**Polymer sequencing by molecular machines: A framework for predicting the resolving power of a sliding contact force spectroscopy sequencing method.**

*Alex Dunlopa, Kate Bowmanb, Olav Aarstadc, Gudmund Skjåk-Brækc, Bjørn T. Stokked, Andrew N. Roundb\**

(a) Dr. A. Dunlop

HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

(b) Dr. K. Bowman, Dr. A. N. Round (\*corresponding author; e-mail: a.round@uea.ac.uk)

School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

(c) Dr. O. A. Aarstad, Prof. G. Skjåk-Bræk

Department of Biotechnology, The Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

(d) Prof. B. T. Stokke

Department of Physics, The Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

Keywords: molecular machines, single molecule forcespectroscopy, nanopores, polymers, sequencing, rotaxanes, glycosylation, post-translational modification

**Abstract**

We evaluate an AFM-based single molecule force spectroscopy method for mapping sequences in otherwise difficult to sequence heteropolymers, including glycosylated proteins and glycans. The sliding contact force spectroscopy (SCFS) method exploits a sliding contact made between a nanopore threaded over a polymer axle and an AFM probe. We find that for sliding α- and β- cyclodextrin nanopores over a wide range of hydrophilic monomers, the free energy of sliding is proportional to the sum of two dimensionless, easily calculable parameters representing the relative partitioning of the monomer inside the nanopore or in the aqueous phase, and the friction arising from sliding the nanopore over the monomer. Using this relationship we calculate sliding energies for nucleic acids, amino acids, glycan and synthetic monomers and predict on the basis of these calculations that SCFS will detect N- and O-glycosylation of proteins and patterns of sidechains in glycans. For these applications, SCFS offers an alternative to sequence mapping by mass spectrometry or newly-emerging nanopore technologies that may be easily implemented using a standard AFM.

Introduction

While the sequencing of DNA is now routine, with more rapid and more accurate approaches under constant development, a method for sequencing long stretches of other polymers, whether naturally occurring (such as polysaccharides) or synthetic, does not exist. There is a pressing, unmet need for a polysaccharide sequence mapping tool, since these polymers lack a canonical sequence and instead the pattern of monomer and branching sequence depends on several factors, determined by cellular metabolism, developmental stage, nutrient availability and others1. Even in polymers with canonical sequences, such as proteins and nucleic acids, post-translational modification of proteins by glycosylation, phosphorylation and other additions2,3, as well as epigenetic modifications of nucleic acids4, occur in micro- and macroheterogeneous patterns that are not always easy to discern. Recently, the emergence of sequence-defined synthetic polymers has created a new class of materials in which relevant properties may be controlled with much greater precision than before, where sequence directly controls function.5-7 Previously the authors8,9 and others10 explored the feasibility of a new route, here called sliding contact force spectroscopy (SCFS), to obtaining sequence information in linear heteropolymers by atomic force microscopy (AFM), but the limits and applicability of the method have not been fully explored. Here we set out to describe the parameters that determine how easy or difficult it is to pass a cyclodextrin (CD)-based macrocycle over a particular monomer, and from that basis derive a framework within which we can predict whether the monomers in a particular copolymer are distinguishable using this method.

The SCFS method uses a cyclodextrin (CD)-based macrocycle tethered to the AFM probe, with the polymer to be interrogated tethered likewise to another surface and induced to form a host-guest complex with the macrocycle to form a polyrotaxane11 or molecular ring-thread complex. Rotaxanes are examples of a broad group of supramolecular complexes that can be induced to do mechanochemical work, and which can be described as ‘molecular machines’12. This rapidly expanding group includes molecular ‘walkers’13, shuttles and switches that can in some cases do work against significant external loads14. The SCFS experiment has parallels with the nanopore sequencing approach that is under continuous development as a DNA sequencing tool15,16 and which has recently been shown to discriminate between different poly(ethylene glycol) (PEG) polymers on the basis of their molecular weight, with monomer resolution.17 In particular, AFM has been used to measure the forces acting on ssDNA as it slides by either a “frictionless” or a “stick-slip” mode in a nanopore18. The most common terminology19 for describing the processes occurring in a rotaxane depicts a macrocyclic ‘bead’ shuttling between ‘stations’ on the polymer axle. In the SCFS method described here and previously,8-10 the bead is α- or β-cyclodextrin (α-CD and β-CD hereafter) and the stations are the individual monomers comprising the polymer axle, while the AFM probe supplies the unidirectional driving force for shuttling the bead between stations (hereafter ‘sliding’) under a load generated by the controlled separation of probe and sample. Figure 1 illustrates the parallels between a conventional rotaxane system and the assembly constructed for SCFS. The concept of manipulating a rotaxane using a local force probe has been explored before: Komiyama and coworkers20 used STM (Scanning Tunnelling Microscopy) to manipulate α-CD beads forming a polyrotaxane with poly(ethylene glycol) (PEG) back and forth along the PEG axle, while Stoddart et al21 and Leigh and Duwez14 have used AFM to measure the force required to drive a bead between two stations in a rotaxane. None of these works addressed the use of a sliding contact between a bead and a polymer as a sequencing tool. Previous analyses of the challenges to polymer sequencing by single molecule force spectroscopy (SMFS), with or without a sliding contact, have focused on DNA sequencing.22,23

