The bacteriophage φ29 portal motor can package DNA against a large internal force

Douglas E. Smith*†, Sander J. Tans*†, Steven B. Smith‡, Shelley Grimes§, Dwight L. Anderson§ & Carlos Bustamante†‡‖¶

† *Department of Physics,* ‖ *Department of Molecular and Cell Biology,* ‡*Howard Hughes Medical Institute,* ¶ *Physical Biosciences Division of Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA*
§ *Departments of Microbiology and Oral Science, University of Minnesota, Minneapolis, Minnesota 55455, USA*
* *These authors contributed equally to this work*

As part of the viral infection cycle, viruses must package their newly replicated genomes for delivery to other host cells. Bacteriophage φ29 packages its 6.6-μm long, double-stranded DNA into a 42 × 54 nm capsid[1] by means of a portal complex that hydrolyses ATP[2]. This process is remarkable because entropic, electrostatic and bending energies of the DNA must be overcome to package the DNA to near-crystalline density. Here we use optical tweezers to pull on single DNA molecules as they are packaged, thus demonstrating that the portal complex is a force-generating motor. This motor can work against loads of up to 57 pN on average, making it one of the strongest molecular motors reported to date. Movements of over 5 μm are observed, indicating high processivity. Pauses and slips also occur, particularly at higher forces. We establish the force–velocity relationship of the motor and find that the rate-limiting step of the motor's cycle is force dependent even at low loads. Notably, the packaging rate decreases as the prohead is filled, indicating that an internal force builds up to ~50 pN owing to DNA confinement. Our data suggest that this force may be available for initiating the ejection of the DNA from the capsid during infection.

The *Bacillus subtilis* phage φ29 is an excellent model system for studying viral assembly[3]. The 19.3-kilobase (kb) genome with its terminal proteins, gp3, is packaged into a preformed capsid (prohead) in an efficient *in vitro* assay[4,5]. The DNA packaging motor lies at a unique portal vertex of the prohead[1,6,7] and contains: (1) the head–tail connector (a dodecamer of gp10), the structure of which has been solved recently by X-ray crystallography[6]; (2) the portion of the prohead shell that surrounds the connector; (3) a ring of 174-base prohead RNAs (pRNA), which surrounds the protruding narrow end of the connector; and (4) a multimer of gp16, an ATPase that first binds DNA-gp3 and then assembles onto the connector/pRNA complex prior to packaging. It has been proposed that this portal complex represents a new class of rotary molecular motors, which couple rotation to DNA translocation[6].

Previously, bacteriophage DNA packaging has been studied in bulk reactions using DNase protection assays[4,5]. Here we present an assay in which force-measuring laser tweezers are used to follow packaging activity of a single complex in real time. Stalled, partly prepackaged complexes are attached to a polystyrene microsphere by means of the unpackaged end of the DNA (see Methods). This microsphere is captured in the optical trap and brought into contact with a second bead that is held by a pipette. This bead is coated with antibodies against the phage, so a stable tether is formed between the two beads (Fig. 1a). In the absence of ATP, the tether displays the elasticity expected for a single DNA molecule[8]. Shortly after addition of ATP, the two microspheres move closer together, indicating packaging activity. Control experiments confirm this conclusion: proheads bound to antibody-coated microspheres efficiently package biotinylated φ29 DNA in a bulk assay; no movement is observed in the measurements using optical tweezers when the proheads are omitted and the DNA is directly attached to the beads; and packaging could be reversibly stalled by a non-hydrolysable ATP analogue (γS-ATP).

The first set of measurements was carried out in 'constant force feedback' mode, in which the microsphere distance is adjusted by feedback to maintain DNA tension at a preset value of 5 pN (Fig. 1a). Figure 1b shows plots of the tether length versus time for a φ29-λ DNA construct with a length of 1.8 times the φ29 genome (see Methods). In all measurements a saturating concentration of ATP was used (0.5 mM). Packaging is highly efficient: in over 95% of the measurements, movements of several micrometres could be followed. Packaging dynamics are recorded with more detail, accuracy and time resolution than has previously been possible in bulk studies. Accounting for the prepackaging, we find that it takes 5.5 min on average to package a length of DNA equal to the φ29 genome, with a standard deviation from complex to complex of 0.8 min. A single complex also shows packaging rate fluctuations of ~20 bp s^{-1} r.m.s. in a 1 Hz bandwidth (Fig. 1c), which are approximately five times larger than the measurement noise of ~4 bp s^{-1} r.m.s., determined when the DNA alone is directly attached between two beads (see Supplementary Information Fig. I). Pauses in movement of variable duration (see Fig. 1b, inset) and slips (see below) are also clearly evident.

