Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1814–1818, April 1987
Biochemistry

GTP-binding domain: Three consensus sequence elements with distinct spacing

(computer search/elongation factor Tu/foot-and-mouth disease virus protein 2C/nucleotide binding)

THOMAS E. DEVER*, MANUEL J. GLYNIAS†, AND WILLIAM C. MERRICK*

Departments of *Biochemistry and †Pharmacology, Case Western Reserve University, School of Medicine, Cleveland, OH 44106

Communicated by Harland G. Wood, November 20, 1986

ABSTRACT **A sequence comparison of nine functionally different GTP-binding protein families has yielded further information on the general characterization of the conservation and importance of amino acid sequences in the GTP-binding domain, including (*i*) a consensus sequence composed of three consensus elements GXXXXGK, DXXG, and NKXD with consensus spacings of either 40–80 or ≈130–170 amino acid residues between the first and second elements and ≈40–80 amino acid residues between the second and third sequence elements; (*ii*) the sequence NKXW in place of NKXD in the sequence element responsible for base specificity allows the use of ITP as well as GTP; (*iii*) dGTP can be used with essentially the same efficiency as GTP; (*iv*) signal transducing proteins and enzymes have been identified in the nine families; and (*v*) family conservations allow the identification of the most probable consensus sequence element when more than one is present. Employing these features we have screened the protein sequence data base of the Protein Identification Resource and have identified only known GTP-binding proteins with the exception of protein 2C from foot-and-mouth disease virus as matching the consensus sequence. Based on this finding we predict that foot-and-mouth disease virus protein 2C binds GTP and, by analogy, that protein 2C from several related viruses (polio, rhino, encephalomyocarditis, and cowpea mosaic) will bind a nucleotide as part of its biologic activity.**

The cloning and sequencing of many proteins have stimulated a search for common primary structure motifs that could be used to predict protein function. Nucleotide binding is a property that has been extensively studied and it would be very valuable if it could be predicted based on the primary structure. ATP-binding proteins have been characterized at the structural level by the Rossmann fold (1, 2) and attempts have been made to characterize a common sequence. However, the proposed common sequence elements, typically a glycine-rich region, cannot be characterized by a consensus sequence nor are they unique enough to have predictive value (3, 4). Recently, the x-ray structure of the GTP-binding domain of elongation factor Tu (EF-Tu) was reported by la Cour *et al.* (5) and a consensus sequence for GTP-binding domains was proposed (6). In contrast to the ATP-binding sequence, the GTP-binding sequence is more extensive and unique. The GTP-binding consensus sequence has been found in a wide variety of proteins performing diverse functions and having a high affinity for GTP. These proteins include the elongation factors, ras p21 protein, phospho*enol*-pyruvate carboxykinase (GTP) (EC 4.1.1.32) (PEPCK), and guanine nucleotide-binding proteins of adenylate cyclase (G proteins). Based on a further review of the literature we have refined the GTP-binding domain consensus sequence and have tested the predictive value of this consensus sequence by using it in a computer search of the Protein Identification Resource (PIR) protein sequence data base. In this paper we report our consensus sequence, a list of known GTP-binding proteins that match the consensus sequence, and the results of our computer search.

MATERIALS AND METHODS

Search of the PIR Protein Sequence Data Base. We screened the protein sequence data base of the PIR, supported by the Division of Research Resources of the National Institutes of Health, using the Bionet program "Quest" (Intelligenetics), looking for proteins that had the three elements of the consensus sequence (see text) in the proper order. In our initial screen we allowed conservative amino acid replacements: A for G, E for D, Q for N. At the time the data base was screened, August 1986 (release 10.0), ≈3800 protein sequences were present. The data base was screened for the following sequence: (G/A)XXXX(G/A)K, then (D/E)XX-(G/A), then (N/Q)KX(D/E) (where G/A means G or A and X can be any amino acid). The spacing between the three sequence elements was allowed to be any number of amino acids.

Calculation of the Probability of the Chance Occurrence of the Consensus Sequence. In the determination of the probability of the occurrence of the consensus sequence strictly by chance, several assumptions were made: (*i*) all amino acids occur with equal frequency (1/20); (*ii*) the length of the average protein was taken as 1000 amino acid residues; (*iii*) the spacing between the consensus elements GXXXXGK, DXXG, and NKXD is about 40–80 amino acids. The approximate probability of the three sequence elements appearing in a protein of 1000 amino acids is thus:

$$(1/20)^3(880)(1/20)^2(40)(1/20)^3(40) = 5.5 \times 10^{-5} \text{ or } 1/18{,}182.$$

Though the assumptions used may not be entirely valid (G and K often exceed 5%, the average molecular size is often taken as 450–600 amino acids), the chance occurrence of the three consensus elements correctly spaced would appear to be between 1/5000 and 1/10,000.

