Classification of doubly wound nucleotide binding topologies using automated loop searches

MARK B. SWINDELLS

Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

(RECEIVED July 14, 1993; ACCEPTED September 17, 1993)

Abstract

A classification is presented of doubly wound α/β nucleotide binding topologies, whose binding sites are located in the cleft formed by a topological switch point. In particular, the switch point loop nearest the N-terminus is used to identify specific structural classes of binding protein. This yields seven structurally distinct loop conformations, which are subsequently used as motifs for scanning the Protein Data Bank. The searches, which are effective at identifying functional relationships within a large database of structures, reveal a remarkable and previously unnoticed similarity between the coenzyme binding sites of flavodoxin and tryptophan synthetase, even though there is no sequence or topological similarity between them.

Keywords: analysis; classification; database searching; doubly wound α/β structures; loop motifs; nucleotide binding; topology

It is now well known that proteins with no detectable sequence similarity frequently adopt similar structures, and as a result their classification into superfamilies is becoming increasingly important. This paper classifies doubly wound α/β nucleotide binding domains, using specific structural features of their binding sites to achieve this goal.

Although extensive personal knowledge has previously formed the basis of classification (Baker et al., 1992b; Schulz, 1992), the number of structures now available renders this approach unsuitable for future research. Consequently, many researchers have developed automated procedures for identifying structural similarities in proteins, in particular global similarities (Taylor & Orengo, 1989; Holm & Sander, 1993). However, these algorithms do not assist a comprehensive classification of nucleotide binding proteins due to the unfortunate anomaly that proteins with no obvious functional relationship (e.g., flavodoxin and CheY) frequently exhibit more extensive global

similarities than those with related functions (e.g., lactate dehydrogenase and glutathione reductase). To resolve this dilemma I have developed a new procedure, which searches for specific invariant features of nucleotide binding sites, based on the following observations.

In general, the active sites of α/β topologies are always located in a cleft formed by a topological switch point involving two β -strands (Brändén, 1980). Furthermore, in nucleotide binding proteins, the loop region following the first of these strands (i.e., the one nearest the N-terminus) is usually involved with nucleoside phosphate binding. As a result, this region may be used as a reference point throughout the classification, with the loop conformations being used to identify specific families of phosphate binding motifs.

Nucleotide binding proteins have already been the subject of extensive research activity, and the developments described in this paper owe their existence to these publications. In particular, previous research has identified the two families of proteins described below.

Classic nucleotide binding domains are observed in numerous multidomain enzymes. Generally these domains are limited to dinucleotide binding, but surprisingly the recent determination of trimethylamine dehydrogenase revealed ADP bound in an analogous manner to FAD in glutathione reductase (Lim et al., 1988). Such domains have two recurrent features: a $\beta\alpha\beta\alpha\beta$ topology (frequently referred to as a Rossmann fold [Rao & Rossmann, 1973;

Reprint requests to: Mark B. Swindells, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

Abbreviations: PDB, Protein Data Bank; RMS, root mean square; NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate; FAD, flavin adenine dinucleotide; AMP, adenosine monophosphate; ADP, adenosine diphosphate; ATP, adenosine triphosphate; GMP, guanosine monophosphate; GTP, guanosine triphosphate; FMN, flavin mononucleotide. Abbreviations for proteins studied are given in the Methods section. The single-letter amino acid code is used, where x describes any amino acid.

Rossmann et al., 1974]; Fig. 1) and a short, structurally conserved sequence motif (GxGxxG) located at the first loop/helix junction (Rossmann et al., 1974; Wierenga & Hol, 1983; Wierenga et al., 1986). This loop region is involved with binding the nucleoside phosphate (Wierenga et al., 1985). Although it was initially believed that these features would be strictly conserved in all nucleotide binding domains, this is not so: the $\beta\alpha\beta\alpha\beta$ topology is absent in glutamate dehydrogenase (Baker et al., 1992a) and the latter two glycines of the GxGxxG motif are frequently mutated to alanine (Adams et al., 1991). It should be noted here that although flavodoxin (Smith et al., 1977) also adopts a Rossmann type fold (Rao & Rossmann, 1973), it binds mononucleotide through a significantly different loop conformation and cannot therefore be considered a true member of this family.

The second family of proteins is known as Type A nucleotide binding domains and includes the adenylate kinase, *ras* p21, and *recA* structures. Within this family, the observed topologies are extremely variable, and even include the regions defined by the nucleotide binding loops, where +2, +2x, and +3x connectivities are observed (Story & Steitz, 1992). Despite these variations, the conformation of the phosphate binding region remains strictly conserved and retains a GxxxxGKT sequence fingerprint (Walker et al., 1982). Unlike the classic nucleotide binding motif, this is located almost entirely within the loop.