On average there are 3.1 pauses per micrometre of DNA packaged (σ = 3.5, 43 complexes) and these pauses have a mean duration of 4.0 s (σ = 5.1 s, n = 425). At higher capsid fillings the pausing frequency increases (correlation coefficient[9] of plots of pausing frequency versus capsid filling ~ 0.8), although the pause duration is not strongly correlated with filling (correlation coefficient of plots of pause duration versus capsid filling ~ 0.01). Details of the pause analysis are presented in Supplementary Information Figs II and III. During a pause, the DNA tether length remains constant despite the applied force, indicating that the motor stays engaged with the DNA. During packaging, however, some complexes occasionally exhibit abrupt increases in DNA tether length (see Fig. 2, inset). Apparently the motor loses grip on the DNA but can grasp the molecule again and resume packaging immediately. The mean length of a slip is 44 bp, with a large standard deviation of σ = 103 bp (n = 99). We have not found any evidence that slips occur at specific locations along the DNA, but they do occur more often at higher applied forces (see Supplementary Information Fig. IV). The motor is highly processive, in the sense that for the ensemble of records the length of DNA packaged divided by the number of slips is large (~7 μm, under a load of 5 pN).

In Fig. 1c the packaging rate is plotted against the percentage of the genome that has been packaged. These measurements show a marked reduction in packaging rate beginning when ~50% of the genome is packaged. Initially the rate is ~100 bp s^{-1} and it gradually drops to zero as the capsid fills up and the motor stalls. Note that the proheads are able to package up to ~5% more DNA than the normal φ29 genome length. These observations suggest that the rate decrease results from the build up of an internal pressure due to DNA confinement, which exerts an opposing force and slows the motor.

Within such a model, this internal force generated inside the capsid could be quantified if the effect of force on the velocity of the motor were known. The force dependence of the motor was investigated using a 'no feedback' mode where the positions of the trap and pipette are fixed (see Fig. 1a). These measurements were done using φ29-*Nco*I-A DNA (0.8× the φ29 genome; see Methods). In this mode the tension in the molecule increases as the motor reels in the DNA (Fig. 2a), and the bead is displaced from the centre of the trap. This force acts directly on the motor and at a certain point causes it to stall. These data allow us to determine how the packaging rate depends on applied force. Three such force–velocity (F–v) traces for different complexes are shown in Fig. 3a. The curve

shapes are similar, although the initial packaging rate and the stall force vary.

The mean F–v behaviour is obtained by first normalizing a number of individual curves to the mean stall force and mean initial velocity, and then averaging them together. Mean F–v curves were obtained at two points: when one-third and two-thirds of the genome is packaged (Fig. 3b, solid red and blue lines, respectively). Figure 1c suggests that in the case of one-third of the genome being packaged, no significant internal force opposes the motor because the velocity has not yet been reduced by DNA filling, although the motor does exhibit sensitivity to external force even at these low fillings (Fig. 3b). F–v data in the two-thirds case, however, suggest that an internal force of ~14 pN is acting on the motor, as it takes ~14 pN less external force to stall the motor in this case (Fig. 3b). Here we rely on the supposition that the internal and external forces add together to act on the motor. This supposition is supported by the good overlap of the F–v relationships over their entire range when the data for two-thirds genome packaging is shifted by +14 pN along the force axis. In this way, the inherent F–v relationship of the motor in the absence of any internal forces can be obtained.

The F–v curve in Fig. 3b shows that the packaging rate decreases even for small forces, indicating that the rate-limiting step in the motor's mechano-chemical cycle is force-dependent and therefore involves a mechanical displacement. Such behaviour is different from that reported for RNA polymerase[10], but is similar to that found in the bacterial flagellar motor[11]. Within a Kramer's model[12] of thermal activation over a single reaction barrier under the influence of force, our data imply that the rate-limiting step produces a conformational change of only ~0.1 nm (see Supple-