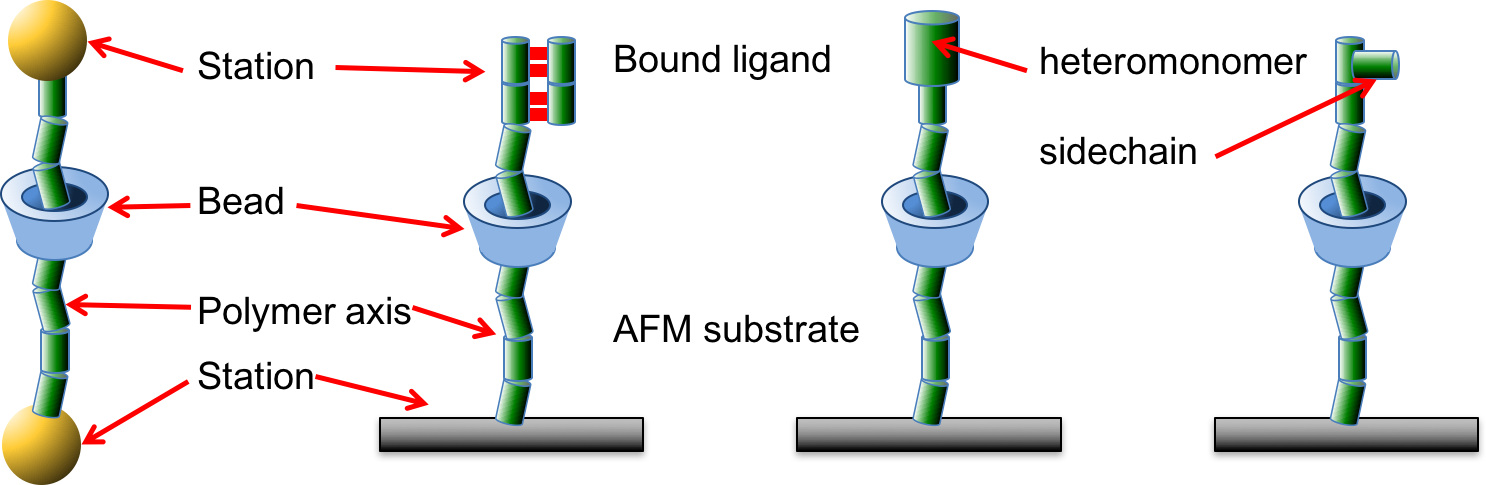
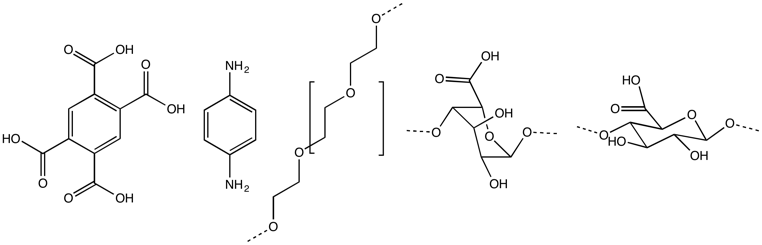
(a) (b) (c) (d) 

Figure 1. Schematic illustration highlighting the parallels between a conventional rotaxane system (a) and various iterations of the sliding contact pseudorotaxane (b-d). The common features (station, bead, axis) are labelled in each, along with examples of the monomer features that may constitute a ‘station’ in the sliding contact experiment: a bound ligand (b), a different monomer (c) or a sidechain (d).

We have shown previously that measurements made with the SCFS approach described here yielded excellent agreement with the predicted positions of aromatic rings substituted into PEG polymers based upon the measured molecular weights of the polymers8, and that the CD bead could be used to unzip interactions between the polymer axle and molecules bound to specific sequences within that polymer.9 Thus the available evidence suggests that SCFS may offer a method for mapping or sequencing long, linear polymers where there are large differences between monomers or blocks, or where specific sequences are recognised by other molecules. However, the limits on the size and nature of the different monomer stations and macrocyclic beads for which differences in force may be distinguished remain undefined.

