STRUCTURAL MOTIF(S) AT THE ACTIVE SITE OF NUCLEOTIDE-BINDING PROTEINS

P. NARAYANAN

Department of Life Sciences, University of Bombay, Vidyanagari, Bombay 400 098, India.

ABSTRACT

The amino acid sequence of the β -subunit of ATP-synthase in the nucleotide-binding region is highly homologous to the amino acid sequence of N-terminal region of structurally-related, 21,000 dalton ras gene proteins, p21. The sequence homology extends to other proteins like yeast gene encoding protein YP2 and other mononucleotide-binding proteins that use GTP (or ATP) in catalysis, like α -transducin, suggesting a common structural motif, $\beta\alpha\beta$ at the nucleotide-binding region of these proteins. Depending on the topology of this motif, a single residue at the β - α junction or a patch consisting of more residues at this region would be important for normal protein function.

INTRODUCTION

The active site of the mononucleotide-binding protein, ATP-synthase is found in the β -subunit of the F1-part. The amino acid sequence of the nucleotide-binding region in β -subunit of F1-part is highly homologous with the amino acid sequence of N-terminal region of structurally-related ras gene proteins, p21¹. Each functional member of the ras gene family, virus and human is split into four exons. The first exon encodes the N-terminal 37 residues of a 21,000 dalton protein, p21. Normal cellular ras gene protein p21 has glycine at position 12. Single amino acid change in the first exon at position 12 (single-point mutation) activates the oncogene²⁻¹⁰.

p21 Protein displays a strong affinity for GTP and GDP and by specifically binding to GTP and GDP¹¹. p21 may resemble G-nucleotide-binding regulator proteins, such as the elongation factors of protein synthesis (EF-Tu & EF-G)¹². The α-subunits of transducin¹³, GTPase of bovine rod outer segments^{14**} and of G-proteins^{15,16} also share homology with ras gene products. These homologous segments correspond mostly to the regions involved in the guanine nucleotide-binding and ADP-ribosylation sites of these proteins.

Comparison of the sequence of 37 N-terminal residues (first exon) of the normal p21 protein with sequences of the dinucleotide-binding proteins with $\beta\alpha\beta$ structural motif and with other proteins that are known to bind nucleotides was carried out ¹⁷⁻²⁰.

Wierenga and Hol¹⁷ suggested structural role to the glycine residue at position 12 (II glycine) in transforming properties, by visual inspection of the sequences and model building. Möller and Amons²⁰ suggested glycine at position 15 (III glycine) as the most important, because of G15 being the most invariant in most of the nucleotide-binding proteins, rather than G12 for (pyro)phosphate binding.

Due to the importance of these studies, and to establish if there exists a common structural motif at the nucleotide-binding regions of the nucleotide-binding proteins, a more systematic analysis of amino acid sequences has been carried out to identify similarities between sequences of proteins known to bind nucleotides and those of suspected nucleotide-binding property.

METHODOLOGY

The residues (23 in number) corresponding to the β - α unit of porcine lactate dehydrogenase (LDH)^{21,22}, equine liveralcohol dehydrogenase (LADH)²³, lobster glyceraldehydephosphate dehydrogenase(GPD)²⁴, glutathione reductase(GR)²⁵ and p-hydroxybenzoate hydroxylase(PHBH)26 were chosen as probe sequences to locate potential nucleotide-binding region in other proteins. FADand NAD-binding proteins have $\beta\alpha\beta$ moiety (see for example Wierenga et al18.27). The probe in the present analysis was limited to 23 residues corresponding to the known β - α moiety of this region of sequences since no deletion or insertion within the segment is observed. For each position of the probe, an average value of Chou-Fasman parameter for ahelix, β -strand and reverse turn potentials was computed28. The probe sequences were compared

^{**}The number of amino acids and their sequences in α -transducin and α -GTPase are the same and, hence, they are the same protein(s) reported by these two groups.

