university of groningen

University Medical Center Groningen

University of Groningen

Prediction of the Occurrence of the ADP-binding βαβ-fold in Proteins, Using an Amino Acid Sequence Fingerprint

Wierenga, Rik K.; Terpstra, Peter; Hol, Wim G.J.

Published in:
Journal of Molecular Biology

DOI:
10.1016/0022-2836(86)90409-2

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1986

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Wierenga, R. K., Terpstra, P., & Hol, W. G. J. (1986). Prediction of the Occurrence of the ADP-binding βαβ-fold in Proteins, Using an Amino Acid Sequence Fingerprint. Journal of Molecular Biology, 187(1). DOI: 10.1016/0022-2836(86)90409-2

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-02-2018

J. Mol. Biol. (1986) **187**, 101–107

Prediction of the Occurrence of the ADP-binding βαβ-fold in Proteins, Using an Amino Acid Sequence Fingerprint

Rik K. Wierenga, Peter Terpstra and Wim G. J. Hol

Biomolecular Study Centre (BIOS)
University of Groningen, Nijenborgh 16
9747 AG Groningen, The Netherlands

(Received 20 March 1985, and in revised form 30 July 1985)

An amino acid sequence "fingerprint" has been derived that can be used to test if a particular sequence will fold into a $\beta\alpha\beta$-unit with ADP-binding properties. It was deduced from a careful analysis of the known three-dimensional structures of ADP-binding $\beta\alpha\beta$-folds. This fingerprint is in fact a set of 11 rules describing the type of amino acid that should occur at a specific position in a peptide fragment. The total length of this fingerprint varies between 29 and 31 residues. By checking against all possible sequences in a database, it appeared that every peptide, which exactly follows this fingerprint, does indeed fold into an ADP-binding $\beta\alpha\beta$-unit.

1. Introduction

It is generally accepted that the three-dimensional structure of a folded polypeptide chain is determined by its amino acid sequence (Creighton, 1983). Nevertheless similar folds can have very different sequences. An example is the very similar structure of the two domains of rhodanese, despite very dissimilar sequences (Ploegman *et al.*, 1978). Somehow these sequences must have some homology at crucial positions in order to ensure a similar folding. Clear insight in the crucial amino acid sequence pattern that determines the three-dimensional structure of a protein is still lacking. This is also reflected in the current methods of secondary structure prediction, which achieve on average less than 60% accuracy (Kabsch & Sander, 1983). Recently, Taylor & Thornton (1984) have outlined a procedure for the recognition of the $\beta\alpha\beta$-super secondary structure with 75% accuracy for α/β proteins. Here we will focus on the recognition of the $\beta\alpha\beta$-structures which are involved in FAD-binding or NAD-binding. It has been shown that in the family of FAD-binding and NAD-binding domains (e.g. Rossmann *et al.*, 1974; Ohlsson *et al.*, 1974; Wierenga *et al.*, 1983) a particular $\beta\alpha\beta$-unit is a recurrent fold which in all known complexes binds the ADP-moiety of the dinucleotide in the same manner (Wierenga *et al.*, 1985). Therefore we will refer to this $\beta\alpha\beta$-unit as the "*ADP-binding $\beta\alpha\beta$-fold*".

From the known structures and sequences it appeared that the amino acid sequence of this $\beta\alpha\beta$-fold has specific features (e.g. Rossmann *et al.*, 1974). We have derived a "fingerprint", which is summarized in Table 1. This pattern was obtained by recognizing essential structural functions of several side-chains of this $\beta\alpha\beta$-fold, as described in detail by Wierenga *et al.* (1985). Altogether, five proteins have been used to construct the fingerprint: lobster glyceraldehyde-phosphate dehydrogenase, spiny dogfish lactate dehydrogenase, horse liver alcohol dehydrogenase, human glutathione reductase and bacterial *p*-hydroxy benzoate hydroxylase. As an example, the ADP-binding $\beta\alpha\beta$-fold of spiny dogfish M-lactate dehydrogenase is depicted in Figure 1. Its sequence follows exactly the fingerprint of Table 1. This fingerprint describes the type of amino acid that consistently occurs at a specific position within the ADP binding $\beta\alpha\beta$-fold. Altogether there are 11 specific positions; the consistent occurrence of specific residues at these positions can be understood from the three-dimensional structure. The hydrophobic residues (indicated by an open square) form the hydrophobic core between the β-strands and the α-helix. The glycine residues (indicated by a filled circle) allow for a sharp turn between the first β-strand and the following α-helix. The middle glycine, at position 8 (Table 1), allows for a close approach of the pyrophosphate to the N terminus of the α-helix. The acid side-chain (indicated by a divided open circle) forms a good hydrogen bond with the 2′-OH of the ADP-ribose (Fig. 1). The total length of the fingerprint is not entirely constant because the loop between the α-helix and the second β-strand has a

0022-2836/86/010101–07 $03.00/0

© 1986 Academic Press Inc. (London) Ltd.

