three common fragments between 12 and 16 kb, possibly representing internal genomic fragments.

We have followed the inheritance of these polymorphic bands in several crosses. Figure 4*a* shows a cross between two CBA/Ca animals; most bands seen in the eight offspring are of a similar size to bands present in the parents. Although we cannot demonstrate an unequivocal relationship between parental and offspring bands, the inheritance pattern is largely consistent with simple mendelian inheritance from both parents. But it is clear from Fig. 4*a* that there are several bands present in the offspring (indicated by arrowheads) that are of a size dissimilar to that of any parental band; four of the eight offspring possess what seem to be new alleles.

To reduce the number of inherited bands in the 70–140-kb region and thus simplify the analysis, we studied several C57BL/6J × DBA/2 crosses, one of which is illustrated in Fig. 4*b*. A number of bands, such as two indicated 'A', seem to correspond to one allele of a heterozygous locus, being inherited by roughly half of the offspring. The inheritance of the band marked 'B' is less clear. Six offspring show bands of a similar but not identical size, and although it is a small sample, it seems that the bands are either of two similar sizes. We cannot so far ascertain whether this reflects limited germline mosaicism or selection from a more extensively heterogeneous sperm population. Furthermore, there is a parental band in Fig. 4*a*, indicated by an arrowhead in parent P1, that is not seen in any of the offspring. This again may reflect mosaicism, further suggested by its apparent presence at a less than unitary amount. But it seems clear that the observed instability of $(TTAGGG)_n$ tract length through generations can account for the high degree of polymorphism seen between individuals of an inbred strain.

It is important to know the site and mechanism of the alteration in length. If variation occurs in somatic tissue, one might be able to detect mosaicism. Our preliminary data from a few animals did not detect any difference in the banding pattern between DNA from liver, brain, spleen or testes of the same animal (data not shown). Mosaicism may be rarely detectable, however, especially as it must occur sufficiently early in a stem-cell lineage so that a new band will be present in a sufficient proportion of cells to allow detection. In addition, highly pure spermatozoan samples obtained by careful dissection show identical banding patterns to liver DNA from the same animal (data not shown). Clearly most of the $(TTAGGG)_n$ tracts remain essentially unchanged on passage through meiosis.

It is not obvious why mice should have such large $(TTAGGG)_n$ tracts. Telomere elongation in the germ line must be closely matched to the rate of sequence loss, if gradual lengthening or shortening over future generations is to be avoided. This implies the existence of a mechanism that can alter the rate of telomere elongation in response to a change in the length of telomeres present in the cell; this could perhaps be controlled by monitoring the total $(TTAGGG)_n$ content of the genome. Whatever the mechanism of this control, it is clear that on an evolutionary timescale it maintains a greater average telomere length in the mouse. This greater average length does not necessarily require a much greater rate of telomere elongation per cell generation in the mouse. Furthermore, a similar length heterogeneity may occur in the mouse as is seen in humans, but is not detected as it is now small in proportion to the much larger telomere size. A genetic control of average telomere length might explain the strain-specific variation in average length illustrated in Fig. 1.

Whether long telomeres are a result of selection or simply a neutral change is not clear. Their size seems largely unchanged on passage to subsequent generations, as well as through somatic cell division, so it is unlikely that the extra length is a defence against rapid loss of sequence. Nor are mouse telomeres significantly reduced in size during the animal's lifespan; a 17-month-old individual still showed normal size distribution of fragments characteristic of its strain (data not shown). This, and the much longer telomeres of this short-lived species, suggests that telomere shortening is unlikely to have any causal role in ageing *in vivo*, in contrast to some recent speculations[9]. The shortening of human telomeres during ageing *in vivo*[10] may instead indicate that telomere maintenance is another metabolic process that senescent cells are unable to perform as efficiently. □

Received 27 June; accepted 2 August 1990.

1. Zakian, V. A. *A. Rev. Genet.* **23,** 579–604 (1989).
2. Moyzis, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85,** 6622–6626 (1988).
3. Meyne, J., Ratliff, R. L. & Moyzis, R. K. *Proc. natn. Acad. Sci. U.S.A.* **86,** 7049–7053 (1989).
4. Allshire, R. C., Dempster, M. & Hastie, N. D. *Nucleic Acids Res.* **17,** 4611–4627 (1989).
5. Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I. & Cooke, H. J. *Nature* **338,** 771–774 (1989).
6. de Lange, T. *et al. Molec. cell. Biol.* **10,** 518–527 (1990).
7. Biessmann, H., Carter, S. B. & Mason, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87,** 1758–1761 (1990).
8. Levis, R. W. *Cell* **58,** 791–801 (1989).
9. Harley, C. B., Futcher, A. B. & Greider, C. W. *Nature* **345,** 458–460 (1990).
10. Hastie, N. D. *et al. Nature* (in the press).
11. Allshire, R. C. *et al. Cell* **50,** 391–403 (1987).
12. Pietras, D. F. *et al. Nucleic Acids Res.* **11,** 6965–6983 (1983).