Name	Residues	5	10 12 15 20	25 27 Reference
LDH	22-44	K	TVVGVGAVGMACAI	SILMKDL 21, 22
LADH .	194-216	T	CAVFGLGGVGLSVIN	AGCKAAGA 23
GPD	2-24	K	GI DGF GRÌ GRLVLF	RAALSCGA 24
GR	22-44	D	YLVI GGGS GGLAS AF	RAAELGA 25
РНВН	4-26	Q	VAII GAGPS GLLLG(LLHKAGI 26
p21	5-27	K	LVVVGAGGVGKSALI	IQLIQNH 2,8
T-24	5-27	K	. V V V G A V G V G K S A L 7	IQLIQNH 2,7
ATPase	151-173	K	GLFGGAGVGKTVFI	MELINNV 30, 31
\ P2	10-32	K	LLI GNS GVGKS CLI	LRFSDDT 32
α-tubulin	138-160	F	HSFGGGTGSGFTSLI	MERLS VD 34
GTP-AMP Phos.	7-29	L	RAIMGAPGSGKGTVS	SRITKHF 40
Prot kinase	44-66	K	KTLGTGSFGRVMLV	KHMETGN 39
Adenylatekinase	9-32	KI	FVVGGPGSGKGTQC	CEKIVQKY 33
ATP-AMP tr. phos.	9-32	ΚI	F V V G G P G S G K G T Q (CEKIVHKY 41
EF-Tu	13-35	N	VGTI GHVDHGKTTL 1	CAAITTVL 35,36
Fe-nitrogenase	4-26	Q	CAIYGKGGIGKSTT	
GTPase (α)	31-53	K	LLLGAGESGKSTI	-
G-protein	35-57	K	LLLGAGESGKSTI	,

Table 1 Sequence homology of (\$\beta\aregion) of nucleotide-binding proteins

GTP-AMP Phos.: GTP-AMP Phosphotransferase Prot. Kinase: c-AMP dependent protein kinase ATP-AMP tr. phos.: ATP-AMP transphosphorylase

Numbering at the top corresponds to that of p21 protein; single-alphabet nomenclature of peptides is followed; A(Ala); C(Cys); D(Asp); E(Glu); F(Phe), G(Gly); H(His); I(Ile); K(Lys); L(Leu), M(Met); N(Asn); P(Pro);

Q(Gln); R(Arg); S(Ser); T(Thr); V(Val), W(Trp); Y(Tyr).

with every possible 23 residues segments of sequences of suspected nucleotide-binding property. The alignment of the sequences of dinucleotide-binding probes with possible nucleotide-binding proteins, based on the sequence and structure, culled from the literature, is shown in table 1.

For each comparison, the following quantities were computed: (i) Correlation coefficient between Chou-Fasman parameters of the chosen segment and the probe. This is defined as

where f_{si} is a Chou-Fasman parameter at the *i*th position in the selected segment and $\langle f_s \rangle$ is the average value of this parameter for the whole segment. Similarly p stands for probe. (ii) The number of residues of the segment which are identical to any one of the residues found at the

corresponding position of probe sequences. (iii) The sum of the scores²⁹ obtained by comparing each residue of the segment to a residue chosen from the corresponding position of the probe so as to maximize the score obtained.

An average value of correlation coefficient and a root-mean-square deviation from this value was obtained using correlation coefficients obtained for all comparisons involving the selected sequence. From (ii) and (iii) a cumulative distribution of identities and scores was obtained. The mean correlation coefficient, as anticipated, was always close to zero and the root-mean-square deviation was ~ 0.04. Local similarities are revealed if (i) correlation coefficient is significant in comparison with its fluctuation; (ii) the scores are significantly higher than the anticipated scores on the basis of cumulative distribution. These tests were chosen to locate segments with a high degree of similarities with the probe sequences. Also, in each case, the segment of the sequences that shows best agreement in terms of each of the parameters was selected. It might be anticipated that each of these selections will point to the same segment if the similarity is highly significant. The three tests measure rather

