Modelling the ATP-binding site of oncogene products, the epidermal growth factor receptor and related proteins

Michael J.E. Sternberg+ and William R. Taylor

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1R 7HX, England

Received 25 July 1984

A three-dimensional model for the ATP-binding site of the oncogene product v-src is proposed by the use of interactive computer graphics. A similar model would apply to sequence related proteins such as other oncogene products, the epidermal growth factor receptor, cAMP-dependent protein kinase and the cell division control protein CDC28. The model was proposed on the basis of the conservation of certain key residues between the oncogene product family of proteins and several nucleotide binding proteins of known structures.

Protein structure prediction *Computer graphics modelling* *Oncogene product* *Kinase activity* *ATP-binding site* *Nucleotide-binding protein*

1. INTRODUCTION

Questions about the relationship between normal and cancerous cell growth are raised by the similarity of amino acid sequence between several oncogene products, such as v-src and v-erb-B, and a variety of other proteins – the epidermal growth space factor receptor (EGF-R), mammalian cAMP-dependent protein kinase (cAMP-K) and a cell division control protein CDC28 [1–10]. Beyond just a sequence similarity many, but not all, of these oncogene products share a tyrosine phosphokinase activity with the EGF-R and a more general kinase activity with cAMP-K and CDC28. As a step towards understanding the relationship between similarity in primary structure and kinase activity, we present a three-dimensional model for the ATP-binding site of the oncogene product v-src. A similar model would be applicable to all the above proteins and is consistent with experimental evidence on the ATP binding site of cAMP-K [12,13]. The proposal for the model was based on the conservation of certain key residues between the oncogene products and several nucleotide binding proteins [15–20] of known sequence and three-dimensional structure (cf. [21]).

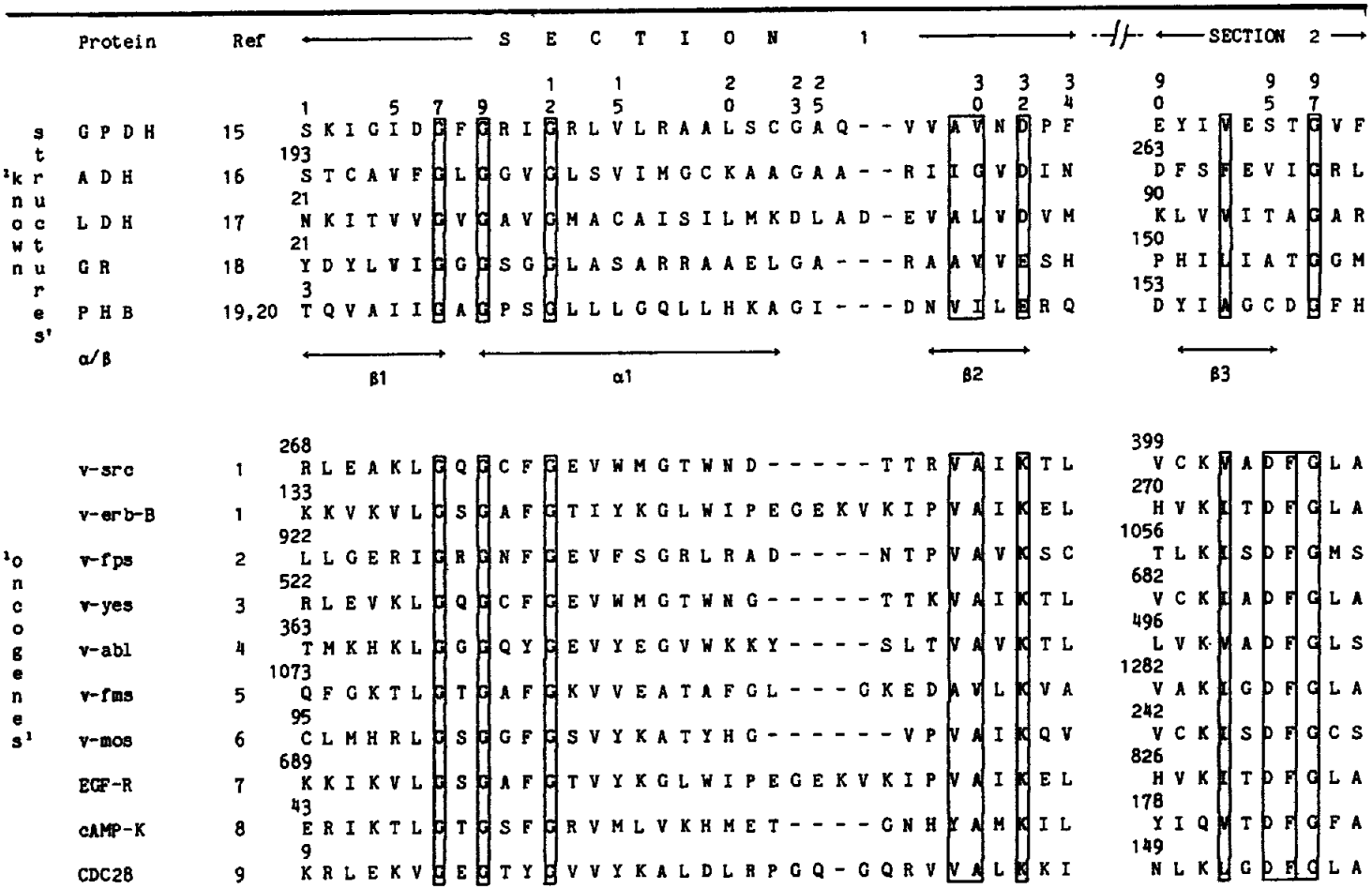
+ To whom correspondence should be addressed