ACKNOWLEDGEMENTS. We would like to thank Ian Jackson and Nick Hastie for their enthusiastic interest in this project and for many helpful discussions, Arthur Mitchell for DNA preparations of pMR 150 and pMR 196, and Sarah Bowen and Vasker Bhattacherjee for comments on the manuscript. This work was supported by a grant from the Medical Research Council Human Genome Mapping Project.

High specificity of a phosphate transport protein determined by hydrogen bonds

Hartmut Luecke & Florante A. Quiocho*

Howard Hughes Medical Institute, and Departments of Biochemistry and Structural Biology and of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030 USA
Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251 USA

TRANSPORT of the essential nutrient phosphorus—primarily in the form of orthophosphate—into cells and organelles is highly specific. This is exemplified by the uptake of phosphate or its close analogue arsenate by bacterial cells by way of a high affinity active transport system dependent on a phosphate-binding protein; this system is unable to recognize other inorganic oxyanions and is, moreover, distinct from the one for sulphate transport[1,2]. The phosphate-binding protein is a member of a family of periplasmic proteins acting as initial high-affinity receptors for the osmotic shock-sensitive active transport systems or permeases for various sugars, amino acids, oligopeptides, and oxyanions[2,3]. We report here the highly refined 1.7 Å resolution X-ray structure of the liganded form of the phosphate-binding protein. The structure reveals the atomic features responsible for phosphate selectivity, either in monobasic or dibasic form, and the exclusion of sulphate. These features are fundamental to understanding phosphate transport systems and molecular recognition of charged substrates or ions in other biological processes.

Phosphate-binding protein (PBP) consists of a single polypeptide chain (relative molecular mass M_r 34,400 estimated from the sequence of 321 residues)[4,5] with one tight phosphate-binding site ($K_d = 1$ μM at pH 8.3)[6,7]. Whereas arsenate is also transported by the same system[1], sulphate is ineffective as a substrate or inhibitor (binding to PBP at least five orders of magnitude less tightly than phosphate).

Crystallization of PBP at pH 4.5 has already been described[7]. The determination of the PBP structure[8] is based on initial phasing from single isomorphous replacement from a four-site KI/I_2 derivative with anomalous scattering contribution (Table

* To whom correspondence should be addressed at: Howard Hughes Medical Institute, One Baylor Plaza, Houston, Texas 77030, USA.

TABLE 1 Crystallographic and refinement data

	Native	KI/I_2
Resolution (Å)	1.7	2.5
Diffraction data*		
Number of observations	102,157	97,758
Number of unique reflections	36,946	11,673
Completeness (%)	95.4	95.9
$\langle I\rangle/\langle\sigma_I\rangle$	22.0	38.6
R_{merge} (%)	5.6	4.2
$R_{Bijvoet}$ (%)		2.1
Phasing statistics		
Number of Bijvoet pairs		7,893
Number of heavy atom sites		4
r.m.s. F_H/residual		2.03
R_{Cullis}		0.57
r.m.s. R_{anom}		0.56
Mean figure of merit		0.69
Restrained least-squares refinement statistics		
R_{cryst} (%)	14.7	
Protein atoms	2,439	
Phosphate atoms	5	
Water molecules	259	
r.m.s. deviations from ideal stereochemistry		
Bond distances (Å)	0.016	
Angle distances (Å)	0.037	
Dihedral distances (Å)	0.045	
Planar group distances (Å)	0.010	
Peptide ω-angle (°)	2.0	
Chiral volume (Å^3)	0.147	

* Each data set was collected from one crystal using an ADSC two area detector system mounted on a Rigaku RU200 rotating anode. The native and derivative crystals are isomorphous: space group $P2_12_12_1$ with cell dimensions $a=41.97$ Å, $b=63.98$ Å, $c=123.97$ Å. The KI/I_2 derivative was obtained following the procedure described in ref. 17. R_{merge} is the conventional R-factor on intensities for symmetry-related reflections. $R_{Bijvoet}$ is the R-factor on intensities for Bijvoet pairs.

$$R_{cryst}=\Sigma|F_{obs}-F_{calc}|/\Sigma|F_{obs}|;\quad R_{Cullis}=\Sigma\||F_{PH}\pm F_P|-F_H|/\Sigma|F_{PH}-F_P|;$$
$$\text{r.m.s.}\quad R_{anom}=[\Sigma(\Delta F^{\pm}_{obs}-\Delta F^{\pm}_{calc})^2/\Sigma(\Delta F^{\pm}_{obs})^2]^{1/2}.$$