Assays. Polyphenylalanine synthesis was monitored by the incorporation of [^{14}C]phenylalanine into hot, trichloroacetic acid-precipitable radioactivity directed by poly(U) using EF-1, EF-2, and ribosomes as described (7). The purification of the protein synthesis factors (EF-1 and EF-2) was as reported (8). PEPCK was provided by T. Nowak (Univ. of Notre Dame, Notre Dame, IN). The PEPCK spectrophotometric coupled assay monitored the conversion of NADH to NAD over time as facilitated indirectly by PEPCK as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EF, elongation factor; PEPCK, phospho*enol*pyruvate carboxykinase (GTP); G protein, guanine nucleotide-binding regulatory protein of adenylate cyclase; PIR, Protein Identification Resource; IF, initiation factor; FMDV, foot-and-mouth disease virus.

described (9), except the assay was performed in the reverse direction as described in ref. 10. The GDP-binding activity was determined as protein-dependent retention of radiolabeled GDP on nitrocellulose filters (7).

RESULTS

GTP-Binding Domain Consensus Sequence. Based on the x-ray data of the GTP-binding domain of EF-Tu (5) and a sequence comparison of other known GTP-binding proteins, a consensus sequence can be found. The consensus sequence contains three consensus elements, GXXXXGK, DXXG, and NKXD, with a consensus spacing requirement of ≈40–80 amino acids between the first and second and between the second and third sequence elements. The first two elements are involved in interactions with the phosphate portion of the GTP molecule and the last element is involved in nucleotide specificity (6). Table 1 shows the alignment of the three elements of the GTP-binding domain consensus sequence for a number of GTP-binding proteins that have been sequenced. Table 1 is spaced to emphasize the different functional families [i.e., proteins from different sources that perform the same function: EF-Tu (EF-1) family, EF-G (EF-2) family, ras family, PEPCK family, G-protein family, plus the three families with one representative each: LepA, initiation factor 2 (IF-2), and GTP:AMP phosphotransferase]. Not all of the known examples for each of the families are listed due to space considerations. It is important to notice that even though there may be extensive amino acid sequence conservation within a family of proteins for the three sequence elements, it would be inappropriate to include these in the consensus sequence since they are not conserved in proteins from other functional families. Though not emphasized in Table 1, in most of the GTP-binding proteins that have been sequenced the spacing between the parts of the consensus sequence is 40–80 amino acids. The only exceptions to the spacing rule are GTP:AMP phosphotransferase, transducin, and the G proteins, which seem to have a spacing of about 150 amino acids between the first two sequence elements but the conserved spacing between the second and third sequence element.

The first consensus element GXXXXGK is similar to the glycine-rich area seen in many ATP-binding proteins (3, 4, 41, 42). However, a significant difference is the variability in the sequence for ATP-binding proteins and the strong conservation in sequence for the GTP-binding proteins (this difference will be further examined in the *Discussion*). For three of the proteins listed in Table 1 (EF-G, EF-2, and LepA), the first glycine in GXXXXGK is replaced by alanine. This finding of a conservative amino acid replacement was important in defining the rules for our computer search, which will be discussed later.

For two of the proteins in Table 1 (rho and G_o), only incomplete sequence data are available; however, based on their similarity to the family sequence they seem to match the elements of the consensus sequence. As shown in Table 1, transducin has two matches to the second conserved sequence element, DXXG. We would predict that DVGGQ (196–200) is most likely the proper element and not DSAGY (146–150) as proposed by McCormick *et al.* (6), since the DVGGQ sequence is also found in the G-protein family. This choice of DXXG sequence also allows for a conservation in the second spacing of 40–80 amino acids for all of the proteins