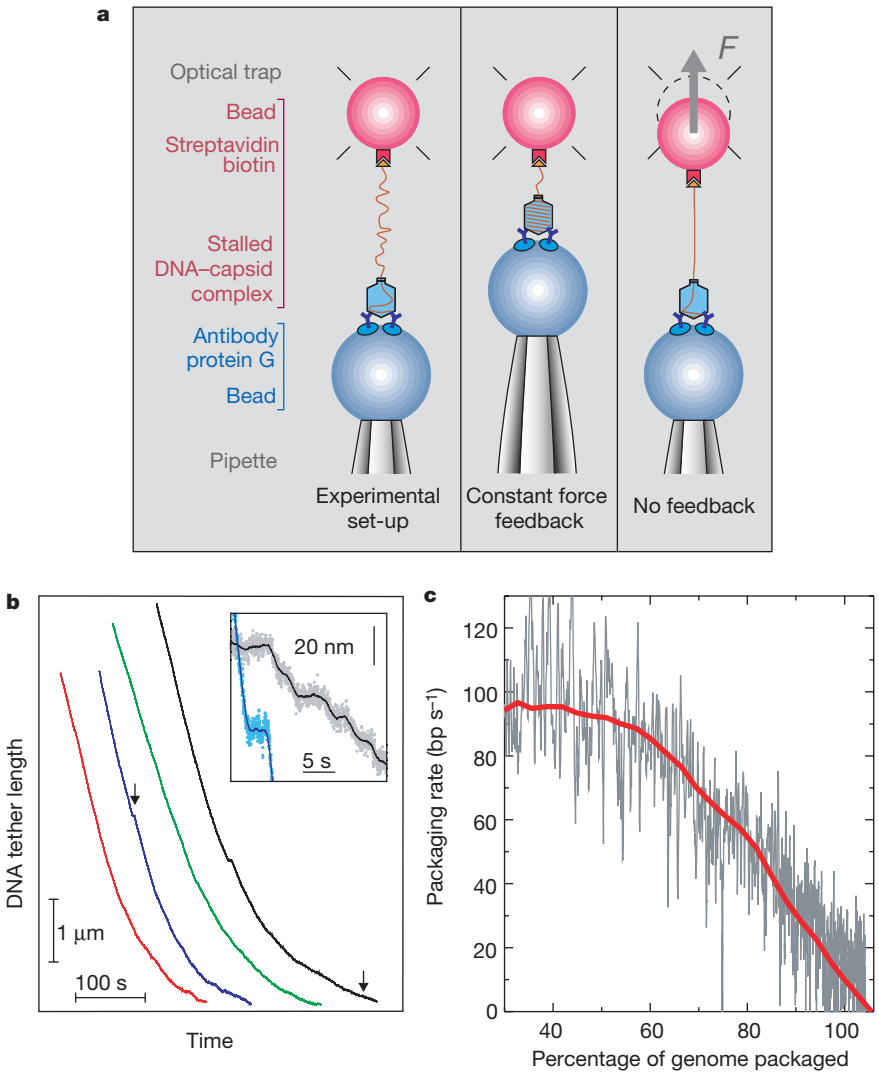

Figure 1 Set-up and initial results. **a**, Diagrams showing the experimental set-up at the start of a measurement (left), constant force feedback mode (middle) and no feedback mode (right) measurements. A single ϕ29 packaging complex is tethered between two microspheres. Optical tweezers are used to trap one microsphere and measure the forces acting on it, while the other microsphere is held by a micropipette. To insure measurement on only one complex, the density of complexes on the microsphere is adjusted so that only about one out of five–ten microspheres yielded hook-ups. Such attachments break in one discrete step as the force is increased, indicating that only one DNA molecule carries the load. **b**, DNA tether length against time for four different complexes with a constant force of ~5 pN using a 34.4-kb ϕ29-λ DNA construct. Inset, increased detail of regions, indicated by arrows, showing pauses (curves have been shifted). The solid lines are a 100-point average of the raw data. **c**, Packaging rate against the amount of DNA packaged, relative to the original 19.3-kb ϕ29 genome. Grey line, trace for a single complex (derived from black trace in **b**). Rates were obtained by linear fitting in a 1.5-s sliding window. The red line is an average of eight such measurements. Large pauses (velocity drops >30 bp s^{-1} below local average) were removed, and the curves were horizontally shifted to account for differences in microsphere attachment points. The red line was smoothed using a 200-nm sliding window. The standard deviation for the ensemble of measurements varies from ~20 bp s^{-1} at the beginning, down to ~10 bp s^{-1} at the end.