It should be borne in mind that while GxGxxG and GxxxxGKT sequence motifs recur, they are insufficient to infer structural properties from sequence alone. One particularly famous application led to conflicting predictions of the *ras* p21 oncogene product (Gay & Walker, 1983; Wierenga & Hol, 1983), whose sequence —GAGGVGKS—

Fig. 1. Ribbon diagram of lactate dehydrogenase, indicating the classic nucleotide binding topology. The two $\beta\alpha\beta\alpha\beta$ motifs, which are typically referred to as Rossmann folds, are emphasized by differential shading.

was compatible with both motifs. Other errors, such as the misprediction of the cyclic AMP-dependent protein kinase structure (Sternberg & Taylor, 1984), have also occurred due to an overdependence on these data. In the following work it will be shown that even these simple sequence motifs are far more variable than previously expected. In particular, the classic GxGxxG sequence is highly changeable in all positions except for the first glycine.

Results

The central thesis of this work is that proteins can be segregated into families using the conformation of conserved nucleotide binding loops rather than the overall topology. However, the large increase in structures available for analysis renders a manual search unsuitable for comprehensive classification. As a consequence, I have adopted a procedure that automatically identifies strictly conserved loop conformations within a structural database. The method, which uses a simple least square superposition procedure, derives its power from the observation that the functional loops investigated here only superimpose well with regions from other proteins when there is a concomitant functional relationship (whether this observation can be more generally applied is the subject of ongoing work). The method was tested by first searching a database of low-homology structures (Orengo et al., 1993) for recurrence of the classic and Type A nucleotide binding motifs. In both cases, a 12-residue probe containing the loop motif was able to distinguish between functionally related loops and those with no relationship, as the former all superimposed with RMS deviations of less than 1.0 Å (Fig. 2).

As this procedure proved effective in detecting all relevant functional relationships in a limited test, the same procedure was subsequently implemented using a variety of search probes on all structures in release 64 of the Protein Data Bank. The results of these searches are shown in Figures 3–5, and the following section highlights specific similarities that were detected by this procedure.

Figure 3 shows all structures that bind nucleoside phosphate using a loop that is structurally analogous to LDH. This rigorous classification reveals that the second two glycine positions are even more variable than previously observed (Adams et al., 1991), with both accepting serine as well as alanine substitutions. Thus, the description of this sequence motif as GxGxxG is misleading because it describes a preference for glycine residues rather than an absolute requirement. Only the first glycine remains completely conserved, and this can be easily explained through an analysis of the ϕ, ψ plot, where it is the only residue to reside in a region of “glycine only” conformational space (in this case + ϕ , + ψ).

In previous analyses, the variant loop conformation of MDH, which contains an extra residue, has puzzled re-

Fig. 2. Results of searches using loop motifs on a low-homology data set. **A:** Overall distribution of C α superpositions made with residues 24–35 of lactate dehydrogenase. **B:** Details of lactate dehydrogenase superpositions with RMS deviations less than 2.0 Å. **C:** Overall distribution of C α superpositions made with residues 64–75 of recA. **D:** Details of the recA superpositions with RMS deviations less than 2.0 Å.

searchers, first because it appeared to be the only example of a (generalized) GxxGxxG loop motif, and second because the topology of MDH is so similar to LDH. However, a rigorous search for compatible loop motifs identified UDP-galactose 4-epimerase (UDP; Bauer et al., 1992) as a second member of this family. As a result, the MDH loop can now be considered the prototype of a distinct family of nucleotide binding loops. It is worth noting that liver dihydropteridine reductase (DHR) will probably also be a member of this loop family, as recent comparisons have revealed detailed structural similarities between UDP and DHR (L. Holm & C. Sander, pers. comm.).

The conformation of the FNR loop is radically different from that of the previous families, with the first glycine occupying positive ϕ , negative ψ and the remaining residues occupying right-handed alpha-helix. The data-

base search identifies phthalate dioxygenase reductase (PDR) as another member of this family, which is in line with previous observations (Correll et al., 1992).