Table 1. Components of the GTP/GDP-binding site

PHOSPHORYL BINDING SEQUENCES					GUANINE SPECIFICITY BINDING SITE		
CONSENSUS SEQUENCE	Gly X X X X Gly Lys		Asp X X Gly		Asn Lys X Asp		REFERENCES
EF-Tu, E. coli	Gly His Val Asp His Gly Lys	(18-24)	Asp Cys Pro Gly His	(80-84)	Asn Lys Cys Asp	(135-138)	(11)
EF-Tu, Euglena chloro.	Gly His Val Asp His Gly Lys	(18-24)	Asp Cys Pro Gly His	(80-84)	Asn Lys Glu Asp	(135-138)	(12)
EF-Tu, yeast mito.	Gly His Val Asp His Gly Lys	(55-61)	Asp Cys Pro Gly His	(117-121)	Asn Lys Val Asp	(172-175)	(13)
EF-1α, yeast	Gly His Val Asp Ser Gly Lys	(14-20)	Asp Ala Pro Gly His	(91-95)	Asn Lys Met Asp	(153-156)	(14, 15)
EF-1α, A. salina	Gly His Val Asp Ser Gly Lys	(14-20)	Asp Ala Pro Gly His	(91-95)	Asn Lys Met Asp	(153-156)	(16)
EF-1α, human	Gly His Val Asp Ser Gly Lys	(14-20)	Asp Ala Pro Gly His	(91-95)	Asn Lys Met Asp	(153-156)	(17)
EF-G, E. coli	Ala His Ile Asp Ala Gly Lys	(16-22)	Asp Thr Pro Gly His	(87-91)	Asn Lys Met Asp	(141-144)	(18)
EF-2, hamster	Ala His Val Asp His Gly Lys	(26-32)	Asp Ser Pro Gly His	(104-108)	Asn Lys Met Asp	(158-161)	(19)
LepA, E. coli	Ala His Ile Asp His Gly Lys	(11-17)	Asp Thr Pro Gly His	(77-81)	Asn Lys Ile Asp	(131-134)	(20)
IF-2, E. coli	Gly His Val Asp His Gly Lys	(398-404)	Asp Thr Pro Gly His	(444-448)	Asn Lys Ile Asp	(498-501)	(21)
RAS 1, yeast	Gly Gly Gly Gly Val Gly Lys	(17-23)	Asp Thr Ala Gly Gln	(64-68)	Asn Lys Leu Asp	(123-126)	(22, 23)
RAS 2, yeast	Gly Gly Gly Gly Val Gly Lys	(17-23)	Asp Thr Ala Gly Gln	(64-68)	Asn Lys Ser Asp	(123-126)	(23, 24)
YP2, yeast	Gly Asn Ser Gly Val Gly Lys	(15-21)	Asp Thr Ala Gly Gln	(63-67)	Asn Lys Cys Asp	(121-124)	(25)
H-ras, N-ras, K-ras, human	Gly Ala Gly Gly Val Gly Lys	(10-16)	Asp Thr Ala Gly Gln	(57-61)	Asn Lys Cys Asp	(116-119)	(26-28)
p29 ras, rat	Gly Ala Arg Gly Val Gly Lys	(69-75)	Asp Thr Ala Gly Gln	(116-120)	Asn Lys Cys Asp	(175-178)	(29)
v-ras, mouse	Gly Ala Lys Gly Val Gly Lys	(10-16)	Asp Thr Ala Gly Gln	(57-61)	Asn Lys Cys Asp	(116-119)	(30)
v-ras H, mouse	Gly Ala Arg Gly Val Gly Lys	(10-16)	Asp Thr Thr Gly Gln	(57-61)	Asn Lys Cys Asp	(116-119)	(31)
v-ras K, mouse	Gly Ala Ser Gly Val Gly Lys	(10-16)	Asp Thr Thr Gly Gln	(57-61)	Asn Lys Cys Asp	(116-119)	(32)
rho, Aplysia	Gly Asp Gly Ala Cys Gly Lys	(12-18)	Asp Thr Ala Gly Gln	(59-63)	Asn Lys Lys Asp	(117-120)	(33)
rho, human	not determined		Asp Thr Ala Gly Gln	(same)	Asn Lys Lys Asp	(same)	(33)
PEPCK, chicken	Gly Asn Ser Leu Leu Gly Lys	(237-243)	Asp Glu Leu Gly Asn	(318-321)	Asn Lys Asp Trp	(388-391)	(34)
PEPCK, rat liver	Gly Asn Ser Leu Leu Gly Lys	(237-243)	Asp Ala Gln Gly Asn	(318-321)	Asn Lys Glu Trp	(388-391)	(35)
GTP:AMP phosphotransferase bovine	Gly Ala Pro Gly Ser Gly Lys	(12-18)	Asp Leu Thr Gly Glu	(150-154)	Asn Lys Ile Trp	(200-203)	(36)
Transducin α, bovine	Gly Ala Gly Glu Ser Gly Lys	(36-42)	Asp Ser Ala Gly Tyr Asp Val Gly Gly Gln	(146-150) (196-200)	Asn Lys Lys Asp	(265-268)	(37, 38)
G_s protein, bovine adrenal	Gly Ala Gly Glu Ser Gly Lys	(47-53)	Asp Val Gly Gly Gln	(223-227)	Asn Lys Gln Asp	(292-295)	(39)
G_s protein, rat brain	Gly Ala Gly Glu Ser Gly Lys	(47-53)	Asp Val Gly Gly Gln	(223-227)	Asn Lys Gln Asp	(292-295)	(40)
G_o protein, rat brain	– – – – – Gly Lys		Asp Thr Leu Gly Val Asp Val Gly Gly Gln		Asn Lys Lys Asp		(40)
G_i protein, rat brain	Gly Ala Gly Glu Ser Gly Lys	(40-46)	Asp Leu Ser Gly Val Asp Val Gly Gly Gln	(123-127) (201-205)	Asn Lys Lys Asp	(270-273)	(40)

The numbers in parentheses represent the amino acid residue number determined by protein sequencing or inferred protein sequence from recombinant DNAs. E. coli, *Escherichia coli*; A. salina, *Artemia salina*; chloro., chloroplast; mito, mitochondrion.

in Table 1, whereas the DSAGY sequence requires a second spacing of 115 amino acids.