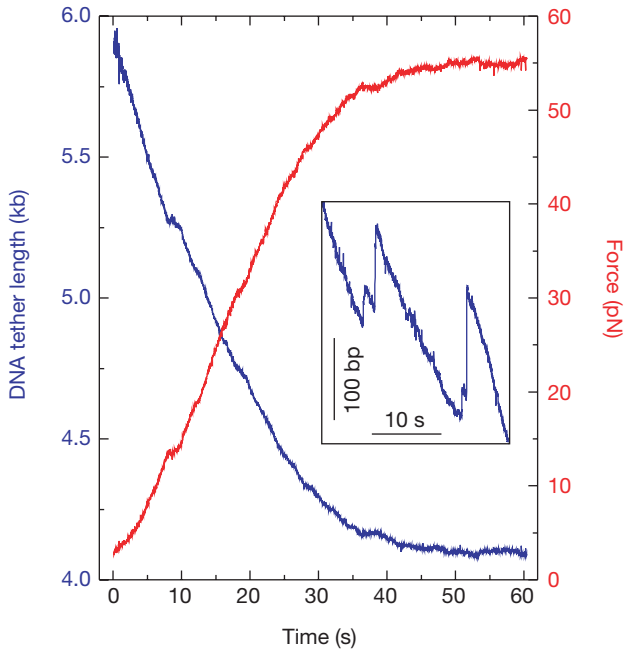

Figure 2 Measurements of packaging in no feedback mode. The tether shortens (blue line) and the force increases (red line) as the portal motor pulls in the DNA. In this example, the motor stalled at a force of ~55 pN. Although in many measurements the linkage broke before stalling was observed, similar maximum forces could be projected from these curves. We suspect that such breakage events occur at the antibody–prohead connection. Inset, spontaneous slipping events, where the DNA comes out of the capsid, were accompanied with abrupt decreases in force.