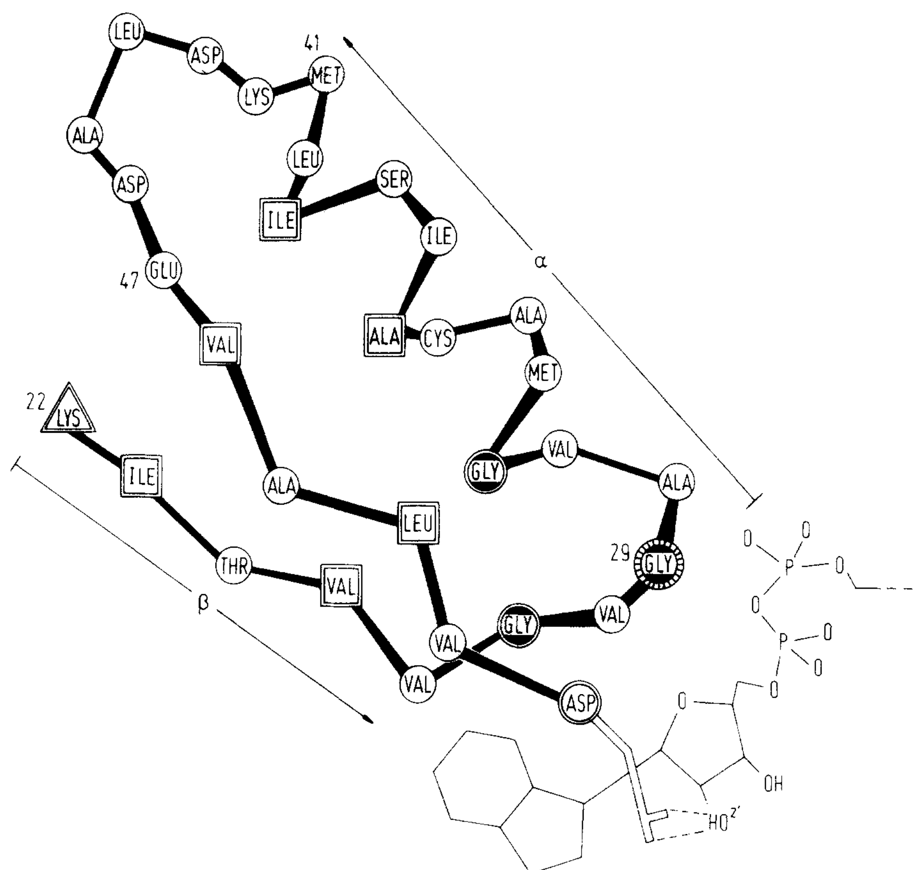

Figure 1. Schematic drawing of the ADP-binding $\beta\alpha\beta$-fold of spiny dogfish M-lactate dehydrogenase (Taylor, 1977; Eventoff *et al.*, 1977). The residues at the fingerprint positions are indicated by special symbols, as also used in Table 1 (△, basic or hydrophilic; □, small and hydrophobic). At these 11 positions the residue names are framed by 2 lines. Gly29 allows for a close approach of the pyrophosphate to the N terminus of the α-helix. Residues 43, 44, 45 and 46 form the loop between the α-helix and the second β-strand. The total length of this ADP-binding $\beta\alpha\beta$-fold is 31 residues.

variable length in the known structures. Here we report investigations into the uniqueness of this fingerprint.

2. Methods

We used the amino acid sequences which were available in the Protein Sequence Database of the Protein Identification Resource (PIR; Barker *et al.*, 1984). This database (released April 1984) contains 2676 sequences. Altogether 5 searches were carried out with fingerprints of, respectively, 28, 29, 30, 31 and 32 residues. This variable length of the probe allows for different lengths of the loop between the α-helix and the second β-strand (Table 1 and Fig. 1). The searches were performed by the program SEARCH (Dayhoff *et al.*, 1983), using a proper scoring matrix and a corresponding probe sequence. A very simple scoring scheme was used. The minimal score of a peptide is 0. The score is raised by 1 whenever at a specific position within the peptide the amino acid at that position corresponds with the requirements of the fingerprint. The maximal score therefore is 11, since there are 11 specific positions.