2. MATERIALS AND METHODS

Table 1 gives the sequence alignment of five nucleotide binding regions from five proteins of known sequence and structure. This set will be referred to as the 'known structures' and consists of the NAD-binding regions of glyceraldehyde-3-phosphate dehydrogenase (GPDH), lactate dehydrogenase and alcohol dehydrogenase and the FAD-binding regions of glutathione reductase and *p*-hydroxybenzote hydroxylase. For reference the sequence numbering of GPDH will be used. This alignment and the observations about the role of conserved residues have been noted by other workers [32–33]. Each binding region comprises two sections. The first consists of a β-strand–α-helix–β-strand unit where the two strands (called 1 and 2) lie parallel and adjacent. The second section consists of a β-strand (called 3) and a few subsequent residues. β-Strand 3 lies parallel and adjacent to 1 to give the strand order 312. The major

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00 © 1984 Federation of European Biochemical Societies

Table 1
The sequence alignment of five known structures



α/β denotes the secondary structure assignment which is taken from GPDH but would apply in general to all five chains. The role of glycines at positions 7, 9 and 12 has been described by Wierenga and Hol [21]. Gly 7 is required for the binding of the ribose moiety, Gly 9 enables the close approach of the negatively-charged pyrophosphate group to the positively charged N-terminus of the α-helix dipole [27], both Gly 7 and Gly 12 enable the chain to make the sharp turn between β-strand 1 and the α-helix. The oncogene sequences are also given. The residues whose conservation are central to the arguments are enclosed in the boxes. In addition, note the similarity of sequence between the 'known structures' and the 'oncogenes': the preferences for hydrophobic residues at positions 15, 20 and 31 and for Gly at 25

feature of the first section is the presence of the sequence pattern Gly-X-Gly-X-X-Gly (Gly is G in table 1, X is any residue) in the region between the end of the β-strand 1 and the start of the connecting α-helix. These glycines are essential for the binding of the ribose moiety and for the formation of the tight turn between the β-strand and the α-helix (see legend to table 1). Another important feature is the presence of a conserved acidic side chain (Asp(D) or Glu(E)) at the carboxyl end of β-strand 2 (position 32). This side chain forms a hydrogen bond with the 2'-hydroxyl of the adenine ribose. In the second section there is a conserved Gly at the end of β-strand 3 whose structural and functional role is not clear.