Type A mononucleotide binding motifs (Walker et al., 1982) are summarized in Figure 4. Three different topological connectivities have been observed between the strands enclosing the nucleotide binding loop, which makes an evolutionary relationship difficult to establish. In particular, the 3x connectivity in REC is incompatible with earlier proposals of a common core (Milner-White et al., 1991). Type A loop conformations remain strictly conserved. (Note that the refined coordinates of ETU used here differ significantly from the unrefined set 1etu in PDB.) Analyses of the ϕ , ψ map suggest that both glycines in the GxxxxGKT motif are required for this loop conformation, as both occupy restricted regions of conformational space. A third site (x) features residues with high

Fig. 3. Data for classic nucleotide binding, malate dehydrogenase variant, and ferredoxin NADP reductase prototype loop motifs. Topology diagrams are presented for selected structures belonging to each group. N-terminal strands are shaded and marked appropriately. Every loop motif detected is listed together with its bound nucleotide, and the location of each motif relative to the subsequent loop/helix junction is represented by a line/cylinder diagram. A ϕ , ψ plot of each motif is presented. Because the classic nucleotide/MDH variant family is large, LDH and MDH loops are chosen as representative structures.

propensities for the γ_L conformation, namely glycine and aspartate.

Three other doubly wound α/β nucleotide binding families are shown in Figure 5. All three have 1x connectivities between their first strands and at least one site at which glycine adopts a restricted ϕ , ψ angle. These variations contribute to the unique conformations observed. Although database searches using the loop motifs of PGK and PFK did not reveal any other similarities, the FXN motif identified two loops in the β subunit of tryptophan synthetase (WSY) (Fig. 6). Both loops of WSY are located in the active site; one is directly involved with pyridoxal phosphate binding, whereas the other reflects a twofold repeat in this subunit. Structural alignments of the coenzyme binding loops from FXN and WSY reinforce this similarity, with the phosphate moieties overlapping almost perfectly (Kinemage 1). The phosphate oxygens of each cofactor form similar patterns of hydrogen bonds with loop main-chain nitrogens; two pairs of structurally aligned main-chain nitrogens (FXN-12, WSY-B236; FXN-8,

WSY-B232) locate within 2.9 Å of cofactor oxygens, whereas the final pair (FXN-11, WSY-B235) is slightly more distant but still within 3.2 Å of the remaining cofactor oxygen (Fig. 7; Kinemage 1). This compares favorably with previous analyses of classic nucleotide binding proteins in which hydrogen bonds were found to be limited and frequently longer than 3.2 Å (Wierenga et al., 1985). Although it was initially surprising to discover that it was the coenzyme binding loop that had a slightly higher RMS deviation, analyses of the observed binding mode revealed that Lys-87 of WSY is also required to form a Schiff base with the aldehyde group of pyridoxal phosphate. As Lys-87 has no structural equivalence within the twofold repeat, coenzyme binding by the other loop is inappropriate.

Discussion

In this paper I have analyzed a variety of nucleotide binding proteins, many of which retain similar activities while

Fig. 4. Type A nucleotide binding proteins. For each protein, the bound nucleotide is listed. ϕ , ψ plots of the sequence motifs reinforce the structural similarities within the loops. The layout of this figure is the same as for Figure 3.

adopting radically different topologies. The classification is based solely on the recurrence of analogous binding structures and is not intended to provide any insights into the possible evolutionary relationships between these proteins. This is a major difference between the current classification and that presented recently by Schulz (1992), in which the differences in global topological features have been discussed at length. With the exclusion of topology, the classification becomes much simpler and can be easily extended to include a variety of new structures and families.

Although recent publications have suggested that only a limited number of topologies are available (Chothia, 1992), it is still commonly believed that loop regions are extremely variable. Although in general this may be true, the work described here shows that functional loops remain highly conserved even across unrelated topologies. Given the recent identification of recurrent active site structures (Pearl, 1993), it may be possible to consider proteins in a modular manner in which recurrent topologies, loops, and side-chain orientations are recombined to tailor a specific function.

Methods

Coordinates for most proteins were taken from the Protein Data Bank (Bernstein et al., 1977). The following list of proteins studied is written in the manner: protein name (abbreviation used in paper), PDB code (reference).