After the scores of all possible peptides have been calculated the sequences with the highest scores are printed. Here we will discuss only features of those sequences which have scores of 11, 10 or 9. The scores of 10 and 9 occur whenever the amino acid sequence of a peptide deviates from the fingerprint at one or two positions, respectively. Careful analysis of the sequences of the ADP-binding $\beta\alpha\beta$-folds of which the structures are known, indicates that the type of residue at some positions is strictly conserved, whereas at other positions exceptions to the fingerprint occur. The 3 glycine residues, which are present in all known structures of NAD- or FAD-binding $\beta\alpha\beta$-units, as well as the acid residue, at the end of the fingerprint (Table 1) are strictly required. Some variability, however, exists at the first position of the fingerprint and at the positions of the hydrophobic residues (Table 1). Therefore, we have defined a "core" fingerrpint consisting of the 3 glycine and the acid residues. This allows for some variability at the first position as well as at the positions of the hydrophobic residues. Sequences with a score of 10 or 9 are only considered if they still agree with this core fingerprint.

The statistics of the 5 searches are summarized in Table 2. Some peptides are found, with different scores, in more than one search (with different probe lengths); in those cases only the highest score is tabulated. For all 5 searches, the average score is 2·6, and the standard deviation (σ) is 1·4. Therefore, a score of 11 is 6σ above average. This corresponds to a chance of 1 out of $\sim 10^9$. Similarly a score of 10 is $5{\cdot}3\sigma$ above average (1 out of $\sim 10^6$) and a score of 9 is $4{\cdot}6\sigma$ above average (1 out of $\sim 10^5$).

Table 1
The amino acid sequence fingerprint

Position	Secondary structure	Fingerprint position (indicated by a special symbol)	Residues obeying the fingerprint	Comments	Exceptions, as observed in known structures
1	β	△	K, R, H, S, T, Q, N	Basic or hydrophilic	D (Schulz *et al.*, 1982)
2	β	□	A, I, L, V, M, C	Small and hydrophobic	Y (Schulz *et al.*, 1982)
3	β				
4	β	□	A, I, L, V, M, C	Small and hydrophobic	—
5	β				
6	β	●	G	Glycine	
7					
8	α	⬤	G	Glycine	—
9	α				
10	α				
11	α	●	G	Glycine	
12	α				
13	α				
14	α				
15	α	□	A, I, L, V, M, C	Small and hydrophobic	G (Wierenga *et al.*, 1983) F (Biesecker *et al.*, 1977)
16	α				
17	α				
18	α	□	A, I, L, V, M, C	Small and hydrophobic	
19	α				
20	α				
21	α				
22–26	Loop†				
27	β				
28	β	□	A, I, L, V, M, C	Small and hydrophobic	N (Wierenga *et al.*, 1983)
29	β				
30	β	□	A, I, L, V, M, C	Small and hydrophobic	G (Eklund *et al.*, 1981)
31	β				
32	β	⊖	D, E	Acid	

This fingerprint was derived from a comparison of the ADP-binding βαβ-folds in the following 5 proteins: lobster glyceraldehyde-3-phosphate dehydrogenase (NAD; "GPDLO", see Table 3); spiny dogfish lactate dehydrogenase (NAD, M-chain; "LDHMS"); horse liver alcohol dehydrogenase (NAD; "ADHHO"); human glutathione reductase (FAD; "GRHUM") and bacterial *p*-hydroxybenzoate hydroxylase (FAD; "PHBHP") (Wierenga & Hol, 1983; Wierenga *et al.*, 1985).

† The length of the loop can vary somewhat.

3. Results

Let us now discuss the peptides which gave the highest scores.

(a) *Sequences with a score of 11*

This score, an exact match with the fingerprint, is found for only 11 peptides. If the fingerprint is a truly unique amino acid sequence pattern of the ADP-binding βαβ-fold then these 11 peptides should all fold into this structure. Indeed, as can be seen from Table 3A, the 11 peptides all occur in proteins that bind NAD. Our prediction is confirmed by the known crystallographic analysis of these proteins themselves or because the structure of a homologous protein is known. Only for L-3-hydroxyacyl CoA-dehydrogenase the prediction of the ADP-binding βαβ-fold of its peptide 17 to 45 cannot yet be supported by a crystallographic analysis. The score of 11 occurs only in peptides with a total length of 29, 30 or 31 residues. Therefore, in these ADP-binding βαβ-folds the loop consists of 2, 3 or 4 residues, respectively.