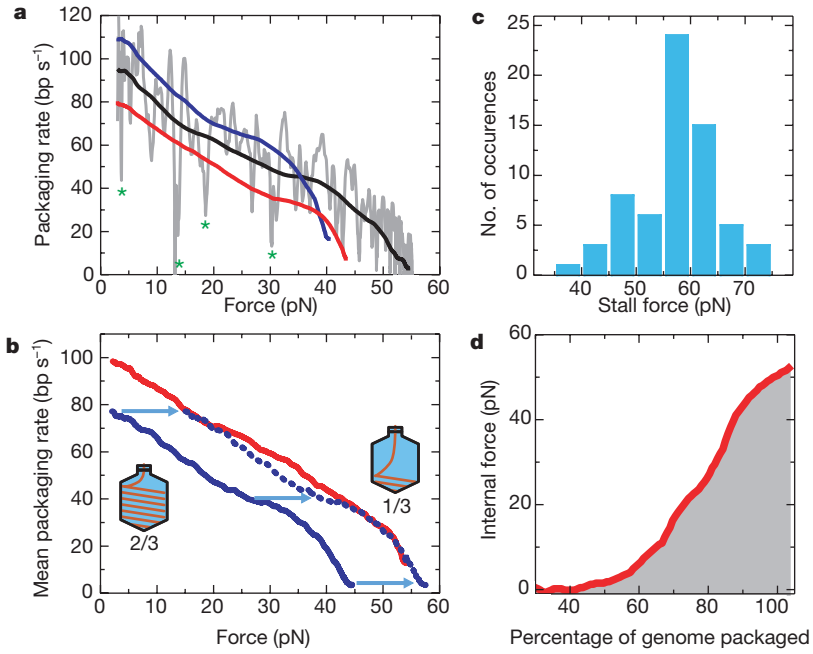

Figure 3 Force–velocity (*F*-*v*) analysis. **a**, The packaging rate for a single complex (grey line) was determined by linear fitting of the data in Fig. 2 in a 0.7-s sliding window. The black line is obtained by editing out large pauses (asterisks indicate where velocity drops >30 bp s^{-1} below local average) and smoothing (50-point sliding window). These long pauses were removed so as not to perturb the general trend of the *F*–*v* behaviour. The red and blue lines are data from two other complexes. **b**, External force against velocity curves when about one-third (red line) and about two-thirds (blue line) of the genome is packaged. Curves were obtained from averaging 14 and 8 individual traces, respectively. If, in the case of two-thirds of the genome being packaged, an additional 14 pN of internal force were acting on the motor (see text), the dashed blue line would show the expected behaviour. The red line and the dashed blue lines would then represent the inherent (total) *F*–*v* curve for the motor. **c**, Stall force measured for 65 individual complexes indicates an average stall force of ~57 pN. Stall force refers to the total force (external force plus, in the case of two-thirds of the genome being packaged, the inferred internal force of 14 pN), which is needed to stop further packaging. Clear plateaux (slope <5 bp s^{-1} for >5 s) in the force against time plots were sometimes observed (for example, in Fig. 2). The average force over this plateau is taken as the stall force. In some cases the packaging rate dropped by more than 80% (to below <20 bp s^{-1}), but the tether broke before a full plateau was reached. In these cases we extrapolated to obtain an estimate of the stall force, as the shapes of truncated curves were similar to those that fully stalled. DNA over-stretching was not observed because linkages always broke at external forces lower than 65 pN. However, stall forces above 65 pN could be determined in cases where internal force added to the total force. **d**, Internal force against percentage genome packaged. This plot is obtained by relating the packaging rate, as obtained from the rate against percentage of genome packaged curve (Fig. 1c), to the total force required to produce the same packaging rate, as given by the rate against force curve (**b**, average of red and dashed blue lines). To obtain the internal force, we subtracted the 5 pN of external force that is present in the experiment of Fig. 1c.

mentary Information Fig. V). This value is much smaller than the net movement of ~0.68 nm (2 bp) per ATP inferred from bulk studies[2]. Therefore, to the extent that this simple model applies, the rate-limiting step is sensitive to force, but is not the main translocation step. Note that the velocity begins to decrease more sharply with force at ~45 pN, suggesting that a second step in the mechanochemical cycle of the motor, associated with a larger movement, becomes rate-limiting at these higher forces.

Figure 3c shows a histogram of the total force needed to stall the motor. The stall force ranges between 40 and 70 pN and has an average force of ~57 pN, making ϕ29 one of the strongest molecular motors reported to date. This stall force is approximately eight times higher than that reported for conventional kinesin[13,14] or skeletal muscle myosin II motors[15], and approximately two times higher than that for RNA polymerase[10]. By multiplying the average stall force (57 pN) with the distance moved per ATP[2] (~0.68 nm), we obtain a work done per ATP of ~39 pN nm. Assuming the free energy of ATP hydrolysis in our buffer[16] is ~ 120 pN nm, then the energy conversion efficiency is ~30%. This value falls within the range of efficiencies reported for myosin[15], kinesin[13,14] and RNA polymerase motors[10]. Additionally, a maximum force of 70 pN implies that the motor's step size must be less than 5 bp (120 pN nm/70 pN ≈ 1.7 nm).

Combining the rate dependence on external force (Fig. 3b) and on the fraction of DNA packaged (Fig. 1c), we can make a quantitative estimate of the build-up of internal force as the prohead is filled. Figure 3d shows that when the full genome has been packaged, this internal force reaches a large value of about 50 pN. In many phages it is believed that the DNA is ordered as a spool in which the strands are packaged in a hexagonal lattice. By dividing the measured force by the hexagonal cell surface area, as determined by X-ray diffraction studies[17], one may obtain a rough estimate of a pressure of 6 MPa inside the phage capsid, assuming that there is no significant energy dissipation in packaging. An osmotic pressure of a similar order (~1 MPa) has been reported to be necessary to condense DNA in solution to a density similar to that in the phage capsid[18]. If 6 MPa of pressure is transmitted to the capsid shell, the shell (thickness ~1.6 nm[1]) must have a tensile strength of at least 100 MPa, a value similar to the bulk tensile strength for a typical aluminum alloy.

As seen in Fig. 3d, the force starts building only after about half of the ϕ29 genome is already packaged, a trend that has recently been seen also in simulations of DNA packaging into phage capsids (J. Kindt, S. Tzlil, A. Ben-Shaul and W. M. Gelbart, personal communication). Our finding suggests that the DNA does not immediately adopt its final condensed state inside the capsid, but instead is progressively compressed. Furthermore, both the pause frequency and the internal pressure increase with DNA filling, indicating that internal pressure may have a function in inducing pauses.

The compression of DNA in phage capsids has long been considered an intriguing phenomenon because it is thought to involve large energetic penalties[19,20]. An estimate of the total work done in packaging the ϕ29 genome is obtained by integrating the force curve in Fig. 3d, which yields 7.5×10^{-17} J (or 2×10^4 kT). This value can be compared with theoretical predictions for the equilibrium free energy change of the DNA. One model[19] gives 4×10^4 kT, whereas a preliminary estimate (T. Odijk, personal communication) based on another model[20] yields ~5.6×10^3 kT. These predictions, which include electrostatic, bending and entropic penalties, but no dissipative effects, are of the same order of magnitude as our estimate of the total work done. Thus it is possible that a large fraction of the work done by the motor is reversible work, and that dissipation is not dominant. Moreover, the observations of rapid slipping of the DNA out of the capsid on a timescale below 0.014 s with no evidence of slower relaxation also suggest that there is little viscous resistance to DNA movement within the capsid.