Lactate dehydrogenase (LDH), 6ldh (Abad-Zapatero et al., 1987); alcohol dehydrogenase (ADH), 8adh (Colanna-Cesari et al., 1986); malate dehydrogenase (MDH), 4mdh (Birktoft et al., 1989); glyceraldehyde-3-phosphate dehydrogenase (GAH), 1gd1 (Skarzynski et al., 1987); glutathione reductase (GRS), 3grs (Karplus & Schulz, 1987); hydroxybenzoate hydrolase (PHH), 1phh (Schreuder et al., 1989); ferredoxin NADP reductase (FNR), 1fnr (Karplus et al., 1991); phthalate dioxxygenase reductase (PDR), 2pia (Correll et al., 1992); adenylate kinase (ADK), 3adk (Dreusicke et al., 1988); guanylate kinase (GKY), 1gky (Stehle & Schulz, 1990); *ras* p21 (P21), 5p21 (Pai et al., 1989); *recA* (REC), 2reb (Story & Steitz, 1992); phosphofructokinase (PFK), 1pfk (Skirakihara & Evans, 1988); flavodoxin (FXN),

Mononucleotide Binding Proteins			
1x Topology	Sequence motif	Nucleotide	phi/psi plot for sequence motif
PGK: ¹⁹⁴ GGAKVKDK	ATP

PFK: ¹⁰ GGDAPGMN	ATP

FXN: ⁸ GTGNTKEM	FMN

Fig. 5. Other nucleotide binding proteins. All three proteins clearly adopt radically different loop conformations. The layout of this figure is similar to Figure 3.

Flavodoxin: ⁶WSGTGNTKMAE

Tryptophan: ¹⁰⁹ETGAGQHGVASA 0.6 Å

Synthetase: ²³⁰CVGGGSNAIGMF 0.9 Å

Fig. 6. Details of the database searches using residues 6–17 of FXN, which identify two motifs from the β -subunit of tryptophan synthetase. The loop starting at residue 230 binds pyridoxal phosphate.

4fxn (Smith et al., 1977); thioredoxin reductase (TRB), 1trb (Kuriyan et al., 1991b); cholesterol reductase (COX), 1cox (Vrielink et al., 1991); 6-phosphogluconate dehydrogenase (PGD), 1pgd (Adams et al., 1991); tryptophan synthetase (WSY), 1wsy (Hyde et al., 1988); UDP-galactose 4-epimerase (UDP), 1udp (Bauer et al., 1992); trimethylamine dehydrogenase (TMD), 1tmd (Lim et al., 1988); trypanothione reductase (TPR), 2tpr (Kuriyan et al., 1991a); NADH peroxidase (NPX), 1npx (Stehle et al., 1991).

The following coordinates were received directly from the authors: nitrogenase (NIT; Georgiadis et al., 1992) from D.C. Rees, *Bacillus stearothermophilus* phosphoglycerate kinase (PGK; Davies et al., 1993) from H.C. Watson, and elongation factor-TU (ETU; Kjeldgaard & Nyborg, 1992) from Jens Nyborg.

Secondary structure assignments were made using DSSP (Kabsch & Sander, 1983), and diagrams generated from atomic coordinates were produced using MOLSCRIPT (Kraulis, 1991).

Searches for analogous loop structures were implemented using the structural alignment program SARF

Fig. 7. Superposition showing residues 8–13 from flavodoxin (black) and residues 232–237 from the tryptophan synthetase β -subunit (red). Flavin mononucleotide (black) and pyridoxal (red) phosphates are separated by <0.5 Å. Phosphate oxygens are stabilized by the main-chain nitrogens of residues 8, 11, and 12 in FXN and the structurally equivalent residues in WSY. See text for more details.

(Alexandrov et al., 1992), which is able to perform a wide range of structural comparisons based on alpha carbon superpositions. In this paper 12-residue fragments from the site of interest were compared with all fragments of similar length (no insertions or deletions allowed). Searches using different probe lengths were also tested, but their results were less clear. Comparisons were performed using the rigid body superposition procedure of SARF and applied to structures in release 64 of the Protein Data Bank. The RMS deviation of each comparison was recorded.

Descriptions of strand connectivities follow the procedures of Richardson (1981). In this manner, lactate dehydrogenase (Fig. 1) has strand connectivities $+1x, +1x, -3x, -1x, -1x$. This connectivity is also represented by a simplified topology diagram in Figure 2.

Acknowledgments

It is a pleasure to thank Nikolai Alexandrov for the use of the SARF program, Christine Orenko for low-homology data sets, Liisa Holm and Chris Sander for making their data available before publication, Yo Matsuo for helpful discussions, and the crystallographers for depositing their coordinates in the PDB.