Protein	Residues		Std. dev	Segment with highest c.c	Correla- tion of c.c	Seg. with highest McLachlan Score	McLa- chlan Score	Seg. with largest identical residues	No. of identical	References
p21	5-27	0.0	0.17	5-27	0.69	5–27	48	5-27	13	2, 8
ATPase	151-173	0.0	0.17	151-173	0.48	151-173	50	151-173	14	30, 31
YP2	10-32	0.0	0.18	10-32	0.58	10-32	48	10-32	14	32
α-tubulin	138-160	0.0	0.17	138-160	0.52	375-397	44	139-161	12	34
GTP-AMP Phos.	7-29	0.0	0.17	7-29	0.52	7-29	43	127-129	11	40
Adenylatekinase ATP-AMP	9–32	0.0	0.18	112-134	0.54	112-134	43	112-134	11	33~
tr. phos.	9-32	0.0	0.18	9-32	0.52	170-192	42	6-28	11	41
EF-Tu Fe-nitrogenase	13-35	0.0	0.18	101–123	0.51	370-392	45	370–392	13	35, 36
(A. Vinelandii) α-GTPase	4–26	0.0	0.19	4–26	0.57	4–26	48	4-26	14	37, 38
(a-transducin)	31-53	0.0	0.17	31-53	0.51	31-53	46	65-87	12	13, 14
G-protein	35-57	0.0	0.16	35-57	0.51	35-57	46	71-93	11	15, 16

Table 2 Correlation parameters for nucleotide-binding proteins with β-α moiety

different aspects of protein sequence. Chou-Fasman parameters measure the propensity of the sequence to assume a particular structure while the sequence comparisons measure the possible evolutionary convergence or divergence to a particular suitable sequence. The results obtained are given in table 2.

DISCUSSION

From the sequence comparison and model building studies of $\beta\alpha\beta$ region of the dinucleotide-binding proteins, Wierenga et al^{17,19} suggested a set of "thumb rules" to test whether a particular sequence will have $\beta\alpha\beta$ motif. These rules can be summerized as:

- (i) The presence of hydrophilic residue (usually lysine) at the beginning of the I β -strand. (ii) The sequence G-X-G-X-X-G in the region adjoining the I β -strand and the N-terminus of the α -helix.
- (iii) The occurrence of several neutral, often hydrophobic residues at these positions, which are involved in forming the hydrophobic core of the $\beta\alpha\beta$ -unit.
- (iv) The presence of a negatively charged residue (Asp or Glu) at the C-terminus of the II β -strand.

According to Wirenga et al a "core finger-print" consists of three glycine residues and an acid residue (D or E). Wirenga and Hol suggested a structural role to the position 12, that glycine at position 12 (II glycine) is essential for the proper function of protein p21. But, Möller and Amons²⁰ contended that glycine at position 15 (III glycine) rather than

the residue at position 12 is more important for (pyro)phosphate binding and hence substitution at the equivalent position 15 is expected to have a greater influence on GDP binding in mononucleotide-binding proteins. They suggested the sequence-G(1st)-X-X(2nd)-X-X-G(3rd)-K to occur in mononucleotide-binding proteins and G-X-G-X-X-G in dinucleotide-binding proteins (table 3).

Though β -subunit of F1-ATPase^{30,31}, yeast gene encoding protein YP2³², adenylate kinase³³, α tubulin³⁴, elongation factors^{35,36}. Fe- and FeMonitrogenases^{37,38} and other nucleotide-binding proteins have similar homologies at the nucleotidebinding region, only normal p21 and other GTPbinding proteins like α-transducin and G-proteins adhere to the criteria as suggested by Wierenga et al for $\beta\alpha\beta$ motif. The first criterion, namely, hydrophilic residue (K) at the beginning of the I β -strand is adhered to by some nucleotide-binding proteins, like ATPase, YP2, G-proteins, but are not adhered to by other nucleotide-binding proteins like α-tubulin, Fe-nitrogenase^{37,38}, GTP-AMP phosphotransferase and others. There are also exceptions to the second criterion that glycine occurs at the equivalent position 12 (II glycine). Again, exceptions are found to the fourth criterion also, that is, to the occurrence of negatively charged residue at the C-terminus of the second β -strand (examples.) ATPase, YP2, EF-Tu, Fe-nitrogenase etc.).