In the present work we compare the forces measured during the forced sliding of α- and β-CD beads along PEG-based polymers possessing one or more of 4 different stations representative of 2 classes of monomer: aromatic rings and glycans. As well as representing common polymers, these groups encompass a wide range of molecular cross-sectional areas, aqueous solubilities and affinities for complexation with α- and β-CDs. We apply the Friddle-Noy-de Yoreo (F-N-Y)24,25 method for analysing single molecule force spectra in order to extract the energies involved in the bead-station interactions and consider the parameters that have predictive power in determining the resolution of the SCFS sequencing approach. Finally, we consider the potential and the limits of the method for sequencing common polymers.

Results and discussion

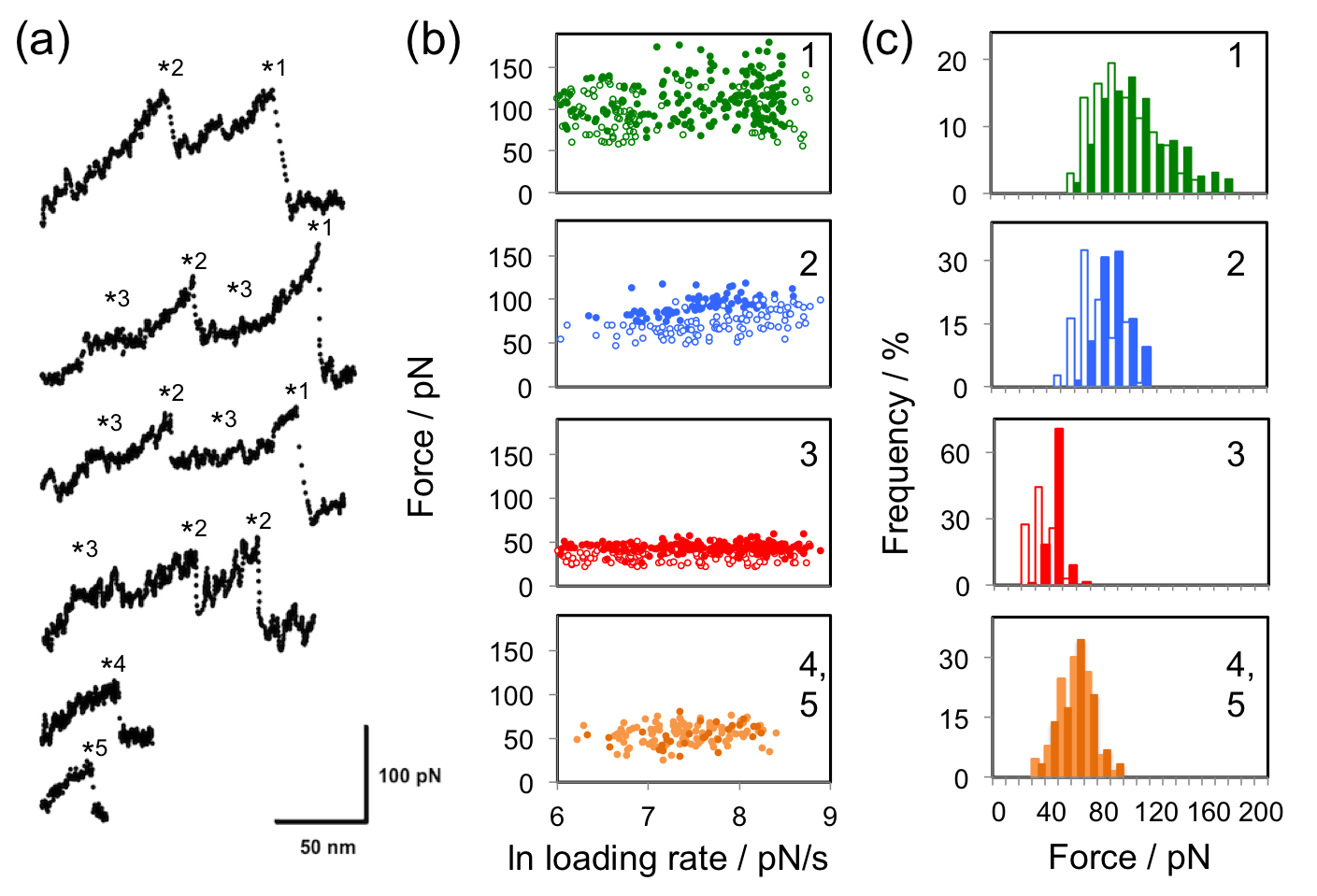


**1. 2. 3. 4. 5.**

**Figure 2.** The monomers used as stations in this study: pyromellitic acid (1), aminoaniline (2), (poly)ethylene glycol (3), guluronic acid (4) and mannuronic acid (5).