Table 1 also gives a sequence alignment of two sections of several oncogene products, the EGF-R, cAMP-K and CDC28 and similar alignments have been published previously [1–11]. All these sequences will be referred to 'oncogene' sequences and their similarity implies that all the chains will adopt similar three-dimensional structures. We report how sections of the 'oncogene' sequences can be aligned with the 'known structures' and hence a three-dimensional model proposed. In the first section all the 'oncogene' sequences share a common Gly-X-Gly-X-X-Gly which we have aligned with three Gly of the 'known structures'. Some 17 to 23 residues beyond the third Gly there is an invariant Lys(K) in the 'oncogenes' which we have aligned with the conserved acidic residue of the 'known structures'. In this alignment positions 29 and 30 are always occupied by a hydrophobic side chain or by a neutral Gly. The second section of the 'oncogene' sequences includes the region before the tyrosine that in many of these proteins has been identified as the site of phosphorylation. Of interest is the presence of the conserved sequence Asp-Phe-Gly (DFG) whose total conservation is suggestive of a functional role. A secondary structure prediction for this region by the algorithm of Taylor and Thornton [23,24] shows that in most of the 'oncogene' sequences the residues before the Asp-Phe-Gly are predicted to adopt a β-strand conformation. Accordingly we have aligned the invariant Gly of the 'oncogenes' with the Gly 97 just after β-strand 3 of the 'known structures'. Although this is the more tentative part of the sequence alignment, this is the only region in which a persistent β prediction precedes an absolutely conserved Gly.

The sequence alignment of table 1 implies that the two sections of the 'oncogene' polypeptide chain will adopt a similar conformation to the corresponding region of the 'known structures'. Accordingly the crystallographic coordinates of GPDH [25] were used to model-build the two sections of the oncogene v-src. The model building was performed using interactive computer graphics [26] and involved altering the chemical type of the side chains followed by adjustment of side-chain torsion angles to obtain allowed stereochemistry. The coordinates of GPDH included the bound NAD whose chemical structure was altered to ATP and the mode of binding of ATP to v-src was investigated. Central to this study was the experimental evidence [13] from affinity reagents which suggests that in cAMP-K the Lys at aligned position 32 interacts with the phosphate section of ATP.

3. RESULTS AND DISCUSSION

The sequenced changes were made from GPDH to v-src and it was found that Lys 32 could not span the distance to interact with any oxygen atom on the phosphate section of ATP when the ATP adopted an analogous conformation to the bound NAD. In addition the change from an acidic group at position 32 of the 'known structures' to a Lys in the 'oncogenes' resulted in the loss of a stabilising hydrogen bond with the 2'-hydroxyl of the ribose. We concluded that the ATP/'oncogene' interaction would be different from that of the adenosine part of NAD with the 'known structures' and we modelled a possible ATP/'oncogene' interaction that would include functional roles for the conserved residues (see fig.1b and table 2). In this interaction the N^{ϵ} of Lys 32 forms a salt bridge with an oxygen on the β-phosphate of ATP. The conserved Gly at position 7 ensures the close approach of the N1 and C2 atoms of the adenine while a terminal oxygen of the conserved Asp 95 forms a hydrogen bond with the NH_2 group at position 6 of the adenine. The conserved Phe 99 is proposed to make hydrophobic interactions with the adenine ring. In the model the positively-charged amino terminus of the α-helix dipole [27] stabilises the γ-phosphate group of ATP and this interaction is made possible by the presence of the conserved Gly