Table 2
Summary of the search with the fingerprint using the 2676 sequences in the database

Probe length	Number of peptides with: score = 11	score = 10†	score = 9†
28	0	0	0
29	1	0	4
30	2	4	2
31	8	3	2
32	0	0	0

About 435,000 comparisons were made in each search. The average score is 2·6 and the standard deviation is 1·4.

† Only those peptides are counted which agree with the "core" fingerprint (see the text).

(b) *Sequences with a score of ten*

Each sequence with a score of ten deviates at only one position from the fingerprint. These deviations are explicitly shown in Table 4. The seven peptides with a score of ten most likely are folded as ADP-binding βαβ-units (Table 3B). For horse liver alcohol dehydrogenase and bacterial glyceraldehyde phosphate dehydrogenase (GPDBS) this is confirmed by the known structures them-

Table 3
Sequence scores

Code	Enzyme	Probe length	Position
A. Sequences with a score of 11			
GPDLO	Glyceraldehyde phosphate dehydrogenase (lobster)	30	2–31
GPD3B	Glyceraldehyde phosphate dehydrogenase-3 (baker's yeast)	30	2–31
GPD2B	Glyderaldehyde phosphate dehydrogenase-2 (baker's yeast)	31	2–32
LDHMS	Lactate dehydrogenase, M-chain (spiny dogfish)	31	22–52
LDHXM	Lactate dehydrogenase, X-chain (mouse)	31	20–50
LDHHP	Lactate dehydrogenase, H-chain (pig)	31	22–52
LDHHC	Lactate dehydrogenase, H-chain (chicken)	31	21–51
LDHMP	Lactate dehydrogenase, M-chain (pig)	31	21–51
LDHMC	Lactate dehydrogenase, M-chain (chicken)	31	21–51
LDHLC	Lactate dehydrogenase (*Lactobacillus casei*)	31	10–40
HCDHP	L-3-Hydroxyacyl-CoA-dehydrogenase (pig)	29	17–45
B. Sequences with a score of 10			
ADHHO	Alcohol dehydrogenase (horse) E-chain	30	194–223
GPDPI	Glyceraldehyde phosphate dehydrogenase (pig)	31	2–32
GPDHU	Glyceraldehyde phosphate dehydrogenase (human)	31	4–34
GPDBS	Glyceraldehyde phosphate dehydrogenase (*Bacillus stearothermophilus*)	31	3–33
GPDTA	Glyceraldehyde phosphate dehydrogenase (*Thermus aquaticus*)	30	6–36
DAOPI	D-Amino acid oxidase (pig)	30	2–31
GDHBO	Glutamate dehydrogenase (bovine)	30	246–275
C. Sequences with a score of 9			
GRHUM	Glutathione reductase (human)	29	22–50
PHBHP	*p*-Hydroxybenzoate hydroxylase (*Pseudomonas fluorescens*)	29	4–32
ADHFF	Alcohol dehydrogenase (fruit fly)	30	10–39
P21HU	p21 (gene product of H-ras-1, human)	29	5–33
CATBO	Catalase (bovine)	29	72–100
CYCAV	Cytochrome *c*-551 (*Azotobacter vinelandii*)	30	47–76
IGLHU	Bence-Jones protein Del (v-domain)	31	18–48
PKABO	c-AMP-dep. Protein kinase, catalytic chain (bovine)	31	45–75

Table 4
The deviations from the fingerprint, as observed in the sequences of Table 3B and C

Sequence	Fingerprint positions △	□	□	●	⬤	●	□	□	□	□	⊖	Score	Reference†
ADHHO										G		10	Eklund *et al.* (1981)
GPDPI							T					10	
GPDHU							T					10	
GPDBS							F					10	Biesecker *et al.* (1977)
GPDTA							F					10	
DAOPI										P		10	
GDHBO		F										10	
	△	□	□	●	⬤	●	□	□	□	□	⊖		
GRHUM	D	Y										9	Schulz *et al.* (1982)
PHBHP							G		N			9	Wierenga *et al.* (1983)
ADHFF	V						S					9	
P21HU								Q		E		9	
CATBO	V							D				9	Murthy *et al.* (1981)
CYCAV								N	K			9	
IGLHU							H	Q				9	
PKABO			T					H				9	

† Known structures are described in these references.