Analysing sliding contact force spectra

In addition to previously published8,9 datasets using α-CD and stations 1, 2, 4 and 5 on PEG axles (pyromellitic acid, aminoaniline, guluronic and mannuronic acid respectively; see **Figure 2** for structures), we have conducted SCFS experiments using β-CD beads in order to probe the effect of differences in pore size. Recently Friddle, Noy and De Yoreo introduced a new model24,25 for analysing single molecule force spectroscopy experiments that describes both the near-equilibrium (at low loading rates) and far-from-equilibrium (at high loading rates) regimes of the dynamic force spectrum (plot of most probable rupture force vs. instantaneous loading rate at rupture). The model has been shown to apply to interactions between ligands and receptors, small molecules and bulk surfaces.24 In common with the established Bell-Evans model,26 the method is used to extract the parameters koff and xβ, the intrinsic unbinding rate of the bond and the distance to the transition state, from the force spectrum. In cases where the near-equilibrium regime is reached, a third parameter, the equilibrium force *f*eq (the minimum force required to move the binding pair apart by the distance xβ, beyond which they can no longer instantaneously rebind) may be obtained and from it ΔGbu, the equilibrium unbinding free energy, for the bond. The term ‘equilibrium’ here is used in the sense used by Friddle et al24,25 and denotes a process that is occurring rapidly in both forward and reverse direction with respect to the travel of the AFM probe and the CD bead attached to it. Here we treat the process of shuttling (sliding a bead over a station in a polymer) in the same way as breaking a conventional ligand-receptor bond and so we use the values of *f*eq we have recorded to calculate ΔGsl, the sliding free energy, in analogy with the ΔGbu term described above. Our justification for taking this approach lies in the common features of both processes: the elastic polymer tethers will act as entropic springs at low forces and undergo enthalpic bond stretching at higher forces until the tension is released, either by breaking a bond or by forcing the bead to slide over the monomer station. Before applying the model we follow Akhremitchev’s method27 of using the fitted Kuhn lengths to distinguish between single and multiple polymer stretches, selecting only single polymer stretches (those with Kuhn lengths equal to or greater than the Kuhn length of a single PEG chain) for further analysis. This approach was recently applied to the crosslinking of DNA by intercalators.28 The analysis of the data is presented in supporting information.



**Figure 3.** (a) Example force curves for the interactions investigated here: from the bottom, the first two curves were collected when sliding α-CD over oligoguluronic and mannuronic acids (stations 4 and 5); the remaining curves were collected when (top two curves) α-CD and (middle two curves) β-CD were pulled over a polymer consisting of PEG (station 3) and individual monomers of aminoaniline (station 2) and pyromellitic acid (station 1). Asterisks mark the rupture points at which forces and loading rates are measured. For more details of the polymer characterisation see reference [1]. (b) Dynamic force spectra for the interactions of α-CD (filled circles) and β-CD (open circles) with each of the five stations. For stations 4 and 5, all data is for α-CD; lighter symbols are for guluronic acid and darker symbols for mannuronic acid. (c) Histograms of the most probable sliding force, equivalent to feq, the equilibrium force, for the five stations. Colours of bars follow those described for symbols in part (b).

Table 1. Values of calculated and measured parameters (log K, the binding constant; feq, the equilibrium force; kc, the cantilever spring constant; and ΔGsl, the free energy of sliding) for interactions of stations 1 to 5 with α- and β-CD.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Station** | **Bead** | **Log K, literature values** | **Log K, UV values [a]** | **feq ± sd (pN)** | **kc (pN/nm)** | **ΔGsl ± sd (kJ/mol)** |
| 1 | α-CD | - | 1.4 | 98 ± 8 | 25 | 159 ± 17 |
| β-CD | - | 2.81 | 86 ± 12 | 22 | 134 ± 26 |
| 2 | α-CD | 1.41 [23] | 1.47 | 76 ± 6 | 25 | 95 ± 13 |
| β-CD | 2.05 [23] | 1.74 | 63 ± 6 | 22 | 72 ± 14 |
| 3 | α-CD | 2.51 - 4.43 [24] | - | 29 ± 4 ; 37 ± 3 | 14 ; 25 | 22 ± 4 ; 23 ± 4 |
| β-CD | 3.03 - 3.17 [24] | - | 34 ± 5 | 22 | 21 ± 6 |
| 4 [b] | α-CD | 1.56 [25] | - | 45 ± 5 | 14 | 54 ± 12 |
| 5 [b] | α-CD | 1.56 [25] | - | 46 ± 7 | 16 | 50 ± 15 |