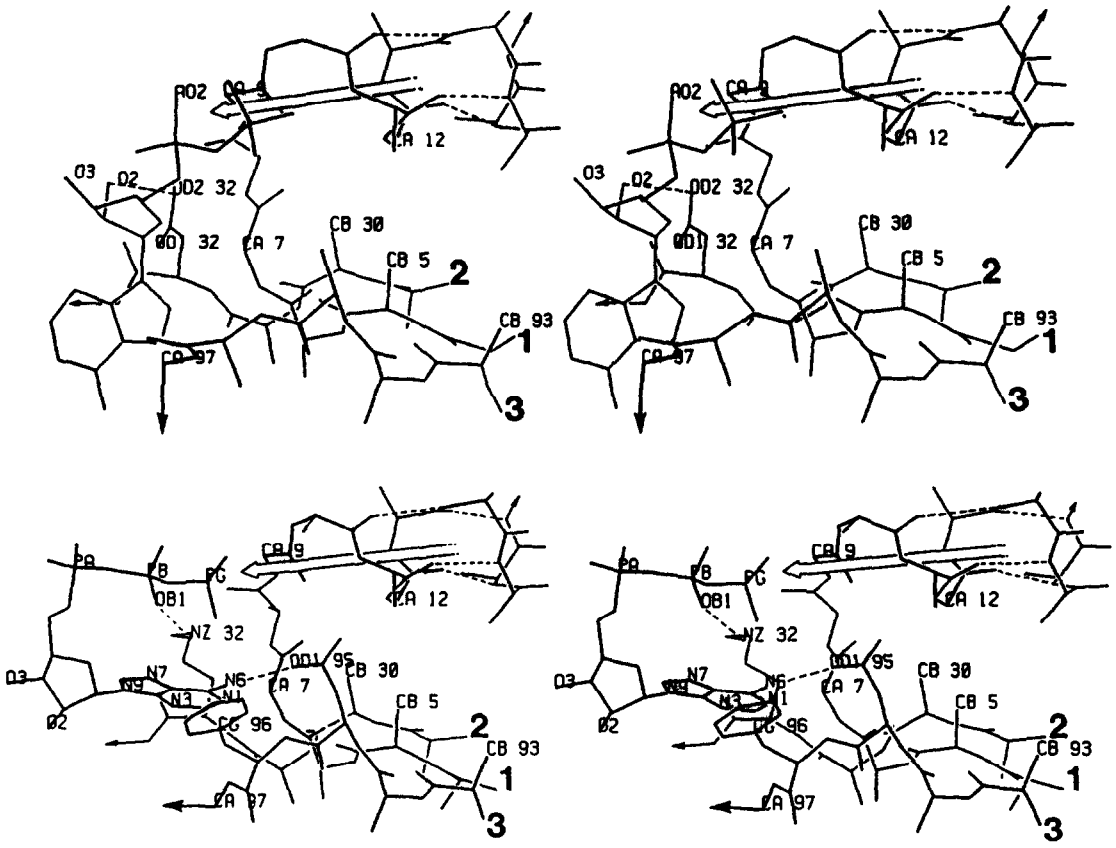


Fig.1. (A) A stereo diagram of the interaction of NAD and parts of GPDH. The polypeptide backbone and Cβ atoms are drawn and the side chain of Asp 32 is shown. The four conserved glycines at positions 7, 9, 12 and 97 are labelled together with residues 30, 5 and 93 which indicate the strand alignment. The role of the α-helix dipole in interacting with the pyrophosphate is denoted by an arrow. (B) A stereo diagram of the proposed ATP-binding site of v-src drawn in a similar way to (A). The roles of Gly 7, Gly 9, Gly 12, Gly 97, Lys 32, Asp 95 and Phe 96 together with the α-helix dipole (an arrow) in binding ATP are shown. The model was constructed using the program FRODO which was written by Dr T.A. Jones [26] and modified for the Evans and Sutherland Picture System 2 and PDP11/60 by Dr I.J. Tickle.

9 which facilitates the close approach of the γ-phosphate group. In this model the tyrosine residue that is phosphorylated in many of the proteins is located about 9 residues from the Asp-Phe-Gly sequence of the second section. Thus if the mechanism involves intrachain autophosphorylation, the tyrosine could be spatially close to the γ-phosphate of the ATP.

After this model building study was completed, our attention was drawn to experimental work on the binding of ATP analogous to cAMP-K [14]. Our model for the ATP/protein interaction is consistent with these experimental results. Substitution of the NH_2 at the 6 position of the adenine by a larger group decreased the affinity of the binding probably due to steric effects and this was confirmed in the proposed model. Replacement of the NH_2 at position 6 by oxygen lead to a decrease of affinity due to electronic effects which is consistent with the proposed hydrogen bond of the adenine NH_2 with Asp 95. Other studies suggested that the glycosidic bond is in the *anti* conformation rather than the *syn* arrangement and in the model the torsion angle is 81°, which is the high *anti* conformation. Replacement of either 2′- or the 3′-hydroxyl by a hydrogen was not deleterious for binding; this

Table 2

Interactions between v-src and ATP

Protein atom	ATP atom	Distance (Å)	Interaction
Asp 95 OD1	N6 on adenine	2.8	hydrogen bond
Gly 7 CA	N1 on adenine	3.1	van der Waals
Gly 7 CA	C2 on adenine	3.0	van der Waals
Phe 96 CD	C5 on adenine	3.9	hydrophobic
Lys 32 NZ	O on β phosphate	3.1	salt bridge

is consistent with the model in which the ribose hydroxyls are exposed to solvent and do not interact with the protein. The removal of the γ-phosphate only moderately affected the binding while removal of both the β- and the γ-phosphates markedly affected the binding. This is consistent with the proposed salt bridge between the β-phosphate and the Lys 32.

