-strands correspond to strands βD and βF identified by Stellwagen (1976). It is a general feature of enzymes possessing eight-stranded α/β-barrels that the substrate-binding sites are located at the *N*-termini of the helices forming the barrel (Muirhead, 1983). The observation that Arg-55 of aldolase is probably at the *N*-terminus of a helix would suggest that aldolase also conforms to this general pattern. We are awaiting the publication of the detailed structures of the three different aldolases with considerable interest.

We thank Dr. Hilary Muirhead and Dr. Elias Eliopoulos for supplying versions of the prediction programs, and the Science and Engineering Research Council for financial support.

REFERENCES

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I. & Wilson, I. A. (1975) Nature (London) **255**, 609–614

Besmond, C., Dreyfus, J.-C., Gregori, C., Frain, M., Zakin, M. M., Sala Trepat, J. & Kahn, A. (1983) Biochem. Biophys. Res. Commun. **117**, 601–609

Brenner-Holzach, O. & Smit, J. D. G. (1982) J. Biol. Chem. **257**, 11747–11749

Burgess, A. W., Ponnuswamy, P. K. & Scheraga, H. A. (1974) Isr. J. Chem. **12**, 239–286

Chou, P. Y. & Fasman, G. D. (1974) Biochemistry **13**, 222–245

Clayton, C. E. (1985) EMBO J. **4**, 2997–3003

Costanzo, F., Castagnoli, L., Dente, L., Arcari, P., Smith, M., Costanzo, P., Raugei, G., Izzo, P., Pietropaolo, T. C., Bougueleret, L., Cimino, F., Salvatore, F. & Cortese, R. (1983) EMBO J. **2**, 57–61

Dufton, M. J. & Hider, R. C. (1977) J. Mol. Biol. **115**, 177–193

Eliopoulos, E. E., Geddes, A. J., Brett, M., Pappin, D. J. C. & Findlay, J. B. C. (1982) Int. J. Biol. Macromol. **4**, 263–268

Fasman, G. D., Hoving, H. & Timasheff, S. N. (1970) Biochemistry **9**, 3316–3324

Freemont, P. S., Dunbar, B. & Fothergill-Gilmore, L. A. (1988) Biochem. J. **249**, 779–788

Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) J. Mol. Biol. **120**, 97–120

Geisow, M. J. & Roberts, R. D. B. (1980) Int. J. Biol. Macromol. **2**, 387–389

Go, M. (1981) Nature (London) **291**, 90–92

Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. (1978) Nature (London) **273**, 443–446

Hsu, L. & Neet, K. E. (1975) J. Mol. Biol. **97**, 351–367

Kabat, E. A. & Wu, T. T. (1973) Biopolymers **12**, 751–774

Lai, C. Y. (1975) Arch. Biochem. Biophys. **166**, 358–368

Lenstra, J. A., Hofsteenge, J. & Beintema, J. J. (1977) J. Mol. Biol. **109**, 185–193

Leonard, K. R. & Walker, I. O. (1967) Biochim. Biophys. Acta **133**, 366–369

Lim, V. I. (1974) J. Mol. Biol. **88**, 873–894

Lubini, D. G. & Christen, P. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 2527–2531

Magar, M. E. (1967) J. Biol. Chem. **242**, 2517–2521

Malek, A. A., Suter, F. X., Frank, G. & Brenner-Holzach, O. (1985) Biochem. Biophys. Res. Commun. **126**, 199–205

Mavridis, I. M., Hatada, M. H., Tulinsky, A. & Lebioda, L. (1982) J. Mol. Biol. **162**, 419–444

McLachlan, A. D. (1977) Int. J. Quantum Chem. **12**, Suppl. **1**, 371–385

Millar, J. R., Shaw, P. J., Stammers, D. K. & Watson, H. C. (1981) Philos. Trans. R. Soc. London B **293**, 209–214

Muirhead, H. (1983) Trends Biochem. Sci. **8**, 326–329

Nagano, K. (1973) J. Mol. Biol. **75**, 401–420

Pathy, L., Varadi, A., Thesz, J. & Kovacs, K. (1979) Eur. J. Biochem. **99**, 309–313

Sawyer, L., Tollin, P. & Wilson, H. (1987) J. Gen. Virol. **68**, 1229–1232

Saxena, V. P. & Wetlaufer, D. B. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 969–972

Stellwagen, E. (1976) J. Mol. Biol. **106**, 903–911

Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. (1979) J. Mol. Biol. **134**, 109–142

Sygusch, J., Boulet, H. & Beaudry, D. (1985) J. Biol. Chem. **260**, 15286–15290

Sygusch, J., Beaudry, D. & Allaire, M. (1986) Biophys. J. **49**, 287a

Tolan, D. R., Amsden, A. B., Putney, S. D., Urdea, M. S. & Penhoet, E. E. (1984) J. Biol. Chem. **259**, 1127–1131

Tsutsumi, K., Mukai, T., Hidaka, S., Miyahara, H., Tsutsumi, R., Tanaka, T., Horis, K. & Ishikawa, K. (1983) J. Biol. Chem. **258**, 6537–6542

Received 17 July 1987/24 September 1987; accepted 28 September 1987