Whereas it has been argued that T7 phage relies on a motor for DNA injection into the host[21], other phages are known to passively eject their genome when exposed to the bacterial membrane receptor[17,22]. Our results lend support to a model for ϕ29 in which internal pressure provides the driving force for DNA injection into the host cell for the first half of the injection process. □

Methods

Stalled packaging complexes

Components of the ϕ29 *in vitro* packaging assay, including proheads, DNA-gp3 and gp16, were purified as described previously[23]. The 19.3-kb ϕ29 DNA-gp3 (6.6 μm in length) was cut with *Nco*I to yield a 15-kb left end (ϕ29-*Nco*I-A DNA) and a 4.3-kb right end. The left end is preferentially packaged[5,24]. These fragments were biotinylated using the Klenow fragment of DNA polymerase I (exo$^-$ mutant; New England Biolabs) to incorporate biotin-14-dATP and biotin-14-dCTP (Gibco). We constructed the longer 34.4-kb ϕ29-λ DNA by ligating together the ϕ29 and λ *Nco*I-A fragments. Before cutting, the cos sites (cohesive ends) of λ-DNA were biotinylated as described above. To initiate a packaging reaction we added 4 μl of 0.25 mM ATP (Roche) to a 13.5 μl solution of buffer A (50 mM Tris-HCl buffer (pH 7.8), 50 mM NaCl, 5 mM $MgCl_2$), 0.1 μg of biotinylated DNA-gp3, 0.5 μg gp16 and 10 μg ϕ29 proheads. This mixture is incubated for approximately 30 s, and then the packaging is stalled by adding 4 μl of 2.5 mM γS-ATP (Roche), such that the genome is partly prepackaged. These complexes are stable for >10 h and are reactivated during the measurement by exposing them to buffer A plus 0.5 mM ATP, 5 μM ADP, 5 μM NaH_2PO_4 (source of phosphate ions) and 50 μg ml^{-1} BSA. This ATP concentration is saturating (increasing it by a factor of ten does not significantly increase the packaging rate, that is, the reaction is not limited by diffusion).

Microsphere preparation

Polystyrene microspheres that were coated with streptavidin (2.2 μm diameter, 0.5% w/v; Spherotech) were washed twice and pre-blocked for 5 min with 5 mg ml^{-1} BSA in buffer A. Stalled complexes were attached to the microspheres (by means of the biotinylated DNA) by adding approximately 0.3 μl of the stalled reaction mixture and 1 μl RNase inhibitor (SUPERase-IN, Ambion) to 20 μl of microspheres, and incubating them for 15–20 min. We then diluted the microspheres in 1 ml buffer A. Polystyrene microspheres coated with protein G (2.8 μm diameter, 5% w/v; Spherotech) were washed twice in phosphate-buffered saline and incubated for >20 min with a ~1/10 dilution of rabbit antisera prepared against ϕ29. We added 5 μl of microspheres (washed five times) and 1 μl of RNase inhibitor to 1 ml of buffer A.

Optical tweezers

We used a dual-beam optical tweezers apparatus similar to that described previously[8]. The measured trap stiffness is ~0.12 pN nm^{-1} and the force is recorded at a rate of ~70 Hz. The displacements of the trapped microsphere were derived from the measured forces. The pulling angle could be varied over a 120° range but did not seem to affect the packaging rate. The pipette position is recorded using an optical lever having a resolution of ~1 nm. The contour length of the DNA (in bp) was determined from the measured force and end-to-end distance, and through using the worm-like chain model assuming a persistence length of 53 nm, a stretch modulus of 1,200 pN and a distance per bp of 0.34 nm[25].

Received 20 March; accepted 24 August 2001.