A careful examination of Table 1 will show that PEPCK and GTP:AMP phosphotransferase do not adhere to the final (nucleotide specificity) consensus sequence element with a tryptophan in place of the consensus aspartic acid, NKXW (not D). From x-ray studies the asparagine residue in this sequence is proposed to interact with the keto group of the guanine ring, the lysine forms part of the hydrophobic pocket, and the aspartic acid interacts with the amino group of the guanine ring (5, 6, 43). The deviation in consensus sequence is consistent with the ability of both enzymes to use either guanine or inosine nucleotides (44, 45), whereas most of the other proteins cited will use only guanine nucleotides (46, 47). Further support for the importance of this sequence is the finding by Feig *et al.* (48) that a ras p21 mutant protein with the aspartic acid of NKXD replaced by asparagine (NKXN) has a lower affinity for GTP by a factor of 100. Additionally, Clanton *et al.* (49) have recently shown that in a ras p21 mutant protein a lysine or tyrosine in place of the asparagine in NKXD abolishes GTP-binding activity. Therefore, the altered sequence of the ITP-utilizing proteins and the GTP-binding properties of the ras p21 mutants verify the importance of the sequence NKXD for GTP binding.

Search of the PIR Protein Sequence Data Base. To examine the validity of the proposed GTP-binding domain consensus sequence, we have screened the PIR protein sequence data base. Since in EF-G, EF-2, and LepA the first glycine was conservatively replaced by alanine, we allowed for conservative amino acid substitutions, as defined in *Materials and Methods*, during the screening. In the initial screening of the data base we did not include our spacing restriction, and therefore we visually checked the spacing of the positive matches. Table 2 is a list of those proteins, not listed in Table 1, that were identified as potential GTP-binding proteins by means of the computer search. Based on our spacing rules, only consensus sequences 80–160 amino acids in length are acceptable (except G protein family members, which are ≈190–225 amino acids in length), and this restriction will eliminate many of the candidates in Table 2.

Two proteins, which had not previously been determined to be GTP-binding proteins, were found that exactly match our three consensus elements: foot-and-mouth disease virus (FMDV) protein 2C and α_2-macroglobulin. In FMDV the consensus elements are GKSGQGK (110–116), DDLG (160–163), and NKLD (243–246), whereas in α_2-macroglobulin the consensus elements are GLYTYGK (229–235), DCHG (254–257) or DGHG (350–353) or DEHG (377–380) or DMKG (496–499) or DVIG (527–530), and NKVD (543–546). The numbers in parentheses represent the amino acid residue number taken from the published sequences of FMDV protein 2C (50) and α_2-macroglobulin (51). Even though α_2-macroglobulin has multiple DXXG sequences and two meet the second spacing requirement of 40–80 amino acids, based on the overall spacing requirement we would predict that α_2-macroglobulin does not bind GTP since the consensus length of 317 amino acids is at least 100 amino acids too long. When tested, α_2-macroglobulin (and bovine serum albumin) lacked GDP-binding activity (although no activity was detected, the maximal possible level of activity as statistical error was <1% of the level observed with the control eukaryotic IF-2; data not shown). The FMDV protein 2C meets our spacing requirement and therefore matches our consensus GTP-binding domain sequence. Unfortunately, protein 2C is not well characterized for function or properties and we were unable to test it for GTP-binding activity.

The FMDV protein 2C belongs to a family of viral protein 2Cs. Table 3 is taken from Argos *et al.* (52) and we have added the rhino virus protein 2C sequence (53) to the table. All of the 2C proteins contain the first consensus element and

Table 2. Search of the PIR protein sequence data base

PIR code no.	Protein	Length of consensus sequence
No mismatch		
GNNY2F	Protein 2C, FMDV	137*
MAHU	α_2-Macroglobulin, human	317
One mismatch		
EGMSMG	Epidermal growth factor, mouse	451
JGECM	Maltose-binding protein, *E. coli*	202*
PWBHB	ATPase (β chain), barley chloroplast	95*
PWZMB	ATPase (β chain), maize chloroplast	95*
QQBE11	M_r 140,000 ribonucleotide reductase, Epstein–Barr virus	400
R3EC1	Ribosomal protein S1, *E. coli*	274
VMUT8B	Variant surface glycoprotein 117 precursor, trypanosome	189*
Two mismatches		
EZHU	Coagulation factor VIII, human	642
FNHU	Fibronectin, human	1932
FOMVVB	Core shell protein, baboon endogenous virus	274
MWKW1	Myosin heavy chain 1, nematode	632
RDECFF	Fumarate reductase, *E. coli*	153*
RNECB	DNA-directed RNA polymerase (β chain), *E. coli*	867
RNECC	DNA-directed RNA polymerase (β' chain), *E. coli*	740
TFHUP	Transferrin, human	163*
TVMVHZ	Kinase-related transforming protein (kit), FeSV	337
UDHUS	Stefin, human	89*
VMUT7R	Variant surface glycoprotein 7, trypanosome	271
WMTM18	M_r 183,300 protein, tobacco mosaic virus	478
WMTM8T	M_r 180,000 protein, tomato mosaic virus	478
XHHU3	Antithrombin III, human	220*