1), solvent-flattening[9], model building, phase combination[10], simulated annealing[11], and, finally, restrained least squares refinement[12] at 1.7 Å resolution which yielded an R-factor of 0.147. The crystallographic results are summarized in Table 1 and the legend to Fig. 1. We have also obtained a well-refined 1.8 Å isomorphous structure of PBP at pH 6.2 which is virtually identical to the pH 4.5 structure. The 1.7 Å electron density of the phosphate-filled substrate-binding site is exceptionally well defined (Fig. 1*a*). This is consistent with analysis, which shows that the central region of the molecule surrounding the binding site has the lowest mean thermal B value (Figs 1*b* and 2*a*) and, hence, has the most accurate atomic coordinates.

Binding proteins vary in size (M_r values 25,000–50,000) and in sequence[2,3]. Nevertheless, the structure of PBP (Fig. 1*b*) is similar overall to the six other binding protein structures determined in our laboratory[2,13]. (A detailed description of the structure will be published elsewhere.) It consists of two separate similarly folded globular domains. The deep cleft between the domains contains the bound phosphate which is totally dehydrated and completely sequestered ~8 Å below the protein surface.

The buried phosphate is held tightly in place by 12 strong hydrogen bonds, all formed with protein groups—5 NHs of the main chain, 2 NHs of the Arg 135 side chain (which is also salt-linked with Asp 137), 4 OHs of two serines and two threonines, and 1 oxygen of Asp 56 side chain (Fig. 2 and Table 2). Three of the phosphate oxygens (O1, O2, and O3) are heavily involved by accepting ten hydrogen bond donors, six of which originate from a seven-residue segment folded into a loop (residues 135–138) and the first turn of helix VI (139–141). The O2 is the recipient of four hydrogen bonds in a rare pyramidal arrangement with the heavy atoms (2 N and 2 O) of the four donor groups forming the base and the O2 atom the apex. There is pairing of all polar groups and total involvement of all donatable protons in the binding site region, giving rise to extensive networks or arrays of hydrogen bonds, and to precisely and stably oriented functional residues.

The interactions associated with the phosphate O4 are not only distinctive but are also the key to molecular understanding of the exquisite specificity of PBP (Fig. 2). The O4 accepts a hydrogen bond from an NH of a backbone peptide unit whose carbonyl oxygen is in turn the recipient of two hydrogen bonds. It is also in close contact (2.45 Å) with $O_{\delta 2}$ of the Asp 56 carboxylate side chain. This close association demands that the two oxygens share a proton on the O4 phosphate, resulting in a hydrogen bond between the phosphate and the carboxylate side chain. This accounts for the slight preference for dibasic phosphate[6,7]. Also, the hydrogen bonding between the phosphate and carboxylate is reminiscent of the short H bonds in the carboxylate-carboxyl interactions found in small compounds[14] and in proteins[15].

TABLE 2 Hydrogen bonds between phosphate and PBP

Phosphate atom	Hydrogen bond partner Domain	Residue	Group	Distance (Å) H···Y	X···Y	Angle (°) X—Ĥ···Y
O1	N	Thr 10	NH	1.85	2.81	156.0
	N	Thr 10	$O_\gamma H$	1.73	2.69	167.2
	C	Arg 135	$N_{\eta 1}H$	1.82	2.84	172.3
O2	C	Arg 135	$N_{\eta 2}H$	1.96	2.90	148.9
	C	Ser 139	$O_\gamma H$	1.80	2.77	169.0
	C	Thr 141	NH	1.84	2.83	159.2
	C	Thr 141	$O_\gamma H$	1.79	2.68	149.7
O3	N	Ser 38	NH	1.68	2.67	163.9
	N	Ser 38	$O_\gamma H$	1.69	2.63	160.9
	C	Gly 140	NH	1.74	2.76	177.1
O4	N	Phe 11	NH	1.92	2.92	163.3
O4—H	N	Asp 56	$O_{\delta 2}$	1.57	2.45	151.2
Mean				1.78 (0.11)	2.75 (0.13)	161.6 (8.7)

In order to obtain hydrogen bond distances and angles, all hydrogen atoms were placed and energy-minimized with the program AMBER 3.0[23] while freezing the refined positions of the heavy atoms and heavily constraining the hydrogens fixed by geometry (such as those in peptide units, Arg, etc). X is the hydrogen bond donor group and Y the acceptor group (see Fig. 2*b*). Note that the hydrogen bond between the phosphate O4—H and Asp 56 $O_{\delta 2}$ is significantly shorter than the mean value. Also, the main chain NH and side chain OH groups of three of the hydroxyl-containing residues (10, 38, 141) simultaneously serve as H-bond donors, and the hydroxyl oxygen (X) to phosphate oxygen (Y) distances tend to be short.