Table 3 *(continued)*

Reference (sequence/structure)	Dinucleotide	Comment
Davidson *et al.* (1967; Moras *et al.* (1975)	NAD	Agreement with the known structure
Holland *et al.* (1983)	NAD	Agreement with the homologous structure GPDLO (Hocking & Harris, 1980; Holland & Holland, 1979)
Holland & Holland (1979)	NAD	Agreement with the homologous structure GPDLO (Hocking & Harris, 1980; Holland & Holland, 1979)
Taylor (1977); Eventoff *et al.* (1977)	NAD	Agreement with the known structure
Pan *et al.* (1980); Musick & Rossmann (1979)	NAD	Agreement with the known structure
Kiltz *et al.* (1977); Grau *et al.* (1981)	NAD	Agreement with the homologous structures LDHMS (Eventoff *et al.*, 1977)
Torff *et al.* (1977)	NAD	Agreement with the homologous structures LDHMS (Eventoff *et al.*, 1977)
Kiltz *et al.* (1977)	NAD	Agreement with the homologous structures LDHMS (Eventoff *et al.*, 1977)
Torff *et al.* (1977)	NAD	Agreement with the homologous structures LDHMS (Eventoff *et al.*, 1977)
Hensel *et al.* (1983)	NAD	Agreement with the homologous structures LDHMS (Hensel *et al.*, 1983)
Bitar *et al.* (1980)		NAD-binding enzyme
Jörnvall (1970); Eklund *et al.* (1981)	NAD	Agreement with the known structure
Harris & Perham (1968)	NAD	Agreement with the homologous structure GPDLO (Hocking & Harris, 1980)
Nowak *et al.* (1981)	NAD	Agreement with the homologous structure GPDLO (Nowak *et al.*, 1981)
Biesecker *et al.*, 1977)	NAD	Agreement with the known structure
Hocking & Harris (1980)	NAD	Agreement with the homologous structure GDPBS (Hocking & Harris, 1980)
Ronchi *et al.* (1982)	FAD	Agreement with a published alignment (Swenson *et al.*, 1982)
Moon & Smith (1973)	NAD	Agreement with a published alignment (Rossmann *et al.*, 1974; Wootton, 1974)
Krauth-Siegel *et al.* (1982); ; Schulz *et al.* (1982)	FAD	Agreement with the known structure
Weyer *et al.* (1982); Wierenga *et al.* (1979)	FAD	Agreement with the known structure
Kreitman (1983)	NAD	Agreement with the homologous structure (Thatcher & Sawyer, 1980)
Capon *et al.* (1983)		βαβ-fold predicted after model-building (Wierenga & Hol, 1983)
Schroeder *et al.* (1982); Murthy *et al.* (1981)		Not a βαβ-fold (Murthy *et al.*, 1981)
Ambler (1973)		Not a βαβ-fold, as observed in homologous structures (Dickerson *et al.*, 1976)
Eulitz (1974)		Not a βαβ-fold, as observed in a homologous structure (Epp *et al.*, 1974)
Shoji *et al.* (1981)		Unknown structure

selves. The sequences of pig and human glyceraldehyde phosphate dehydrogenases are sufficiently similar to the sequences of lobster and bacterial glyceraldehyde phosphate dehydrogenase, whose structures are known, to be confident about their folding into ADP-binding βαβ-units.

There are no convincing structural data with respect to the peptides of D-amino acid oxidase and glutamate dehydrogenase. If the D-amino acid oxidase peptide does fold as a βαβ-unit, then the second β-strand will have a proline residue near its C terminus (Table 4). For both D-amino acid oxidase and glutamate dehydrogenase our predictions agree with previous predictions (Swenson *et al.*, 1982; Rossmann *et al.*, 1974; Wootton, 1974).

(c) *Sequences with a score of nine*

Altogether eight sequences with a score of nine have been found (Table 3C). The two peptides of human glutathione reductase and bacterial *p*-hydroxybenzoate hydroxylase do form βαβ-units as is shown by their known structures (Schulz *et al.*, 1982; Wierenga *et al.*, 1983). The peptide from alcohol dehydrogenase of the fruit fly most likely folds as a βαβ-unit, as is indicated by the general sequence homology with the sequence of horse liver alcohol dehydrogenase (Kreitman, 1983; Thatcher & Sawyer, 1980). However, the peptides of catalase, cytochrome *c*-551 and a Bence Jones protein do not fold as βαβ-units, as can be concluded from the available structural data (Table 3C). The structures of the p21-protein, a protein with GTPase activity (McGrath *et al.*, 1984) encoded by the "ras"-proto oncogene, and of the catalytic chain of the cAMP-dependent protein kinase are not known. For the peptide of p21, model-building studies have shown that this peptide can be built as an ADP-binding βαβ-fold (Wierenga & Hol, 1983). In this model, the deviating Glu31 contributes *via* the hydrophobic part of its side-chain to the hydrophobic core, while its carboxyl moiety can form a salt bridge with Lys16.