A mismatch represents a conservative amino acid substitution, and the length of the consensus sequence is the number of amino acid residues between the first and third consensus elements. FeSV, feline sarcoma virus.

*These proteins meet our spacing requirement.

Table 3. Viral protein 2C alignments

		GXXXXGK			
FMD	RPEPVVVCLR	GKSGQGKSFL	ANVLAQAIST	HF-TGRIDSV	WYCPPDPDHF
EMC	RCEPVVIVLR	GDAGQGKSLS	SQVIAQAVSK	TI-FGR-QSV	YSLPPDSDFF
Polio	RIEPVCLLVH	GSPGTGKSVA	TNLIARAIAE	--- REN-TST	YSLPPDPSHF
Rhino	RTEPVCVLIH	GTPGSGKSLT	TSIVGRAIAE	- - - HFN-SAV	YSLPPDPKHF
CPM	RKMPFTIFFQ	GKSRTGKSLL	MSQVTKDFQD	HYGLGG-ETV	YSRNPCDQYW
		DXXG			
FMD	DGYNQQSTVV	MDDLGQ--NP	DGKDFKYFAQ	MVSTTGFIPP	MASLEDKGKP
EMC	DGYENQFAAI	MDDLGQ--NP	DGSDFTTFCQ	MVSTTNFLPN	MASLERKGTP
Polio	DGYKQQGVVI	MDDLNQ--NP	DGADMKLFCQ	MVSTVEFIPP	MASLEEKGIL
Rhino	DGYQQQEVVI	MDDLNQ--NP	DGQDISMFCQ	MVSSVDFLPP	MASLDNKGML
CPM	SGYRRQPFVL	MDDFAAVVTE	PSAEAQ-MIN	LISSAPYPLN	MAGLEEKGIC
FMD	FNSKVI IATT	NLYSGFTPRT	- MVCPDAL-N	RRFH-FDIDV	SAKDGY- KIN
EMC	FTSQLVVATT	NL- PEFRPVT	- I AHYPAV - E	RRIT-FDYSV	SAGPVCSKTE
Polio	FTSNYVLAST	NS- SRISPPT	- VAHSDAL - A	RRFA-FDMDI	QVMNEYSR-D
Rhino	FTSNFVLAST	NS-NTLSPPT	- I LNPEAL - V	RRPG-FDLDI	CLHTTYTK - N
CPM	FDSQFVFVST	NF-LEVSPEA	KVRDDEAFKN	RRHVIVQVSN	DPAKAYDAAD
		NKXD			
FMD	NKLD I IKALE				
EMC	AGYKVLDVER				
Polio	GKLNMAMATE				
Rhino	GKLNAGMSTL				
CPM	FASNQ IYTI L				

The consensus elements are shown above the sequences and spaces have been inserted as proposed by Argos *et al.* (52) to maximize the alignment of the homologous regions of the proteins. The FMDV protein 2C sequence extends from amino acid residue 100 to 252. FMD, foot-and-mouth disease; EMC, encephalomyocarditis; CPM, cowpea mosaic.

FMDV and EMC virus have the second consensus element. As seen in Table 3, only the FMDV protein 2C contains the third element of the consensus sequence specifying guanine and therefore we would predict that only this protein would bind GTP. As there is a functional correlation along the genome for each protein in each of the viruses listed in Table 3 and we would predict FMDV protein 2C binds GTP, we would also predict that protein 2C from each of the other viruses binds nucleotide triphosphates, although we cannot predict the base specificity. Based on the conservation of the GXXXXGK sequence, Gorbalenya *et al.* (41) have also suggested a nucleotide-binding property for viral protein 2C from FMD, polio, encephalomyocarditis, and cowpea mosaic viruses. In a review of the literature, the sequence of another serotype of FMDV was found. This serotype has several amino acid substitutions in protein 2C and one of the substitutions is the asparagine in NKXD to a serine (54). We would continue to predict that this protein 2C binds a nucleotide; however, it is not certain that the protein would bind GTP.