a) values for K were calculated from UV measurements using the Benesi-Hildebrand method; b) For stations 4 and 5, the value of K is the mean of values for several pyranoses. Details of method for the calculation of log K are available in supporting information.

Figure 3 shows examples of force curves, the dynamic force spectrum and histogram of forces for each dataset, while the values of equilibrium force *f*eq and free energy of sliding ΔGsl for the interactions studied are presented in Table 1. The equilibrium forces for sliding over the stations ranged from 29 to 98 pN and are always larger for the α-CD interaction than for the β-CD interaction. They occupy a comparable range to that predicted and observed for single molecule ligand-receptor unbinding events24 and intercalation into DNA.28 Correspondingly, the energies calculated for the sliding interactions (ΔGsl) are found to range from 20 to 160 kJ/mol; equivalent to between approximately 1 and 8 hydrogen bonds (free energy of hydrogen bond in water = 23.3 kJ/mol)29,30. In subsequent analyses presented below we use the free energy of sliding ΔGsl rather than equilibrium force *f*eq since the latter quantity is dependent on the spring constant of the particular cantilever used,24 making direct comparison of values obtained with different cantilevers more difficult. ΔGsl is calculated from the value of *f*eq and the spring constant of the cantilever used and is therefore directly comparable across different experiments. However, direct comparison with literature data can only be made when the spring constants of the cantilevers used for each specific data set are reported. As an example, the range of magnitudes of the forces observed in the present study is consistent with those found for the unbinding of host-guest complexes between β-CD and a range of aromatic groups,31 although when the equilibrium forces and spring constants reported in that work are used to calculate ΔGbu by Friddle and Noy’s method the values range between 13 and 87 kJ/mol (see supporting information). The discrepancy in values of ΔGbu calculated from the data in [22] and the ΔGsl values obtained in the present work highlights the distinction between the dissociation of equilibrated host-guest complexes and the forced sliding of the host cyclodextrin ring over the guest monomers.

Additional comparison can be made with the values predicted and observed for the sliding of a β-CD bead over single- stranded DNA: Lindsay and Williams23 predicted that the force required to drive a β-CD bead along a single strand of DNA was 75-78 pN (corresponding to 31-33 kJ/mol using the F-N-Y relation between *f*eq and ΔGbu and a reported spring constant of 0.3 N/m, see Table S2 in supporting information), depending on whether the base passed over was a purine or a pyrimidine. Therefore, no distinguishing force signature between purine and pyrimidine nucleotides would be detected above instrumental noise (~15 pN or more). These authors proposed that this rather low value was due to the mobility of the bases around their point of attachment to the (deoxy)ribose backbone, allowing them to fold flat against the phosphate-deoxyribose backbone to pass through the CD pore. The same group subsequently published experimental data10 showing force plateaus for the sliding of β-CD along DNA somewhat larger than this value, at approximately 125 pN. We report here most probable sliding forces of 63-98 pN for differently-substituted aromatic groups and 45-46 pN for a monosaccharide (passing through α-CD), at comparable or higher loading rates (instantaneous loading rates from 400 to 8000 pN/s).