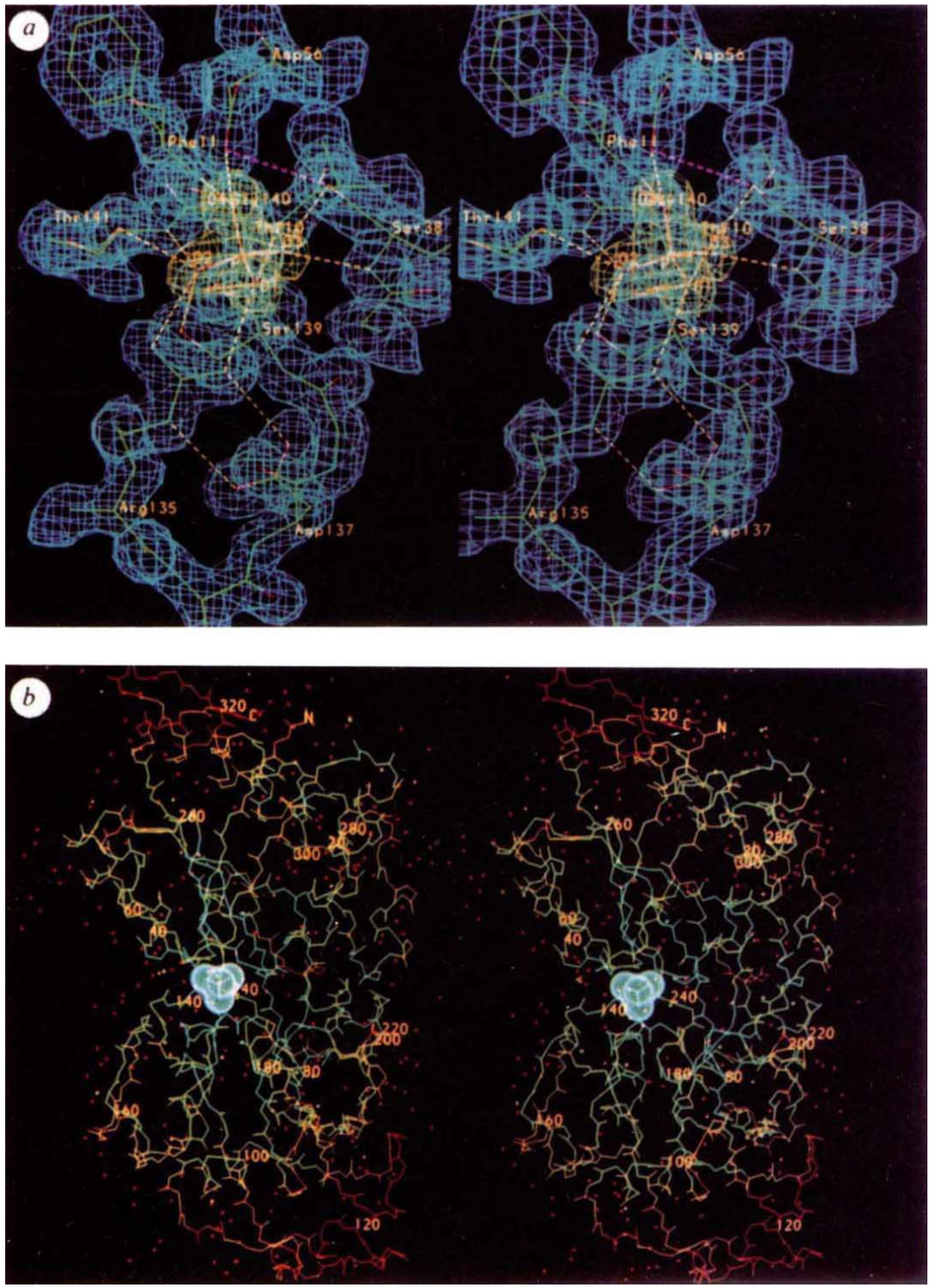


FIG. 1 *a*, Stereoscopic view of the binding site of PBP with bound phosphate. This view shows the 1.7 Å electron density (calculated with $(2|F_o| - |F_c|)$ coefficients and final α_c phases) and superimposed refined model. The density surface (phosphate in yellow and protein in blue) was contoured at 0.5 e/Å^3 and a grid-spacing of 0.5 Å. Dashed lines represent hydrogen bonds. CHAIN, a density-fitting and molecular graphics program developed by Dr J. S. Sack, was used in this study. *b*, The refined main chain atoms of PBP (total of 321 residues), the phosphate molecule shown with van der Waals surface and water molecules represented by dots. Every 20th α-carbon is numbered, and N and C designate the amino- and carboxy-terminal ends, respectively. The atoms are coloured according to their temperature factors: blue for *B*-values below 10 Å^2, green between 10 Å^2 and 15 Å^2, yellow between 15 Å^2 and 20 Å^2, and red over 20 Å^2. Starting from the binding site region which has the lowest average *B*-factor (Fig. 2*a*), the *B*-factor increases radially so that eventually both extremities of the ellipsoidal protein exhibit the highest motion. The molecule has dimensions of 35 Å × 40 Å × 70 Å, and consists of two domains (designated N and C) which are connected by two different peptide segments. The first domain (N domain) is made up of two segments (residues 1–72 and 257–321) and the second domain (C domain) is composed of a single segment (90–209).