Since the first two consensus elements are involved with phosphate binding (6), those proteins in Table 2 that have conservative amino acid replacement(s) in the consensus sequence could reflect the ability to bind phosphate, sugar-phosphate, RNA, DNA, or nucleotides. For example, the ATPases, which match the consensus sequence with one mismatch, could reflect the similarity in the phosphate-binding sequences found in ATP- and GTP-binding proteins.

GTP-Binding Pocket. From the x-ray studies on EF-Tu, the 2′ and 3′ hydroxyls of the GTP point away from the protein and are exposed to the solvent (5). Therefore, it should be possible to make substitutions at these positions without affecting activity. We have tested this prediction with some GTP-binding proteins by substituting dGTP for GTP in functional assays. We have found that EF-1 and EF-2 in polyphenylalanine synthesis utilize dGTP at least 93% as efficiently as GTP, whereas PEPCK in a functional assay utilizes dGTP at ≈85% the efficiency of GTP. Other proteins listed in Table 1 have been tested for activity with dGTP and the following proteins have been found to use dGTP with equal or nearly equal efficiency as GTP: EF-Tu (46); ras p21 (55, 56); PEPCK (44); and GTP:AMP phosphotransferase (47). Therefore, we predict that all of the proteins listed in Table 1 should use dGTP with about the same efficiency as GTP.

DISCUSSION

In this paper we have characterized a consensus amino acid sequence that identifies a protein as a GTP-binding protein. The consensus sequence includes the sequence elements GXXXXGK, DXXG, and NKXD with spacings of 40–80 amino acids between the first and second and between the second and third sequences. As described in *Materials and Methods*, the odds of a chance occurrence of the consensus sequence in a protein containing 1000 amino acids is approximately 1 in 5000 to 1 in 10,000. During a testing of the predictive value of the sequence in a screening of the PIR protein sequence data base, only known GTP-binding proteins were found to match this consensus sequence with the exception of FMDV protein 2C. Other groups have looked for homology among the GTP-binding proteins and have identified regions of homology containing limited amino acid sequence conservation (57, 58). However, our consensus sequence is significantly different from these regional homologies (which do not define specific amino acid sequences) in that our sequence is a true consensus based on the sequence of many GTP-binding proteins with dramatically different functions. Though it has been suggested that many of these guanine nucleotide-binding proteins (EFs, G proteins, tubulin, transducin) may be evolutionarily related, as they have a similar subunit composition, they can be ADP-ribosylated, and they undergo conformational changes (signal transduction) depending on the nucleotide bound (59), the finding of the consensus sequence also in PEPCK and GTP:AMP phosphotransferase extends the consensus sequence to proteins having enzymatic functions.

However, there are known GTP-binding proteins that have been sequenced and fail to match the consensus sequence. α- and β-tubulin bind GTP and yet fail to match the consensus sequence elements presented in this paper (60, 61). α-Tubulin has at least one property that may explain why it does not match the consensus sequence. α-Tubulin binds GTP in what is described as a "nonexchangeable" manner (62). This difference in GTP-binding property is consistent with the lack of a match to the consensus sequence; however, β-tubulin does not have any unusual properties that might explain its lack of the GTP-binding domain consensus sequence. Due to these known exceptions and the potential for more, we would suggest that there are at least several different GTP-binding domains, one characterized by α-tubulin, one by β-tubulin, and one by the proteins that match our consensus sequence with the possibility of a subset for those proteins that do not follow the "standard spacing" rule of 40–80 amino acids (i.e., GTP:AMP phosphotransferase, G protein, and transducin). Before any conclusive statements about different GTP-binding domains can be made, the various GTP-binding proteins must be studied by x-ray crystallography.

Recently, several laboratories have characterized a common glycine-rich sequence found in many ATP-binding and other nucleotide-binding proteins (3, 4, 41, 42). This sequence would match our first consensus element and therefore could be important for phosphate binding (6). However, there is a major difference between the glycine-rich region seen in the ATP-binding proteins and the consensus sequence element described here for the GTP-binding proteins. In the ATP-binding proteins, the glycine-rich region cannot be characterized by a defined sequence and used as a predictive tool. The ATP-binding proteins, however, have been characterized at the secondary and tertiary structure level by the Rossmann fold (1, 2). An interesting comparison to make is the lack of a true consensus sequence for the ATP-binding proteins and the strong conservation in sequence among the

different GTP-binding proteins in relation to the tightness of the binding of these nucleotides. A typical K_d for ATP-binding proteins is 50–200 μM, whereas a typical K_d for GTP-binding proteins is 1–10 μM. This much tighter binding by the GTP-binding proteins may explain their stronger conservation of sequence as compared to the ATP-binding proteins.