4. Discussion

In the PIR sequence database, a number of peptides have been found (Table 3A) whose sequences agree completely with the fingerprint sequence of the ADP-binding $\beta\alpha\beta$-fold. These peptides, with a score of 11, always belong to NAD-binding enzymes. Whenever the structures of these enzymes are known, it can be confirmed that these peptides fold as an ADP-binding $\beta\alpha\beta$-fold. Therefore, the presence of this fingerprint in the sequence of a protein apparently ensures the occurrence of an ADP-binding $\beta\alpha\beta$-fold in the structure of that protein.

Sequences with a score of ten or nine have also been tabulated (Table 3B and C). Only sequences which do not disagree with the "core"-fingerprint have been considered. This "core"-fingerprint demands glycine residues at positions 6, 8 and 11. Moreover, an acid residue should occur at the end of the fingerprint. This requirement implies that we have restricted our search to the NAD- or FAD-binding $\beta\alpha\beta$-folds, and we have not considered the NADP-binding $\beta\alpha\beta$-folds, such as for example occurs in glutathione reductase (Wierenga *et al.*, 1983). In this $\beta\alpha\beta$-fold the acid residue is replaced by a hydrophobic residue. Because of this change a very unfavourable interaction with the negatively charged 2′-phosphate of NADP is avoided.

All accepted sequences with a score of ten most likely fold as ADP-binding $\beta\alpha\beta$-units (Table 3B). Apparently, the agreement of the sequence with the fingerprint is a more powerful criterion for structure prediction than sequence homology. For example the sequence homology between the peptides GPDBS (score of 10, see Table 3B) and ADHHO (score of 10, see Table 3B) is only 16%, nevertheless these peptides are known to be folded as ADP-binding $\beta\alpha\beta$-units.

Sequences with a score of nine do not always fold as $\beta\alpha\beta$-units (Table 3C). In the three sequences which are known to have no $\beta\alpha\beta$-structure, a neutral hydrophobic residue is replaced by a charged residue (Table 4). Apparently this replacement does not favour the folding of a $\beta\alpha\beta$-unit. If the "core"-fingerprint is changed, such that the hydrophobic residues are not allowed to be replaced by charged residues, then all remaining peptides with a score of nine are known to have a $\beta\alpha\beta$-structure. More sequences and structures are needed to verify the allowed variability of these hydrophobic residues.

Thus, the fingerprint search is a convenient way of predicting ADP-binding $\beta\alpha\beta$-folds with great confidence if the score is ten or 11. For scores of nine considerable care has to be exercised. It may also be pointed out that some enzymes with even lower scores may still have an ADP-binding $\beta\alpha\beta$-fold quite similar to the folds used for defining the fingerprint. For instance, yeast alcohol dehydrogenase (Jörnvall, 1970; Bennetzen & Hall, 1982; Table VIII of Wierenga *et al.*, 1985) only has a score of seven and was therefore not found by our search, but is quite likely to be closely related in structure to horse liver alcohol dehydrogenase. This sequence was not found for a second reason: one of the essential glycine residues of the "core"-fingerprint is changed into an alanine residue. Cytoplasmic porcine malate dehydrogenase is also reported to have an alanine substituted for one of these essential glycine residues in our fingerprint (Birktoft *et al.*, 1982) and was therefore not picked up by the search described in this paper. This sequence is an "X-ray sequence", however, and one would like to see this confirmed by an independent method. It is obvious that knowledge of the three-dimensional structure of yeast alcohol dehydrogenase complexed with NAD, and of the definite amino acid sequence of malate dehydrogenase would be of great value for improving our fingerprint, or for deriving alternative fingerprints which are reliable in a search for slightly different (di)nucleotide-binding $\beta\alpha\beta$-folds in protein sequence data bases.