The structure also reveals a simple molecular mechanism for binding—depending on the pH—dibasic (HPO_4^{2-}) or monobasic ($H_2PO_4^-$) phosphate (see Fig. 2*a*, *b*). In binding the monobasic species at acidic pH, Ser 38 γ-OH is the only other group favourably positioned to accept—through its oxygen lone pairs—the second phosphate proton (on O3) whilst donating its proton to Asp 56 $O_{\delta 2}$ (distance 3.28 Å) (Fig. 2*a*). In this binding mode, the negative charge on the phosphate will be concentrated on the two oxygens (O1 and O2) interacting with Arg 135. At pH around 8, where slightly tighter binding for the dibasic species is observed, the hydrogen-bonding interactions involving the Ser 38 γ-OH switches to that shown in Fig. 2*b*. In these interactions, the γ-OH forms a bifurcated hydrogen bond by also donating to the unprotonated phosphate O3 (distance 2.62 Å).

Asp 56 has at least three essential roles. It is responsible for the recognition of a proton on the phosphate. By accepting a hydrogen bond from the Ser 38 hydroxyl, Asp 56 facilitates the hydroxyl oxygen in accepting a second phosphate proton. Lastly and of critical importance, Asp 56 is the only residue capable of totally discriminating sulphate or similar tetrahedral fully ionized divalent oxyanions. Although these oxydianions have structural features very similar to those of phosphate[16], they would be repulsed by the negatively charged side chain.

Typical of phosphate transport systems, sulphate or similar fully ionized oxydianions cannot substitute for phosphate or

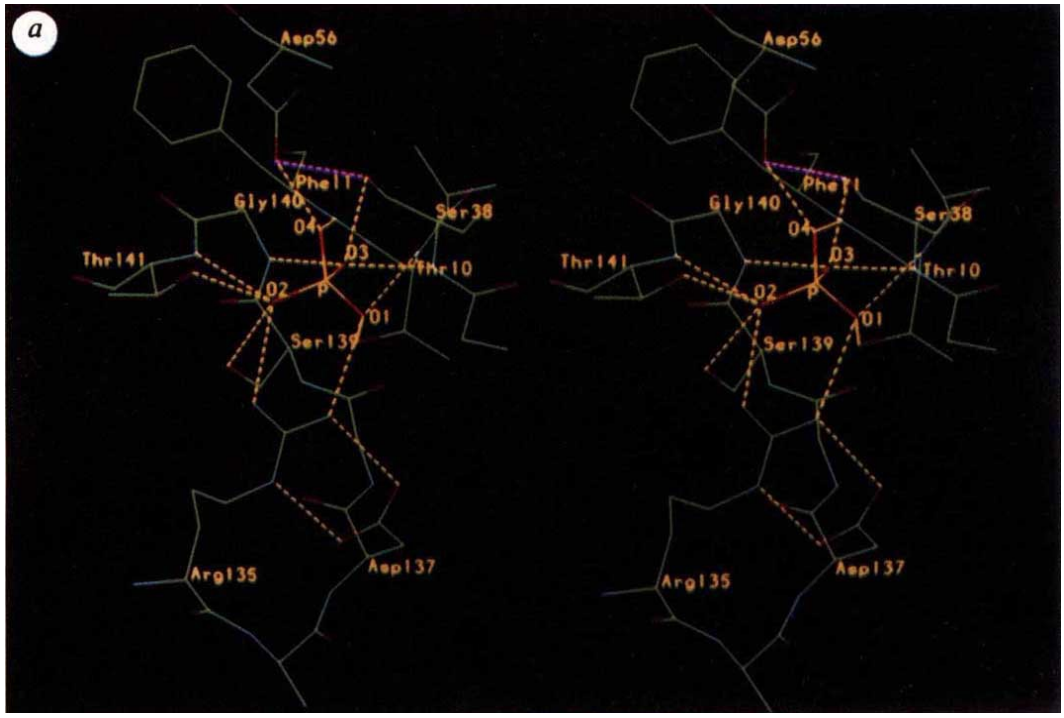
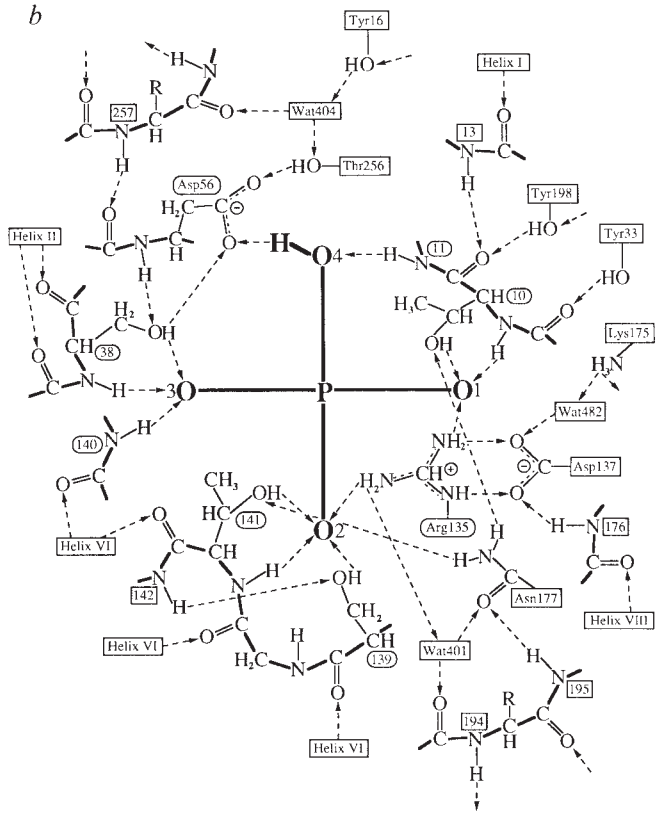