1. Tao, Y. *et al.* Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95,** 431–437 (1998).
2. Guo, P., Peterson, C. & Anderson, D. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage ϕ29. *J. Mol. Biol.* **197,** 229–236 (1987).
3. Anderson, D. & Reilly, B. in *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics* (eds Sonenshein, A., Hoch, J. A. & Losick, R.) 859–867 (American Society for Microbiology, Washington DC, 1993).
4. Guo, P., Grimes, S. & Anderson, D. A defined system for *in vitro* packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage ϕ29. *Proc. Natl Acad. Sci. USA* **83,** 3505–3509 (1986).
5. Grimes, S. & Anderson, D. *In vitro* packaging of bacteriophage ϕ29 DNA restriction fragments and the role of the terminal protein gp3. *J. Mol. Biol.* **209,** 91–100 (1989).
6. Simpson, A. A. *et al.* Structure of the bacteriophage ϕ29 DNA packaging motor. *Nature* **408,** 745–750 (2000).
7. Ibarra, B. *et al.* Topology of the components of the DNA packaging machinery in the phage ϕ29 prohead. *J. Mol. Biol.* **298,** 807–815 (2000).
8. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271,** 795–799 (1996).
9. Barlow, R. *Statistics: a Guide to the Use of Statistical Methods in the Physical Sciences* 15–16 (Wiley, Chichester, 1989).
10. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282,** 902–907 (1998).
11. Berg, H. C. & Turner, L. Torque generated by the flagellar motor *of Escherichia coli*. *Biophys. J.* **65,** 2201–2216 (1993).
12. Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7,** 284–304 (1940).
13. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77,** 773–784 (1994).

14. Coppin, C. M., Pierce, D. W., Hsu, L. & Vale, R. D. The load dependence of kinesin's mechanical cycle. *Proc. Natl Acad. Sci. USA* **94,** 8539–8544 (1997).
15. Finer, J. T., Simmons, R. M. & Spudich, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368,** 113–119 (1994).
16. Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Principles of Biochemistry* 375 (Worth, New York, 1993).
17. Earnshaw, W. C. & Casjens, S. R. DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21,** 319–331 (1980).
18. Rau, D. C., Lee, B. & Parsegian, V. A. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. *Proc. Natl Acad. Sci. USA* **81,** 2621–2625 (1984).
19. Riemer, S. C. & Bloomfield, V. A. Packaging of DNA in bacteriophage heads: some considerations on energetics. *Biopolymers* **17,** 785–794 (1978).
20. Odijk, T. Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophys. J.* **75,** 1223–1227 (1998).
21. Garcia, L. R. & Molineux, I. J. Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. *J. Bacteriol.* **178,** 6921–6929 (1996).
22. Novick, S. L. & Baldeschwieler, J. D. Fluorescence measurement of the kinetics of DNA injection by bacteriophage lambda into liposomes. *Biochemistry* **27,** 7919–7924 (1988).
23. Grimes, S. & Anderson, D. The bacteriophage ϕ29 packaging proteins supercoil the DNA ends. *J. Mol. Biol.* **266,** 901–914 (1997).
24. Bjornsti, M. A., Reilly, B. E. & Anderson, D. L. Morphogenesis of bacteriophage ϕ29 of *Bacillus subtilis*: oriented and quantized *in vitro* packaging of DNA-gp3. *J. Virol.* **45,** 383–396 (1983).
25. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl Acad. Sci. USA* **94,** 6185–6190 (1997).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank W. M. Gelbart, P. Jardine, T. Odijk, V. Bloomfield, D. Frenkel, C. Varga, A. Mehta and M. Young for comments. This research was supported in part by grants from the NIH, DOE and NSF. D.E.S. and S.J.T. are supported by a grant from the Packard Foundation. S.J.T. is supported by the Netherlands Organization for Scientific Research (NWO).

Correspondence and requests for materials should be addressed to C.B. (e-mail: carlos@alice.berkeley.edu).

Uracil-DNA glycosylase acts by substrate autocatalysis

Aaron R. Dinner*†, G. Michael Blackburn‡ & Martin Karplus*§‖

**Central Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QH, UK*
‡ Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK
§ Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
‖ Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 4 Rue Blaise Pascal, 67000 Strasbourg, France