References

- Abad-Zapatero, C., Griffith, J.P., Sussman, J.L., & Rossmann, M.G. (1987). Refined crystal structure of M4 apo-lactate dehydrogenase. *J. Mol. Biol.* 198, 445–467.
- Adams, M.J., Gover, S., Leaback, R., Phillips, C., & Somers, D.O. (1991). The structure of 6-phosphogluconate dehydrogenase refined at 2.5 Å resolution. *Acta Crystallogr. B* 47, 817–820.
- Alexandrov, N.N., Takahashi, K., & Go, N. (1992). Common spatial arrangements of backbone fragments in homologous and non-homologous proteins. *J. Mol. Biol.* 225, 5–9.
- Baker, P.J., Britton, K.L., Engel, P.C., Farrants, G.W., Lilley, K.S., Rice, D.W., & Stillman, T.J. (1992a). Subunit assembly and active site location in the structure of glutamate dehydrogenase. *Proteins Struct. Funct. Genet.* 12, 75–86.
- Baker, P.J., Britton, K.L., Rice, D.W., Rob, A., & Stillman, T.J. (1992b). Structural consequences for sequence patterns in the fingerprint region of the nucleotide binding fold. *J. Mol. Biol.* 228, 662–671.
- Bauer, A.J., Rayment, I., Frey, P.A., & Holden, H.M. (1992). The molecular structure of UDP-galactose 4-epimerase from *Escherichia coli* determined at 2.5 Å. *Proteins Struct. Funct. Genet.* 12, 372–381.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* 122, 535–542.
- Birktoft, J.J., Rhodes, G., & Banaszak, L.J. (1989). Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5 Å resolution. *Biochemistry* 28, 6065–6081.
- Brändén, C. (1980). Relation between structure and function of α/β proteins. *Q. Rev. Biophys.* 13, 317–338.
- Chothia, C. (1992). One thousand families for the molecular biologist. *Nature* 357, 543–544.
- Colonna-Cesari, F., Perahia, D., Karplus, M., Eklund, H., Brändén, C.I., & Tapia, O. (1986). Interdomain motion in liver alcohol dehydrogenase. Structural and energetic analysis of the hinge bending mode. *J. Biol. Chem.* 261, 15273–15280.
- Correll, C.C., Batie, C.J., Ballou, D.P., & Ludwig, M.L. (1992). Phthalate dioxygenase reductase: A modular structure for electron transfer from pyridine nucleotides to 2Fe-2S. *Science* 258, 1604–1610.
- Davies, G.J., Gamblin, S.J., Littlechild, J.A., & Watson, H. (1993). The structure of thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins Struct. Funct. Genet.* 15, 283–289.
- Deusicke, D., Karplus, P.A., & Schulz, G.E. (1988). Refined structure of porcine cytosolic adenylate kinase at 2.1 Å resolution. *J. Mol. Biol.* 199, 359–371.
- Gay, N.J., & Walker, J.E. (1983). Homology between human bladder carcinoma oncogene product and mitochondrial ATP-synthetase. *Nature* 301, 262–264.
- Georgiadis, M.M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J.J., & Rees, D.C. (1992). Crystallographic structure of nitrogenase iron protein from *Azotobacter vinelandii*. *Science* 257, 1653–1659.
- Holm, L., & Sander, C. (1993). Structural alignment of globins, phycoerythrins and colicin A. *FEBS Lett.* 315, 301–306.
- Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., & Davies, D.R. (1988). Three dimensional structure of the tryptophan synthetase multienzyme complex $\alpha_2\beta_2$ from *Salmonella typhimurium*. *J. Biol. Chem.* 263, 17857–17871.

- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure. *Biopolymers* 22, 2577–2637.
- Karplus, P.A., Daniels, M.J., & Herriott, J.R. (1991). Atomic structure of ferredoxin-NADP reductase: Prototype for a structurally novel flavoenzyme family. *Science* 251, 60–66.
- Karplus, P.A. & Schulz, G.E. (1987). Refined structure of glutathione reductase at 1.54 Å resolution. *J. Mol. Biol.* 195, 701–729.
- Kjeldgaard, M. & Nyborg, J. (1992). Refined structure of elongation factor EF-Tu from *Escherichia coli*. *J. Mol. Biol.* 223, 721–742.
- Kraulis, P.J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.
- Kuriyan, J., Kong, X.-P., Krishna, T.S.R., Sweet, R.M., Murgolo, N.J., Field, H., Cerami, A., & Henderson, G.B. (1991a). X-ray structure of trypanothione reductase from *Crithidia fasciculata* at 2.4 Å resolution. *Proc. Natl. Acad. Sci. USA* 88, 8764–8768.
- Kuriyan, J., Krishna, T.S.R., Wong, L., Guenther, B., Pahler, A., Williams, C.H., & Model, P. (1991b). Convergent evolution of similar function in two structurally divergent enzymes. *Nature* 352, 172–174.
- Lim, L.W., Mathews, F.S., & Steenkamp, D.J. (1988). Identification of ATP in the iron-sulfur flavoprotein trimethylamine dehydrogenase. *J. Biol. Chem.* 263, 3075–3078.
- Milner-White, E.J., Coggins, J.R., & Anton, I.A. (1991). Evidence for an ancestral core in nucleotide binding proteins with a type A motif. *J. Mol. Biol.* 221, 751–754.
- Orengo, C.A., Flores, T.P., Taylor, W.R., & Thornton, J.M. (1993). Identification and classification of protein fold families. *Protein Eng.* 6, 485–500.
- Pai, E., Kabsch, W., Krengel, U., Holmes, K., Jacob, J., & Wittin-ghofer, A. (1989). Structure of the guanine nucleotide binding domain of Ha-ras oncogene product p21 in the triphosphate conformation. *Nature* 341, 209–214.
- Pearl, L. (1993). Similarity of active site structures. *Nature* 362, 24.
- Rao, S.T. & Rossmann, M.G. (1973). Comparison of supersecondary structures in proteins. *J. Mol. Biol.* 76, 241–256.
- Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34, 167–339.
- Rossmann, M.G., Moras, D., & Olsen, K. (1974). Chemical and biological evolution of nucleotide binding proteins. *Nature* 250, 194–199.
- Schreuder, H.A., Prick, P.A.J., Wierenga, R.K., Vriend, G., Wilson, K.S., Hol, W.G.J., & Drenth, J. (1989). Crystal structure of the *p*-hydroxybenzoate hydrolase substrate complex refined at 1.9 Å resolution. *J. Mol. Biol.* 208, 679–696.
- Schulz, G.E. (1992). Binding of nucleotides by proteins. *Curr. Opin. Struct. Biol.* 2, 61–67.
- Skarzynski, T., Moody, P.C.E., & Wonacott, A.J. (1987). Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* 193, 171–187.
- Skirakihara, Y. & Evans, P.R. (1988). Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. *J. Mol. Biol.* 204, 973–994.
- Smith, W.W., Burnett, R.M., Darling, G.D., & Ludwig, M.L. (1977). Structure of the semiquinone form of flavodoxin from *Clostridium MP*—Extension to 1.8 Å resolution and some comparisons with the oxidised state. *J. Mol. Biol.* 117, 195–225.
- Stehle, T., Ahmed, S.A., Claiborne, A., & Schulz, G.E. (1991). Structure of NADH peroxidase from *Streptococcus faecalis* 10C1 refined at 2.16 Å resolution. *J. Mol. Biol.* 221, 1325–1344.
- Stehle, T. & Schulz, G.E. (1990). Three dimensional structure of the complex of guanylate kinase from yeast with its substrate GMP. *J. Mol. Biol.* 211, 249–254.
- Sternberg, M.J.E. & Taylor, W.R. (1984). Modelling the ATP binding site of oncogene products, the epidermal growth factor receptor and related proteins. *FEBS Lett.* 175, 387–392.
- Story, R. & Steitz, T.A. (1992). Structure of the recA protein-ADP complex. *Nature* 355, 374–376.
- Taylor, W.R. & Orengo, C. (1989). Protein structure alignment. *J. Mol. Biol.* 208, 1–22.
- Vrielink, A., Lloyd, L.F., & Blow, D.M. (1991). Crystal structure of cholesterol oxidase from *Brevibacterium sterolicum* refined at 1.8 Å resolution. *J. Mol. Biol.* 219, 533–554.
- Walker, J.E., Saraste, M., Runswick, M., & Gay, N.J. (1982). Distantly related sequences in the a and b subunits of ATP synthase, myosin kinases, and other ATP requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951.
- Wierenga, R.K., De Maeyer, M.C.H., & Hol, W.G.J. (1985). Interaction of pyrophosphate moieties with alpha helices in dinucleotide binding proteins. *Biochemistry* 24, 1346–1357.
- Wierenga, R.K. & Hol, W.G.J. (1983). Predicted nucleotide binding properties of p21 protein and its cancer related variant. *Nature* 302, 842–844.
- Wierenga, R.K., Terpstra, P., & Hol, W.G.J. (1986). Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ fold in proteins using an amino acid sequence fingerprint. *J. Mol. Biol.* 187, 101–107.