Normal cellular function may not be possible for ras gene p21 protein when glycine at position 12 (II glycine) is replaced, but is certainly possible in spite

Table 3 Sequence homology at the equivalent position 12 for the $\beta\alpha\beta$ motif

		I 10		II 12	-		111 15		
Protein	Residues	G	х 	X 	X	X	G 	X'	References
LDH	27-33	G	V	G	A	V	G	M	21, 22
LADH	199-205	G	L	G	G	V	G	L	23
GPD	7–13	G	F	G	R	I	G	R	24
GR	27-33	G	G	G	S	G	G	L.	2 5
PHBH	9–15	G	Α	G	P	S	G	L	26
p21	10-16	G	Α	G	G	V	G	K	2, 8
V-ras-H		G	Α	R	G	V	G	K	4
V-ras-K		G	Α	S	G	V	G	K	10
V-ras-Ra		G	Α	R	G	V	G	K	6
T-24		G	Α	V	G	V	G	K	2, 7
Calu-1 lung									
carcinoma		G	Α	C	G	V	G	K	3
SW480- colon									
carcinoma		G	A	V	G	V	G	K	3
ATPase	156-162	G	G	Α	G	V	G	K	30, 31
YP2	15-21	G	N	S	G	V	G	K	32
α-tubulin	143-149	G	G	T	G	S	G	F	34
GTP-AMP Phos.	12-18	G	Α	P	G	S	G	K	40
Prot. Kinase	49-55	G	T	G	S	F	G	R	39
Adenylatekınase	15-21	G	G	P	G	S	G	K	33
ATP-AMP tr. phos.	15-21	G	G	P	G	S	G	K	41
EF-Tu	18-24	G	Н	V	D	H	G	K	35, 36
Fe-nitrogenase	9-15	G	K	G	G	I	G	K	37, 38
GTPase (a)	36-42	G	Α	G	E	S	G	K	13, 14
G-protein	40 –46	G	A	G	E	S	G	K	15, 16

Numbering at the top corresponds to that of p21 protein; single-alphabet nomenciature is followed. X & X' are for any peptide residue.

of the replacement of glycine at the equivalent position 12 (Arg, Pro, Ser or Glu in ATPase; and Ser in YP 2) without any change in protein function for large number of nucleotide-binding proteins. The function of ATPase, transducin, adenylate kinase and YP2 etc may be similar to that of p21, but may have altered specifications. Therefore, the binding region of the nucleotide-binding proteins may have the general $\beta\alpha\beta$ structural motif with different topologies.

The suggestion of Moller and Amons that mononucleotide-binding proteins having the sequence G-X-X-X-G-K may be in general true, but there are exceptions to this observation (α-tubulin³⁴, c-AMP dependent protein kinase⁴⁰). As pointed out by them, the sequence of EF-Tu³⁵ at residues 224-230 is ²²⁴G-X-X-X-X-G-R²³⁰, but the N-terminal region of the EF-Tu is the phosphate-binding region due to better homology with other mononucleotide-binding proteins and the consistency of the proposed location of the (pyro)phosphate group

near the N-terminal end. Another example of this kind worth noting is the sequence in α -tubulin. Even though the sequence of α -tubulin at residues 106-112 is ¹⁰⁶G-H-Y-R-I-G-K¹¹², in accordance with the suggested sequence G-X-X-X-X-G-K, this region is unlikely to be phosphate-binding region due to lack of homology with other nucleotide-binding proteins; but the region at the sequence 143-149, 143G-G-T-G-S-G-F¹⁴⁹, is accepted as the (pyro)phosphatebinding region. With occasional exceptions the sequence pattern-G(1st)-X-X(2nd)-X-X-G(3rd)-X' (X and X' are any amino residues and others are represented by single alphabet nomenclature for amino acids. X' is in general lysine for mononucleotidebinding proteins) can be proposed for both monoand dinucleotide-binding proteins.

For ras gene protein p21, whether the most crucial residue is glycine at position 12 (2nd glycine), according to Wierenga and Hol, or glycine at position 15 (3rd glycine) according to Molfer and Amons, is subjective and arguable. It is true for

(most of) the nucleotide-binding proteins that the glycine at the equivalent position 15 (3rd glycine) is invariant. In all the nucleotide-binding proteins the glycine at the equivalent position 10 (I glycine) is also invariant and has exceptional φ , ψ conformation angles. Even when a particular residue is crucial, the effects of neighbouring residues upon it cannot be over looked. For the oncogenic p21 proteins, to have such a specific topology, that a single-point mutation would lead normal cells to malignant cells, and the glycine at the position 12 could not be replaced by another residue without altering the nucleotide-binding properties of p21, the β - α junction should have a sharp turn (hairpin turn) so that G12 is at the 'apex' of the β - α turn; and perhaps the hydrogen bond between I and II β -strands keeps this 'rigid' $\beta \alpha \beta$ motif. When one considers the possible effects of neighbouring residues (of such a motif) on the 'crucial' residue, it appears likely that instead of a single residue at the β - α junction being crucial a patch consisting of more than one residue at this position would be more relevant in understanding normal protein function. As for ras gene p 21 proteins, a sharp bend (hairpin turn) at the position 12 may be a special case of the general structural motif $\beta \alpha \beta$ for the nucleotidebinding proteins.