In humans, uracil appears in DNA at the rate of several hundred bases per cell each day as a result of misincorporation of deoxyuridine (dU) or deamination of cytosine. Four enzymes that catalyse the hydrolysis of the glycosylic bond of dU in DNA to yield an apyridiminic site as the first step in base excision repair have been identified in the human genome[1]. The most efficient and well characterized of these uracil-DNA glycosylases is UDG (also known as UNG and present in almost all known organisms)[2], which excises U from single- or double-stranded DNA and is associated with DNA replication forks[3]. We used a hybrid quantum-mechanical/molecular-mechanical (QM/MM) approach[4] to determine the mechanism of catalysis by UDG. In contrast to the concerted associative mechanism proposed initially[5–10], we show here that the reaction proceeds in a stepwise dissociative manner[11,12]. Cleavage of the glycosylic bond yields an intermediate comprising an oxocarbenium cation and a uracilate anion. Subsequent attack by a water molecule and transfer of a proton to D145 result in the products. Surprisingly, the primary contribution to lowering the activation energy comes from the substrate, rather than from the enzyme. This 'autocatalysis' derives from the burial and positioning of four phosphate groups that stabilize the rate-determining transition state. The importance of these phosphates explains the residual activity observed for mutants that lack key residues[6–9]. A corresponding catalytic mechanism could apply to the DNA glycosylases TDG and SMUG1, which belong to the same structural superfamily as UDG[13,14].

In the QM/MM simulations, the reactive region was treated with the AM1 method[15] (and higher-level techniques, as described in Methods), while the surrounding non-reactive region was modelled classically with the CHARMM force field[4,16]. Because the enzyme–substrate complex (described below) has many charged groups, solvent shielding is essential, and we adapted a procedure developed for MM free energy simulations[17] to QM/MM calculations (see Methods). The charges of exposed groups were scaled to avoid distortion of the structures during the simulations, and continuum electrostatics methods were applied subsequently to obtain energies that include the correct solvent-shielding effects.

The starting point for the simulations was the structure of an enzyme–inhibitor complex at 1.8 Å resolution[11]. This complex contains the non-hydrolysable C-glycoside deoxypseudouridine (dΨU) in place of the dU substrate (see Supplementary Information Fig. 1). The calculations include the entire recombinant protein (223 residues, indexed according to their positions in the *UNG* gene: 82–304)[18], the bound double-stranded DNA oligonucleotide and 168 crystallographic water molecules. One nucleic acid strand has nine nucleotides (indexed 2–10 with dΨU at 5), whereas the complementary strand has ten (indexed 21–30). Similarities between the crystal structure and the rate-limiting transition state, determined below, suggest that the bound ligand is a transition state analogue. We generated a reactant structure by relaxing the system by energy minimization after interchange of the carbon and nitrogen in positions 1 and 5 of the inhibitor to yield the natural substrate dU. The heavy atom r.m.s. deviation between the minimized and crystal structures was 0.7 Å. The glycosylic bond relaxes from more than 40° to about 25° out of the plane of the pyrimidine base (a superposition of the important protein residues of the X-ray and the minimized structure is shown in Supplementary Information). A natural bond-orbital analysis[19] of model compounds suggests that conformational strain and stereoelectronic effects contribute at most 5 kcal mol^{-1} to catalysis at the calculated angle for the substrate (Supplementary Information).

Two-dimensional adiabatic potential energy surfaces, as functions of the lengths of the bonds that are made and broken (Fig. 1), were calculated by minimizing the energy of the system in the presence of harmonic constraints on those bonds. Figure 1b shows the solvent-corrected energy in terms of the distance between the atoms in the glycosylic bond ($r_{C1'-N1}$) and that between C1′ and the oxygen atom of the attacking water molecule ($r_{C1'-OH_2}$). A transition state (TS_1) with energy $E = 14.9$ kcal mol^{-1} (all energies are relative to the reactants) at $r_{C1'-N1} = 2.05$ Å and $r_{C1'-OH_2} = 3.10$ Å (Fig. 2) leads to a well defined intermediate (I_1), with $E = 7.4$ kcal mol^{-1} at $r_{C1'-N1} = 2.75$ Å and $r_{C1'-OH_2} = 2.85$ Å. The presence of the broad minimum at $r_{C1'-N1} = 2.75$ Å and $r_{C1'-OH_2} = 2.85$ Å demonstrates that the mechanism is stepwise and that the glycosylic bond breaks before the water molecule attacks. The leaving group is anionic because there is no effective proton donor in the surrounding region if H268 is neutral, as indicated by NMR[8]. The second transition state in Fig. 1b (TS_2, $E = 10.3$ kcal mol^{-1} at $r_{C1'-N1} = 2.95$ Å and $r_{C1'-OH_2} = 1.95$ Å) corresponds to the attack by the water molecule. The shoulder at $r_{C1'-OH_2} = 1.65$ Å indicates the point at which transfer of a proton to D145 becomes energetically favourable. An analogous

† Present address: Department of Chemistry, University of California, Berkeley, California 94720, USA.