The peptides known to form the ADP-binding $\beta\alpha\beta$-unit vary in length from 29 to 31 residues (Table 2). Although no structures are known in which the $\beta\alpha\beta$-fold is longer than 31 residues, there is no strong *a priori* reason why this should not be possible. For example it might occur in NADH-dehydrogenase (Young *et al.*, 1981), in which residues 7 to 40 exactly follow a fingerprint of 34 residues, having a loop of seven residues.

In at least one case, alcohol dehydrogenase of the fruit fly, it is known that the position of this loop coincides with the position of an intron in the DNA code (Kreitman, 1983). This is somewhat surprising, since the $\beta\alpha\beta$-fold can be considered as an excellent example of a small structural entity, which might be encoded by one exon. However, an intron at this position agrees with the views of Craick *et al.* (1982), stating that introns always occur at a position in a protein near its surface.

Finally, we would like to make one more comment. In the known structures these $\beta\alpha\beta$-structural entities do always occur near the N terminus of the dinucleotide-binding domains (Table 3). This suggests (Schulz & Schirmer, 1979) that these $\beta\alpha\beta$-folds might function as a nucleation centre for the folding of the complete domain. First the $\beta\alpha\beta$-fold is formed, subsequently the remaining polypeptide curls around this core to complete the dinucleotide binding domain.

Our studies have shown that the amino acid sequence fingerprint of this $\beta\alpha\beta$-structural entity is a powerful tool for structure prediction. A score of ten or 11 is consistently observed for peptides which do occur in dinucleotide proteins, and, whenever this can be checked, these peptides indeed fold as ADP-binding $\beta\alpha\beta$-units. The success of our studies suggests strongly that the use of precisely defined fingerprints may be of general use in the prediction of protein tertiary structure.

We are indebted to Drs B. Witholt and J. Drenth for their interest and support. Computations were carried out on the VAX-11/750 at the Department of Chemistry, University of Groningen.

References

Ambler, R. P. (1973). *Syst. Zool.* **22**, 554–565.

Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C., Seibel-Ross, E. I. & Ledky, R. S. (1984). *Report of the National Biomedical Research Foundation*, Georgetown University, Washington, D.C.

Bennetzen, J. L. & Hall, B. D. (1982). *J. Biol. Chem.* **257**, 3018–3025.

Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. (1977). *Nature (London)*, **266**, 328–333.

Birktoft, J. J., Fernley, R. T., Bradshaw, R. A. & Banaszak, L. J. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 6166–6170.

Bitar, K. G., Perez-Aranda, A. & Bradshaw, R. A. (1980). *FEBS Letters*, **116**, 196–198.

Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. (1983). *Nature (London)*, **302**, 33–37.

Craick, C. S., Sprang, S., Fletterick, R. & Rutter, W. J. (1982). *Nature (London)*, **299**, 180–182.

Creighton, T. E. (1983). *Proteins: Structures and Molecular Properties*, pp. 286–313, Freeman, New York.

Davidson, B. E., Sajgo, M., Noller, H. F. & Harris, J. I. (1967). *Nature (London)*, **216**, 1181–1185.

Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983). *Methods Enzymol.*, **91**, 524–545.

Dickerson, R. E., Timkovich, R. & Almassy, R. J. (1976). *J. Mol. Biol.* **100**, 473–491.

Eklund, H., Samana, J.-P., Wallén, L., Brändén, C.-I., Åkeson, Å. & Jones, T. A. (1981). *J. Mol. Biol.* **146**, 561–587.

Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974). *Eur. J. Biochem.* **45**, 513–524.

Eulitz, M. (1974). *Eur. J. Biochem.* **50**, 49–69.

Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H.-J., Meyer, H., Keil, W. & Kiltz, H.-H. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 2766–2681.

Grau, U. M., Frommer, W. E. & Rossmann, M. G. (1981). *J. Mol. Biol.* **151**, 289–307.

Harris, J. I. & Perham, R. N. (1968). *Nature (London)*, **219**, 1025–1028.

Hensel, R., Mayr, U. & Yang, C. (1983). *Eur. J. Biochem.* **134**, 503–511.

Hocking, J. D. & Harris, J. I. (1980). *Eur. J. Biochem.* **108**, 567–579.

Holland, J. P. & Holland, M. J. (1979). *J. Biol. Chem.* **254**, 9839–9845.

Holland, J. P., Labieniec, L., Swimmer, C. & Holland, M. J. (1983). *J. Biol. Chem.* **258**, 5291–5299.

Jörnvall, H. (1970). *Eur. J. Biochem.* **16**, 41–49.