To conclude, a significant value of this GTP-binding domain consensus sequence is its possible predictive function. We have tested the predictive value of the consensus sequence on the PIR protein sequence data base and have found only one protein that matches the consensus sequence and is not a known GTP-binding protein. From this search, we predict that the FMDV protein 2C should bind GTP and the other virally related 2C proteins should bind a nucleotide (ATP, UTP, or CTP) as part of their function. With the rapid accumulation of more protein sequences, the consensus sequence presented in this paper can be used as a tool to predict whether a protein might bind GTP. Moreover, we would predict that a GTP-binding protein that allows substitutions at the 2′ and 3′ positions of the GTP molecule (indicating direct exposure of these hydroxyls to solvent) would be an excellent candidate to have the consensus sequence described in this paper, and a GTP-binding protein with NKXW in place of NKXD for the third element in the consensus sequence would bind ITP with equal efficiency as GTP. These predictions will be easily testable as more protein sequences become available.

This work was supported, in part, by National Institutes of Health Grants GM26796 (W.C.M.) and AM21594 (Alan G. Goodridge).

1. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) *Nature (London)* **250,** 194–199.
2. Rossmann, M. G., Liljas, A., Branden, C.-I. & Banaszak, L. J. (1975) *The Enzymes,* **11,** 61–102.
3. Fry, D. C., Kuby, S. A. & Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 907–911.
4. Weinmaster, G., Zoller, M. J. & Pawson, T. (1986) *EMBO J.* **5,** 69–76.
5. la Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. (1985) *EMBO J.* **4,** 2385–2388.
6. McCormick, F., Clark, B. F. C., la Cour, T. F. M., Kjeldgaard, M., Norskov-Lauritsen, L. & Nyborg, J. (1985) *Science* **230,** 78–82.
7. Merrick, W. C. (1979) *Methods Enzymol.* **60,** 108–123.
8. Merrick, W. C. (1979) *Methods Enzymol.* **60,** 101–108.
9. Jadus, M., Hanson, R. W. & Colman, R. F. (1981) *Biochem. Biophys. Res. Commun.* **101,** 884–892.
10. Ballard, F. J. & Hanson, R. W. (1967) *Biochem. J.* **104,** 886–871.
11. Arai, K., Clark, B. F. C., Duffy, L., Jones, M. D., Kaziro, Y., Laursen, R. A., L'Italien, J., Miller, D. L., Nagarkatti, S., Nakamura, S., Nielsen, K. M., Petersen, T. E., Takahashi, K. & Wade, M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1326–1330.
12. Montandon, P. & Stutz, E. (1983) *Nucleic Acids Res.* **11,** 5877–5892.
13. Nagata, S., Tsunetsugu-Yokota, Y., Naito, A. & Kaziro, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6192–6196.
14. Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M. & Kaziro, Y. (1984) *EMBO J.* **3,** 1825–1830.
15. Cottrelle, P., Thiele, D., Price, V. L., Memet, S., Micouin, J. Y., Marck, C., Buhler, J. M., Sentenac, A. & Fromagoet, P. (1985) *J. Biol. Chem.* **260,** 3090–3096.
16. Van Hemert, F. J., Amons, R., Pluijms, W. J. M., Van Ormondt, H. & Moller, W. (1984) *EMBO J.* **3,** 1109–1113.
17. Brands, J. H. G. M., Maassen, J. A., Van Hemert, F. J., Amons, R. & Moller, W. (1986) *Eur. J. Biochem.* **155,** 167–171.
18. Ovchinnikov, Y. A., Alakhov, Y. B., Bundulis, Y. P., Bundule, M. A., Dovgas, N. V., Kozlov, V. P., Motuz, L. P. & Vinokurov, L. M. (1982) *FEBS Lett.* **139,** 130–135.
19. Kohno, K., Uchida, T., Ohkubo, H., Nakanishi, S., Nakanishi, T., Fukui, T., Ohtsuka, E., Ikehara, M. & Okada, Y. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4978–4982.
20. March, P. E. & Inouye, M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7500–7504.
21. Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A. & Grunberg-Manago, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7787–7791.
22. DeFoe-Jones, D., Scolnick, E. M., Koller, R. & Dhar, R. (1983) *Nature (London)* **306,** 707–709.
23. Powers, S., Kataoka, T., Fasano, O., Goldfarb, M., Strathern, J., Broach, J. & Wigler, M. (1984) *Cell* **36,** 607–612.
24. Dahr, R., Nieto, A., Koller, R., DeFoe-Jones, D. & Scolnick, E. M. (1984) *Nucleic Acids Res.* **12,** 3611–3618.