FIG. 2 Hydrogen-bonding interactions in the PBP–phosphate complex. *a*, A stereo view of the hydrogen bonds (dashed lines) between PBP and dibasic phosphate. This figure is identical to Fig. 1*a*, without the electron density contour. There are no metallic cations in the binding site, and the closest water molecule to the phosphate is at a distance of 5.3 Å. The mean isotropic *B*-factor for the atoms shown is 8.3 (2.1) $Å^2$ which is to be compared with 19.4 (11.5) $Å^2$ for all non-hydrogen protein atoms. The hydrogen bond parameters are listed in Table 2. Note the salt link between Arg 135 and Asp 137. The hydrogen bond between the donor Ser 38 γ-OH and acceptor Asp 56 $O_{\delta 2}$ is coloured differently (magenta) to indicate its role in binding monobasic and dibasic phosphates (see text). PBP binds phosphate maximally at pH about 8.3, but only about threefold decrease in binding is observed at pH 5 and 9.5 (B. L. Jacobson and F.A.Q., unpublished; ref. 6). This suggests binding of both monobasic ($H_2PO_4^-$) and dibasic (HPO_4^{2-}) phosphate, although the latter may be slightly preferred. In the binding of dibasic phosphate, the Ser 38 γ-OH forms a bifurcated hydrogen bond with the phosphate O4 and Asp 56 $O_{\delta 2}$ (3.28 Å distance). In the binding of monobasic phosphate, Ser 38 hydroxyl accepts a hydrogen bond from the protonated O3 and donates to Asp 56. *b*, Schematic diagram of the hydrogen bonds between PBP and dibasic phosphate and the networks of hydrogen bonds involving other groups in the binding site. With the exception of the peptide carbonyl oxygen of residue 9 which is hydrogen bonded to Tyr 33, the other four peptide units directly associated with the phosphate are further coupled to hydrogen-bond arrays involving other main chain peptide units (indicated by arrows). The two left-handed arrays of alternating hydrogen-bonded peptide units associated with helices II (residues 37–47) and VI (139–150) lead to the bulk solvent (Figs 1*b* and 2*b*; see also refs 17–18). The N termini of helices II and VI are near the phosphate.



arsenate. In fact, the uptake of sulphate by the bacterial high-affinity system requires a sulphate-binding protein (SBP)[2,3,17]. In contrast to PBP, the sulphate-binding protein binds only fully ionized oxyanions, including chromate and selenate[16,17]. We have previously refined the 2 Å structure of the SBP-sulphate complex[18–19]. Many features of the atomic interactions between SBP and sulphate are present in the PBP–phosphate complex. The totally desolvated sulphate is completely sequestered in a similar cleft. The main interaction is through hydrogen bonds with the four sulphate oxygens accepting a total of seven hydrogen bonds—one serine OH, five main chain NH groups, and one NH of a Trp side chain. Hydrogen bond arrays are formed, radiating from all four sulphate oxygens. Finally, in the SBP–sulphate complex and in the PBP–phosphate complex, the constellation of hydrogen bond dipoles and other nearby local dipoles (including those from the N termini of helices or other arrays) are responsible for stabilizing or dissipating the charges on the bound oxyanions (see also ref. 20). The isolated negative charge on the buried Asp 56 of PBP is similarly stabilized (Fig. 2*b*).