**Relationship between complexation constant and sliding energy**

We sought to discern the parameters that might be used to predict the sliding free energy for passing a CD bead over a particular monomer, and with the example of Auletta et al31 in mind, started by considering whether the sliding of the CD over the monomer reflects the well-known host-guest (hereafter H-G) complexation interaction. Here the guest (monomer) forms a complex by penetrating into the pore of the host (CD). Using the data provided in that work, we calculated values of ΔGbu from the most probable rupture forces and spring constants reported, and find a positive dependence of ΔGbu on log K, where K is the binding constant for the H-G complex, derived from ΔGº as measured by Auletta et al using ITC (**Figure 4a**). We then compared values of the same binding constant K for each of the monomers used in the present work, (presented in **Table 1**, and derived from literature reports32-35, and/or measured by a UV spectroscopic method, see **supporting information**), to ΔGsl, and find no clear relationship between ΔGsl and log K, as depicted in **Figure 4b**. The lack of dependence observed in our SCFS data, in contrast to the clear relationship observable in Auletta et al’s data, reflects the distinction between H-G complexation and forced threading of the CD over the monomer in SCFS: in H-G complexation the geometry that favours the lowest (kinetically accessible) energy state for the guest in the host may not involve complete inclusion of that guest in the host; while in SCFS the monomer is forced to pass through the CD pore, driving the complex through energetically unfavourable transition states that constitute the largest energy barriers to the passage of the monomer through the pore. This may be expected to be most relevant when the monomer is large and rigid, as would be the case for substituted aromatic groups. We sketch the differing mechanisms and resulting energy pathways in **Figure 4c**. As an illustrative example, we consider that the ΔGº of rupturing the H-G complex between aniline and β-CD, when the aniline guest has adopted the most energetically favourable configuration, was measured to be 2.3 kcal/mol (9.6 kJ/mol) by ITC (and ΔGbu calculated to be 12.7 kJ/mol using that work’s data and applying the F-N-Y formalism; see **Table S2** in **supporting information**), while we find a ΔGsl for sliding β-CD from a PEO chain over the very similar station p-aminoaniline and on to a subsequent PEO chain to be much higher at 72 kJ/mol (see **Table 1**). Thus, since the binding constants measured or calculated for H-G complexes will not necessarily reflect the main energy barrier to sliding the same CD host over the same monomer guest, we reject using H-G binding constants as a basis for predicting the expected sliding energy.

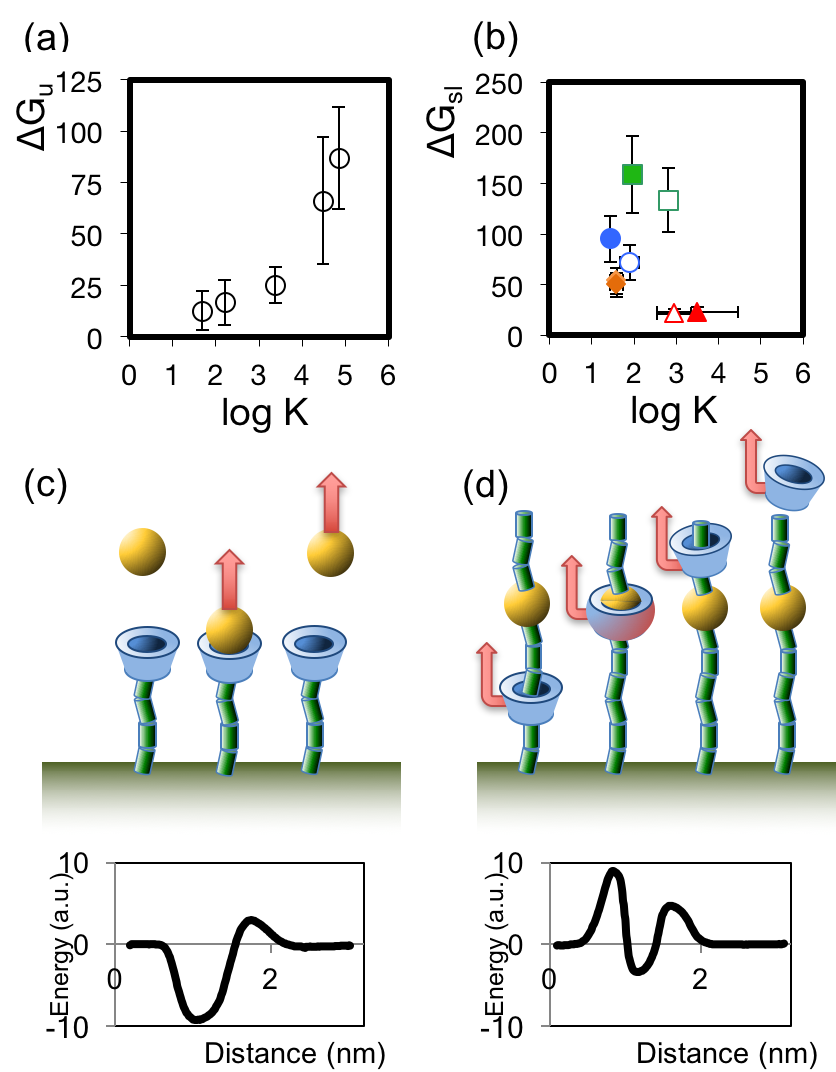


Figure 4. (a) Plot of the relationship between ΔGbu (calculated from the AFM data in [22] using the Friddle-Noy method) and log K (derived from ITC measurements described in [22]) for the host-guest complexes interrogated by Auletta et al31. (b) Plot of the relationship between ΔGsl and log K for the sliding contact experiments described in this work. (c,d) illustrations of the unbinding processes and sketches of possible corresponding energy pathways for disrupting the H-G complex (c), where the depth of the energy well for the H-G complex binding is the principal energy barrier, and the sliding contact experiment (d), where other barriers to the passage of the guest may dominate.