ACKNOWLEDGEMENTS

The author is thankful to Dr M. R. N. Murthy of Indian Institute of Science, Bangalore for the technical help.

29 September 1986

- 1. Gay, N. J. and Walker, J. E., Nature (London), 1983, 301, 262.
- Capon, D. J., Chen, E. Y., Levinson, A., Seeburg, P. H. and Goeddel, D. V., Nature, (London), 1983, 302, 33.
- 3. Capon, D. J., Seeburg, P. H., McGrath, J. P., Hayflick, J. S., Edman, U., Levinson, A. D. and Goeddel, D. V., Nature (London), 1983, 304, 507.
- 4. DeFeo-Jones, D., Scolnick, E. M., Koeller, R. and Dhar, R., Nature (London), 1983, 396, 707.
- 5. Dhar, R., Ellis, R. W., Shih, T. Y., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. and Scolnick, E., Science, 1982, 217, 934.
- 6. Rasheed, S., Normal, G. L. and Heidecker, G., Science, 1983, 221, 155.
- 7. Reddy, E. P., Reynolds, R. K., Santos, E. and

- Barbacid, M., Nature (London), 1982, 300, 149.
- 8. Shih, T. Y., Papageorge, A. G., Stokes, P. E., Weeks, M. O. and Scolnick, E. M., *Nature* (London), 1980, 287, 686.
- 9. Tabin, C. J. et al., Nature (London), 1982, 300, 143.
- 10. Tsuchida, N., Ryder, T. and Ohtsubo, E., Science, 1982, 217, 937.
- 11. Shih, T. Y. and Weeks, M. O., Cancer Invest., 1984, 2, 109.
- 12. Hughes, S. M., FEBS Lett., 1983, 164, 1.
- 13. Tanabe, T. et al., Nature (London), 1985, 315, 242.
- 14. Yatsunami, K. and Khorana, H. G., Proc. Natl. Acad. Sci., USA, 1985, 82, 4316.
- 15. Hurley, J. B., Simon, M. I., Teplow, D. B., Robinshaw, J. D. and Gilman, A. G., Science, 1984, 226, 860.
- 16. Lochrie, M. A., Hurley, J. B. and Simon, M. I., Science, 1985, 228, 96.
- 17. Wierenga, R. K. and Hol, W. G. J., *Nature* (*London*), 1983, 302, 842.
- 18. Wierenga, R. K., Drenth, J. and Schulz, G. E., J. Mol. Biol., 1983, 167, 725.
- 19. Wierenga, R. K., Terpstra, P. and Hol, W. G. J., J. Mol. Biol., 1986, 187, 101.
- 20. Möller, W. and Amons, R., FEBS Lett., 1985, 186, 1.
- 21. Kiltz, H.-H., Keil, W., Griesbach, M., Petry, K. and Meyer, H., Hoppe'Seyler's Z.Physiol. Chem., 1977, 358, 123.
- 22. Taylor, S. S., J. Biol, Chem., 1977, 252, 1799.
- 23. Jörnavall, H., Eur. J. Biochem., 1970, 16, 25.
- 24. Davidson, B. E., Saigo, M., Noller, H. F., and Harris, J. I., *Nature* (*London*), 1967, 216, 1183.
- 25. Krauth-Siegal, R. L., Batterspiel, R., Saleh, M., Schilz, E., Schirmer, R. H. and Untucht-Grau, R., Eur. J. Biochem., 1982, 121, 259.
- 26. Hofsteenge, J., Weijer, W. J., Jekel, P. A. and Beintema, J. J., Eur. J. Biochem., 1983, 133, 91.
- 27. Weirenga, R. K., DeMaeyer, M. C. H., and Hol, W. G. J., Biochemistry, 1985, 24, 1346.
- 28. Palau, J., Argos, P. and Puigdomenech, P., Int. J. Peptide Protein Res., 1982, 19, 394.
- 29. McLachlan, A. D., J. Mol. Biol., 1971, 61, 409.
- Saraste, M., Gay, N. J., Eberle, A., Runswick, M. J. and Walker, J. E., Nucleic Acids Res., 1982, 9, 5287.
- 31. Runswick, M. J. and Walker, J. E., J. Biol. Chem., 1983, 258, 3081.
- 32. Gallwitz, D., Donath, C. and Sanders, C., Nature (London), 1983, 306, 704.