Despite these many common features, there is one crucial difference between the phosphate and sulphate receptors: whereas a carboxylate side chain—having key roles as indicated above—is present in the binding site of PBP, it is absent in the SBP. In fact, there is no group in the binding site of SBP that is suited to be a hydrogen bond acceptor[16]. Thus substrate recognition critically hinges on the protonated or deprotonated state of the oxyanions and on the presence or absence of hydrogen-bond acceptor groups in the binding site.

These findings provide greater understanding of electrostatic interactions in molecular recognition of charged ligands and in protein structures (related results refs 13, 19–20). Charges on buried groups can be stabilized by means other than salt linkages. Hydrogen bonds, involving primarily main chain peptide and

hydroxyl groups, are of importance in conferring specificity and affinity for negatively or positively charged ligands. These bonds are also desirable for efficient active transport and rapid ion movement[2,13,19-20]. Although Arg 135 participates in phosphate binding, salt-linking with Asp 137 (Fig. 2) facilitates transport by diminishing the charge-coupling interaction between the guanidinium and phosphate. This novel mechanism is likely also to be important in ion recognition and rapid movement (see below). There is widespread involvement of hydroxyl side chains in hydrogen bonding interactions with negative and positive groups of charged substrates and side chains as well (the buried Asp 56 and phosphate are best examples, Fig. 2; see also ref. 20). A survey of several refined protein structures reveals similar involvement of hydroxyl-containing residues. These observations are related to the bifunctional nature of hydroxyls; mimicking the dipolar properties of water molecules and having rotational degrees of freedom, hydroxyls are able to accept (through the lone pairs) or donate hydrogen bonds or to wholly or partially substitute hydration of anions or cations. Ion channel receptors, especially those which are gated by ligands such as acetylcholine, glycine and GABA, contain hydroxyl-rich regions (reviewed in ref. 21). Also, channel receptors, especially those that are voltage gated (such as sodium, potassium and calcium channels; reviewed in ref. 22), have regions with high concentrations of charged residues. By being salt-linked (like Arg 135 or Asp 137, Fig. 2) or by dissipating isolated charges, the strength of ion binding by the charged residues in these receptors will be moderated to facilitate fast ion movement. The assemblage of polar groups from these key side chains and/or the main chain is also likely to be important in the selectivity and mechanisms of action of these channel receptors and in a variety of biological processes as well (see also ref. 20).

In conclusion, our results not only constitute prime molecular evidence for the high phosphate specifity of a transport receptor protein but also demonstrate the power of hydrogen bonds in enabling proteins to distinguish between phosphate and sulphate. These findings provide new insights into charged ligand recognition and rapid rates in active transport, a process essential to living cells. □

Received 22 May; accepted 20 July 1990.

1. Medvecsky, N. & Rosenberg, H. *Biochim. biophys. Acta* **241,** 494-506 (1971).
2. Quiocho, F. A. *Phil. Trans. R. Soc.* B**341,** 341-351 (1990).
3. Furlong, C. E. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (ed. Neidhardt, F. C.), 768-796 (American Society for Microbiology, Washington, DC, 1987).
4. Surin, B. P. *et al. J. Bact.* **157,** 772-778 (1984).
5. Magota, K. *et al. J. Bact.* **157,** 909-917 (1984).
6. Medvecsky, N. & Rosenberg, H. *Biochim. biophys. Acta* **211,** 158-168 (1970).
7. Kubena, B. D., Luecke, H., Rosenberg, H. & Quiocho, F. A. *J. biol. Chem.* **261,** 7995-7996 (1986).
8. Luecke, H. thesis, Rice Univ. (1990).
9. Wang, B. C. *Meth. Enzym.* **115,** 90-112 (1985).
10. Read, R. *Acta crystallogr.* **A42,** 140-149 (1986).
11. Bruenger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458-460 (1987).
12. Hendrickson, W. A. *Meth. Enzym.* **115,** 252-270 (1985).
13. Sack, J. S., Saper, M. A. & Quiocho, F. A. *J. molec. Biol.* **206,** 171-191 (1989).
14. Speakman, J. C. *Structure & Bonding* **12,** 141-199 (1972).
15. Sawyer, L. & James, M. N. G. *Nature* **295,** 76-80 (1982).
16. Jacobson, B. L. & Quiocho, F. A. *J. molec. Biol.* **204,** 783-787 (1988).
17. Pardee, A. B. *J. biol. Chem.* **241,** 5886-5892 (1966).
18. Pflugrath, J. W. & Quiocho, F. A. *Nature* **314,** 257-260 (1985).
19. Pflugrath, J. W. & Quiocho, F. A. *J. molec. Biol.* **200,** 163-180 (1988).
20. Quiocho, F. A., Sack, J. S., & Vyas, N. K. *Nature* **327,** 561-564 (1987).
21. Barnard, E. A., Garlison, M. G. & Seeburg, P. *Trends Neurosci.* **10,** 502-509 (1987).
22. Cattrell, W. A. *Science* **242,** 50-61 (1988).
23. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. *J. comp. Chem.* **7,** 230-252 (1986).