Kabsch, W. & Sander, C. (1983). *FEBS Letters*, **155**, 179–182.

Kiltz, H.-H., Keil, W., Griesbach, M., Petry, K. & Meyer, H. (1977). *Hoppe-Seyler's Z. Physiol. Chem.* **358**, 123–127.

Krauth-Siegel, R. L., Blatterspiel, R., Saleh, M., Schilz, E., Schirmer, R. H. & Untucht-Grau, R. (1982). *Eur. J. Biochem.* **121**, 259–267.

Kreitman, H. (1983). *Nature (London)*, **304**, 412–417.

McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. (1984). *Nature (London)*, **310**, 644–649.

Moon, K. & Smith, E. L. (1973). *J. Biol. Chem.* **248**, 3082–3088.

Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975). *J. Biol. Chem.* **250**, 9137–9162.

Murthy, M. R. N., Reid, T. J. III, Sicigano, A., Tanaka, N. & Rossmann, M. G. (1981). *J. Mol. Biol.* **152**, 465–499.

Musick, W. D. L. & Rossmann, M. G. (1979). *J. Biol. Chem.* **254**, 7611–7620.

Nowak, K., Wolny, M. & Banas, T. (1981). *FEBS Letters*, **134**, 143–146.

Ohlsson, I., Nordström, B. & Brändén, C.-I. (1974). *J. Mol. Biol.* **89**, 339–354.

Pan, Y. C. E., Huang, S., Marciniszyn, J. P. Jr, Lee, C.-Y. & Li, S. S.-L. (1980). *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 795–799.

Ploegman, J. H., Drent, G., Kalk, K. H., Hol, W. G. J., Heinrickson, R. L., Keim, P., Weng, L. & Russell, J. (1978). *Nature (London)*, **273**, 124–129.

Ronchi, J., Minchiott, L., Galliano, M., Curti, B., Swenson, R. P., Williams, C. H. & Massey, V. (1982). In *Flavins and Flavoproteins* (Massey, V. & Williams, C., eds), pp. 66–72, Elsevier/North-Holland, Amsterdam.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). *Nature (London)*, **250**, 194–199.

Schroeder, W. A., Shelton, J. R., Shelton, J. B., Robberson, B., Apell, G., Fang, R. S. & Bonaventura, J. (1982). *Arch. Biochem. Biophys.* **214**, 397–421.

Schulz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, pp. 128–129, Springer Verlag, New York.

Schulz, G. E., Schirmer, R. H. & Pai, E. F. (1982). *J. Mol. Biol.* **160**, 287–308.

Shoji, S., Parmelee, D. C., Wade, R. D., Kumar, S., Ericsson, L. H., Walsh, K. A., Neurath, H., Long, G. L., Demaille, J. G., Fisher, E. H. & Titani, K. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 848–851.

Swenson, R. P., Williams, C. H. Jr, Massey, V., Ronchi, S., Minchiotti, L., Galliano, M., Bolognesi, M. & Curti, B. (1982). In *Flavins and Flavoproteins* (Massey, V. & Williams, C., eds), pp. 73–77. Elsevier/North-Holland, Amsterdam.

Taylor, S. S. (1977). *J. Biol. Chem.* **252**, 1799–1806.

Taylor, W. R. & Thornton, J. M. (1984). *J. Mol. Biol.* **173**, 487–514.

Thatcher, D. R. & Sawyer, L. (1980). *Biochem. J.* **187**, 884–886.

Torff, H.-J., Becker, D. & Schwarzwalder, J. (1977). In *Pyridine Nucleotide Dependent Dehydrogenases* (Sund, H., ed.), pp. 31–42, Walder de Gruyter, Berlin.

Weyer, W. J., Hofsteenge, J., Vereijken, J. M., Jekel, P. K. & Beintema, J. J. (1982). *Biochim. Biophys. Acta*, **704**, 385–388.

Wierenga, R. K. & Hol, W. G. J. (1983). *Nature (London)*, **302**, 842–844.

Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1979). *J. Mol. Biol.* **131**, 55–73.

Wierenga, R. K., Drenth, J. & Schulz, G. E. (1983). *J. Mol. Biol.* **167**, 725–739.

Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. (1985). *Biochemistry*, **24**, 1346–1357.

Wootton, J. C. (1974). *Nature (London)*, **252**, 542–546.

Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A. & Shaw, D. C. (1981). *Eur. J. Biochem.* **116**, 165–170.

Edited by A. Klug