Contributions of solvation and friction to sliding force

We then proceed to consider a more general paradigm to account for the origins of this energy penalty to sliding along the polymer chain. Lulevich et al36, when interpreting the flat, plateau-like events they observed when they used AFM to pull single stranded DNA (ssDNA) out of the interior pore of a carbon nanotube, considered that the total work Wtot required to pull a polymer out of a pore requires the actor to overcome two principal barriers: Wfr, the work arising from the friction accompanying motion of the polymer in the pore, and Wadh, the work arising from the strength of the adhesive interaction between the polymer and the pore, so that Wtot = Wfr + Wadh. The first barrier, Wfr, denotes work done opposing relative motion between the polymer and the pore and may be expected to depend upon the ‘tightness of fit’ of the monomer in the pore, while a key component of Wadh is the difference between the solvation energies of the polymer with the pore interior and with the exterior solvent – a hydrophobic molecule will experience the interior of the CD pore as a more favourable environment than the aqueous bulk phase. Both Lulevich36 and, more recently, Nelson et al,18 have found the sliding of ssDNA within nanopores to be frictionless, although in both cases the pores they investigated (between 1 and 3 nm diameter) were significantly larger than the pore of a cyclodextrin (0.5-0.6 nm diameter for α- and β-CD). The phenomenon of solvation as a barrier to polymer unfolding has been observed previously in SMFS experiments as the Raleigh Instability, resulting on the observation of plateaus as individual polymers are pulled out of the globular conformation they adopt in a poor solvent.37 We applied this approach to our data.

In order to estimate the contributions of friction and solvation to the overall energy of passing a particular monomer through the CD pore, we looked for measured or calculable parameters that reflect these two contributions. As already described, the friction component will depend in some part upon the ‘tightness of fit’ of the monomer inside the CD pore, so the ratio of cross-sectional areas of the monomer and the CD pore, called the dimensionless space-filling parameter Φ19 (calculated using the cross-sectional areas of the monomers38 and the cross-sectional area of the interior pore of the bead) and already shown to have predictive power in estimating the stability constants of host-guest complexes where CDs are the host19,39, may be considered as a proxy for the friction component. The calculation of Φ for stations 1 to 5 is presented in supporting information, and ranges from 0.6 to 3 (**Table S3**). Likewise, the solvation component reflects the passage of the monomer from an aqueous environment, into the hydrophobic interior of the CD pore, and then back out into the aqueous phase again, so that P, the dimensionless octanol:water partition coefficient commonly used (in its log form) as a measure of hydrophobicity in drug design, can be used to describe the relative favourability of these two environments for a particular monomer. Values of log P for stations 1-5 are in the range -3.3 to 0.6, corresponding to values of P between 0 and 4.

We then looked for correlations between ΔGsl, Φ and P, as shown by **Figure S7** in **supporting information**. Taken in isolation, Φ does not show a straightforward relationship with ΔGsl, likely due to the anomalously low values of ΔGsl for the two uronic acids, while there is a clear linear dependence of ΔGsl on P (R2 = 0.90). When we look at the dependence of ΔGsl on the sum Φ + P, we find that all datapoints collapse onto a straight line. Using the method of least squares, we can therefore equate the sum Wfr + Wadh to k.(Φ + P), where k = 22.93 kJ/mol (SE = 0.67). The coefficient of determination for this fit is 0.993 and the data are shown in Figure S7c. We can go further and carry out a multiple linear regression analysis to find the values of the constants kΦ and kP in the terms kΦ.Φ = Wfr and kP.P = Wadh to solve the equation ΔGsl = kΦ.Φ + kP.P. This analysis produces values of 19.8 (SE = 1.4) and 25.6 (SE = 1.2) kJ/mol for kΦand kP respectively. **Figure 5** shows the very close correspondence between the value of ΔGsl measured by SCFS (ΔGsl(meas) ) for the series of stations passing through α- and β-CD and the value calculated from kΦ.Φ + kP.P, (ΔGsl(calc)).