ACKNOWLEDGEMENTS. This work was supported by the Howard Hughes Medical Institute and the NIH.

In vitro genetic analysis of the *Tetrahymena* self-splicing intron

Rachel Green, Andrew D. Ellington & Jack W. Szostak

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

THE availability of methods for the amplification of nucleic acid sequences[1,2] allows the genetic analysis *in vitro* of the structural and functional properties of many nucleic acids. We have now developed an *in vitro* selection and amplification system, and used it to analyse the self-splicing *Tetrahymena* ribozyme. A much wider range of selective conditions can be used *in vitro* than is possible with standard *in vivo* methods, and many more variants can be handled *in vitro* than *in vivo*. This method can be used to isolate the wild-type ribozyme, and structural variants that are as active as the wild type, from a pool of over 250,000 variants in only three cycles of selection and amplification.

We have developed an iterative procedure for the selective enrichment of catalytically active RNA molecules, by combining *in vitro* RNA replication with an *in vitro* selection system. The phenotype on which the *in vitro* selection is based is an RNA-catalysed RNA ligation reaction. The ribozyme involved is a derivative of the *Tetrahymena* self-splicing intron[3] that ligates itself to an RNA oligonucleotide complementary to the internal guide sequence[4] (Fig. 1*a*). This reaction is simply the reverse of the first step in self-splicing, with the added oligonucleotide representing the 5′ exon.

The added sequence allows selective amplification of catalytically active molecules, as shown in Fig. 1*b*. The RNA is first copied into complementary DNA by reverse transcriptase extension of a primer complementary to its 3′ end. A second primer that hybridizes exclusively to the added sequence information is then used for exponential amplification of the subset of molecules that have undergone the RNA-catalysed ligation reaction. Inactive molecules that fail to ligate cannot be amplified. DNAs equivalent to the original template are regenerated by a second round of polymerase chain reaction (PCR) with a different 5′ primer that carries the T7 promoter and reinserts the 5′ guanosine lost during the ligation reaction. RNAs transcribed from this DNA are enriched in catalytically active species; iteration of this procedure through many cycles allows the purification of active enzyme variants, even when these are present at very low abundance in a large pool of inactive variants.

To provide a stringent test of this system, we generated a large pool of sequence variants in which wild-type was present at an extremely low level. This was done by completely randomizing the bases in a region of the ribozyme where phylogenetic comparisons of group I introns had revealed an interesting potential covariation. This region included the 3′ end of the base-paired stem P3 and the adjacent joining regions J7/3 and J3/4 (Fig. 1*a*). In 32 out of 35 group Ib introns, the base pair at P3-7 is C·G, the nucleotide at J7/3-2 is an A, and J3/4 contains three bases; however, in three exceptional introns, P3-7 is G·C, J7/3-2 is U or G, and J3/4 contains only two bases (ref. 5; and F. Michel personal communication). In addition, the prevalence of C·G at P3-7 in the group Ib introns can be contrasted with the variety of base pairs at P3-7 in the group Ia introns. On the basis of these phylogenetic data it seemed plausible that nucleotides in these juxtaposed regions interacted with each other. We then used the *in vitro* selection scheme to establish the rules of variation for this region by selecting for all of its possible functional variants.

A pool of sequence variants for this region was generated by randomizing the nine bases that make up P3-7, P3-8, J7/3 and J3/4. Two primers that spanned these regions were prepared with random (equimolar mix of A, T, C and G) sequences at only the nine chosen bases. These two primers were used for PCR amplification of a DNA fragment corresponding to an internal portion of the ribozyme, using 10 ng of plasmid pRG1006 as a template. After amplification, ~1 μg of the fragment was gel-purified, digested with the restriction enzymes *Eco*RI and *Bgl*II, and ligated into plasmid cut with *Eco*RI and *Bgl*II containing the remainder of the ribozyme sequence. The