- 33. Von Zabern, I., Wittman-Leibold, B., Untucht-Grau, R., Schirmer, R. H. and Pai, E. F., Eur. J. Biochem., 1976, 68, 281.
- 34. Ponstingl, H., Krauhs, E., Little, M. and Kempf, T., *Proc. Natl. Acad. Sci., USA*, 1981, 78, 2757.
- Jones, M. D., Peterson, T. E., Nielson, K. M., Magnusson, S., Sotturup-Jensen, L., Gausing, K. and Clark, B. F. C., Eur. J. Biochem., 1980, 108, 507.
- 36. Laursen, R. A., L'Italien, J. J., Nagarkatti, S. and Miller, D. L., J. Biol. Chem., 1981, 256,

8102.

- 37. Hausinger, R. P. and Howard, J. B., J. Biol. Chem., 1982, 257, 2483.
- 38. Robson, R. L., FEBS Lett., 1984, 773, 394.
- 39. Wieland, B., Tomasselli, A. G., Noda, L. H., Frank, R. and Schluz, G. E., Eur. J. Biochem., 1984, 143, 331.
- 40. Shoji, S. et al., Proc. Natl. Acad. Sci., USA, 1981, 78, 848.
- 41. Kuby, S. A., Palmieri, R. H., Frischat, A., Fischer, A. H., Wu, L. H., Maland, L. and Manship, M., Biochemistry, 1984, 23, 2393.

NEWS

OLYMPUS 1 SATELLITE FOR SOLAR SIMULATION TESTS

The Space and Communications Division of British Aerospace has successfully completed the assembly and integration of the world's largest and most powerful civil three-axis stabilized communications satellite.

A specially chartered Belfast transport aircraft took the multipurpose Olympus 1 satellite from Stansted to Pasadena, California, where it will start its solar simulation tests in early June in NASA's jet propulsion laboratory. The tests are designed to prove that the thermal design of the spacecraft copes with severe extremes of temperature encountered in space.

From there it will go to the David Florida Laboratories in Canada for further environmental

testing and will be ready for launch in September 1988.

Assembled and integrated at the Space and Communications Division's Stevenage factory, the body of the satellite measures 2.9 m (9 ft. 6 in.) wide and is 5.6 m (18 ft. 4 in.) long. Its solar array stretches over 25.6 m (84 ft.).

Once in orbit, Olympus 1 will demonstrate its multi-communications payload, pioneering the way for a new generation of satellites which will represent the largest and most powerful range of communications satellites in the world. (BIS: B235, Issued by *British Information Services*, British High Commission, Chanakyapuri, New Delhi 110 021).

NATIONAL SYMPOSIUM ON INDUCTION MELTING SYSTEMS FOR FERROUS ALLOYS

The Director General, Technical Development, Government of India has sponsored holding of the above Symposium under the joint auspices of the Indian Institute of Metals and All India Induction Furnaces Association at New Delhi on 25th & 26th September, 1987.

The symposium will provide a forum for interaction amongst the technologists, entrepreneurs and users of ferrous alloy products. Since Induction melting system is becoming fairly a competitive route for melting ferrous castings, stainless steel ingots and castings and other ferrous alloys as such the two days deliberations all aspects of Induction melting system will be discussed.

Invited papers and contributed papers will be presented in the Symposium. Suitable recommendations will be made for consideration of concerned Government Departments and others. Further details can be had from the Secretariat Office for Symposium, B-31, Deepali, Pitampura, Delhi 110 034.