25. Gallwitz, D., Donath, C. & Sander, C. (1983) *Nature (London)* **306,** 704–707.
26. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. (1983) *Nature (London)* **302,** 33–37.
27. Taparowsky, E., Shimizu, K., Goldfarb, M. & Wigler, M. (1983) *Cell* **34,** 581–586.
28. Shimizu, K., Birnbaum, D., Ruley, M., Fasano, O., Suard, Y., Edlund, L., Taparowsky, E., Goldfarb, M. & Wigler, M. (1983) *Nature (London)* **304,** 497–500.
29. Rasheed, S., Norman, G. L. & Heidecker, G. (1983) *Science* **221,** 155–157.
30. Reddy, E. P., Lipman, D., Andersen, P. R., Tronick, S. R. & Aaronson, S. A. (1985) *J. Virol.* **53,** 984–987.
31. Dhar, R., Ellis, R. W., Shih, T. Y., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. & Scolnick, E. (1982) *Science* **217,** 934–937.
32. Tsuchida, N., Ryder, T. & Ohtsubo, E. (1982) *Science* **217,** 937–939.
33. Madaule, P. & Axel, R. (1985) *Cell* **41,** 31–40.
34. Cook, J. S., Weldon, S. L., Garcia-Ruiz, J. P., Hod, Y. & Hanson, R. W. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7583–7587.
35. Beale, L. G., Chrapkiewicz, N. B., Scoble, H. A., Metz, R. J., Quick, D. G., Noble, R. L., Donelson, J. E., Biemann, K. & Granner, D. K. (1985) *J. Biol. Chem.* **260,** 10748–10760.
36. Tomasselli, A. G., Frank, R. & Schiltz, E. (1986) *FEBS Lett.* **202,** 303–308.
37. Yatsunami, K. & Khorana, H. G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4316–4320.
38. Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H. & Numa, S. (1985) *Nature (London)* **315,** 242–245.
39. Robishaw, J. D., Russell, D. W., Harris, B. A., Smigel, M. D. & Gilman, A. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1251–1255.
40. Itoh, H., Kozasa, T., Nagata, S., Nakamura, S., Katada, T., Ui, M., Iwai, S., Ohtsuka, E., Kawasaki, H., Suzuki, K. & Kaziro, Y. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3776–3780.
41. Gorbalenya, A. E., Blinov, V. M. & Koonin, E. V. (1985) *Mol. Genet. Mikrobiol. Virusol.* **11,** 30–36.
42. Moller, W. & Amons, R. (1985) *FEBS Lett.* **186,** 1–7.
43. Jurnak, F. (1985) *Science* **230,** 32–36.
44. Petrescu, I., Bojan, O., Saied, M., Barzu, O., Schmidt, F. & Kuhnle, H. F. (1979) *Anal. Biochem.* **96,** 279–281.
45. Tomasselli, A. G. & Noda, L. H. (1979) *Eur. J. Biochem.* **93,** 263–270.
46. Miller, D. L. & Weissbach, H. (1977) in *Molecular Mechanisms of Protein Biosynthesis,* eds. Weissbach, H. & Pestka, S. (Academic, New York), pp. 323–373.
47. Northup, J. K., Smigel, M. D. & Gilman, A. G. (1982) *J. Biol. Chem.* **257,** 11416–11423.
48. Feig, L. A., Pan, B.-T., Roberts, T. M. & Cooper, G. M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4607–4611.
49. Clanton, D. J., Hattori, S. & Shih, T. Y. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5076–5080.
50. Carroll, A. R., Rowlands, D. J. & Clarke, B. E. (1984) *Nucleic Acids Res.* **12,** 2461–2472.
51. Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Lonblad, P. B., Magnusson, S. & Petersen, T. E. (1984) *J. Biol. Chem.* **259,** 8318–8327.
52. Argos, P., Kamer, G., Nicklin, M. J. H. & Wimmer, E. (1984) *Nucleic Acids Res.* **12,** 7251–7267.
53. Callahan, P. L., Mizutani, S. & Colonno, R. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 732–736.
54. Forss, S., Strebel, K., Beck, E. & Schaller, H. (1984) *Nucleic Acids Res.* **12,** 6587–6601.
55. Scolnick, E. M., Papageorge, A. G. & Shih, T. Y. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5355–5359.
56. Shih, T. Y., Papageorge, A. G., Stokes, P. E., Weeks, M. O. & Scolnick, E. M. (1980) *Nature (London)* **287,** 686–691.
57. Halliday, K. R. (1984) *J. Cyclic Nucleotide Protein Phosphorylation Res.* **9,** 435–448.
58. Leberman, R. & Egner, U. (1984) *EMBO J.* **3,** 339–341.
59. Hughes, S. M. (1983) *FEBS Lett.* **164,** 1–8.
60. Ponstingl, H., Krauhs, E., Little, M. & Kempf, T. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2757–2761.
61. Krauhs, E., Little, M., Kempf, T., Hofer-Warbinek, R., Ade, W. & Ponstingl, H. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4156–4160.
62. Penningroth, S. M. & Kirshner, M. W. (1978) *Biochemistry* **17,** 734–740.