The proposed model for the ATP binding site of v-src would be applicable to all the 'oncogene' sequences in table 1 as it is based on the conservation of residues. The model should prove helpful in the design of experiments to study the binding of ATP to these proteins and possibly for the design of inhibitors that might be useful clinical agents.

ACKNOWLEDGEMENTS

We thank Professor T.L. Blundell for directing our attention to this problem and Dr J.M. Thornton, Dr M. Waterfield and Mr J. Downward for helpful discussion. M.J.E.S. holds a Royal Society University Research Fellowship and W.R.T. is supported by the SERC.

REFERENCES

[1] Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. and Waterfield, M.D. (1984) Nature 307, 521–527.

[2] Shibuya, M. and Hanafusa, H. (1982) Cell 30, 787–795.

[3] Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. and Yoshida, M. (1982) Nature 297, 205–208.

[4] Reddy, E.P., Smith, M.J. and Srinivasan, A. (1983) Proc. Natl. Acad. Sci. USA 80, 3623–3626.

[5] Hampe, A., Gobet, M., Sherr, C.J. and Galibert, F. (1984) Proc. Natl. Acad. Sci. USA 81, 85–89.

[6] Van Beveren, C., Galleshaw, J.A., Jonas, V., Berns, A.J.M., Doolittle, R.F., Donoghue, D.J. and Verma, I.M. (1981) Nature 289, 258–262.

[7] Ullich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M. and Seeburg, P.H. (1984) Nature 309, 418–425.

[8] Shoji, S., Parmelee, D.C., Wade, R.D., Kumar, S., Ericss, L.H., Walsh, K.A., Neurath, H., Long, G.L., Demaille, J.G., Fischer, E. and Titani, K. (1981) Proc. Natl. Acad. Sci. USA 78, 848–851.

[9] Lorincz, A.T. and Reed, S.I. (1984) Nature 307, 182–185.

[10] Privalsky, M.L., Ralston, R. and Bishop, J.M. (1984) Proc. Natl. Acad. Sci. USA 81, 704–707.

[11] Barker, W.C. and Dayhoff, M.O. (1982) Proc. Natl. Acad. Sci. USA 79, 2836–2839.

[12] Newmark, P. (1984) Nature 307, 499.

[13] Zoller, M.J. and Taylor, S.S. (1979) J. Biol. Chem. 254, 8363–8367.

[14] Flockhart, D.A., Freist, W., Hoppe, J., Lincoln, T.M. and Corbin, J.D. (1984) Eur. J. Biochem. 140, 289–295.

[15] Buehner, M., Ford, G.C., Moras, D., Olsen, K.W. and Rossmann, M.G. (1974) J. Mol. Biol. 90, 25–49.

[16] J. Mol. Biol. (1976) 102, 27–59.

[17] Holbrook, J.J., Liljas, A., Steindel, S.J. and Rossmann, M.G. (1975) in: The Enzymes, vol.11A, 191–292.

[18] Thieme, R., Pai, E.F., Schirmer, R.H. and Schulz, G.E. (1981) J. Mol. Biol. 152, 763–782.

[19] Wierenga, R.K., Jong, R.G. de, Kalk, K.H., Hol, W.G.J. and Drenth, J. (1979) J. Mol. Biol. 131, 55–73.

[20] Weijer, W.J., Hofsteenge, J., Vereijken, J.M., Jekel, P.A. and Beinfema, J.J. (1982) Biochim. Biophys. Acta 704, 385–388.

[21] Wierenga, R.K. and Hol, W.G.J. (1983) Nature 302, 842–844.
[22] Rossmann, M.G., Liljas, A., Branden, C.-I. and Banaszak, L.J. (1975) in: The Enzymes, vol.11A, 61–102.
[23] Taylor, W.R. and Thornton, J.M. (1983) Nature 301, 540–542.
[24] Taylor, W.R. and Thornton, J.M. (1984) J. Mol. Biol. 173, 487–514.
[25] Bernstein, F.C., Koetzle, T., William, G.J.B., Meyer, E. jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[26] Jones, T.A. (1978) J. Appl. Crystallogr. 11, 268–272.
[27] Hol, W.G.J., Van Duijnen, P.Th. and Berendsen, H.J.C. (1978) Nature 273, 443–446.