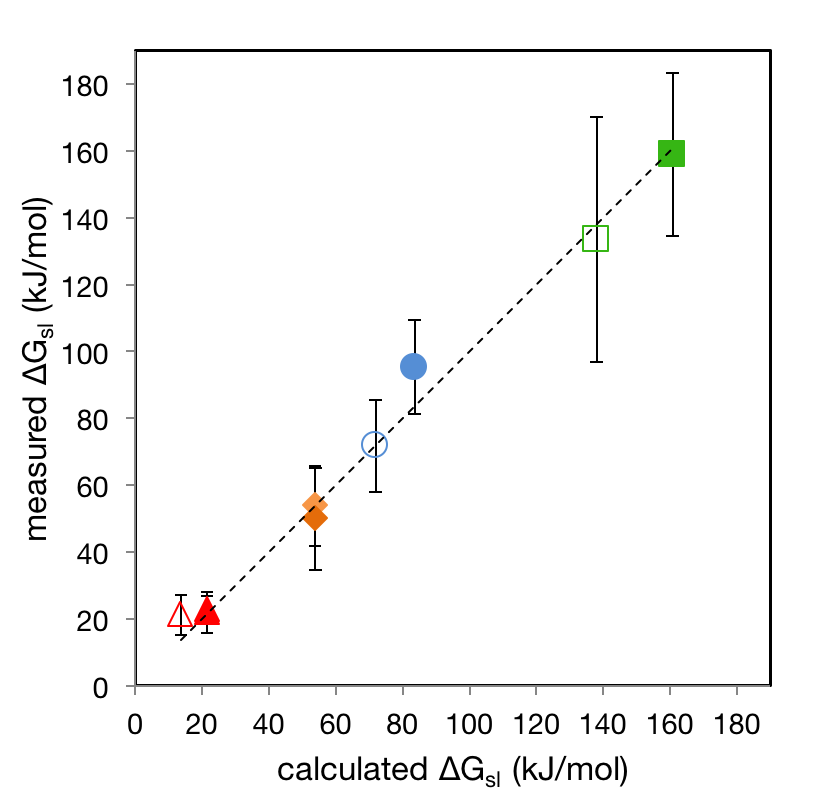


Figure 5. Plot of ΔGsl(meas) vs. ΔGsl(calc) for stations 1-5 passing through α- and β-CD. The dashed line is the fit to the line y = x. The colours of the datapoints follow the pattern used in figure 3. Error bars are 2×SD of the ΔGsl data, encompassing 95% of data assuming normal distribution.

The data used to construct this relation encompasses large and small monomers that are hydrophilic or mildly hydrophobic (max. log P = 0.6). For small, hydrophobic monomers such as ethylene, or for the guest molecules investigated by Auletta et al31 the relation predicts very high energies (more than 500 kJ/mol) due to the dependence on P rather than log P, so clearly our empirical model is applicable over a limited range. Replicating the above analysis, replacing the dependence on P with a dependence on Log P, yields the equation ΔGsl = (52.9×Φ) + (28.4×Log P), which still corresponds reasonably well to ΔGsl(meas) but does predict negative energies in some cases, while providing more likely values of ΔGsl(calc) for hydrophobic monomers. Nevertheless, other monomers that fall within the model’s applicable range include those constituting many important linear and short-branched heteropolymers (essentially all monosaccharides, amino acids, nucleic acids and biocompatible polyhydroxyalkanoates), some of which are not amenable to conventional sequencing methods.

Towards single molecule polymer sequencing

The very clear predictive equation described above ΔGsl = kΦ.Φ + kP.P allows us to consider whether this approach may serve as an alternative, or a first, mapping or sequencing tool for epigenetic modifications of nucleic acids, post-translational modifications of proteins and sidechain patterns in linear glycans. Table S4 in supporting information lists the values of P, Φ, ΔGsl and feq for the four DNA nucleotides, several biocompatible polyhydroxyalkanoates, the 24 standard amino acids and some common post-translational modifications, including phosphorylation of serine, threonine and tyrosine, N-glycosylation of asparagine, O-glycosylation of serine and threonine and methylation of DNA. A check on the applicability of the method may be made by comparing the value of ΔGsl it predicts for ssDNA with the simulated and measured values found by Lindsay and Williams.10,23 Using calculated values of Φ and P we find predicted values of ΔGsl for the four nucleotides and β-CD to fall between 67 and 76 kJ/mol. This is rather larger than the values of ΔGsl calculated using the simulated forces and spring constants reported by Lindsay and Williams10,23 (31-33 kJ/mol), but as noted above, the experimental data for the sliding of β-CD along ssDNA published subsequently10 shows force plateaus of approximately 110pN, corresponding to a ΔGsl of 67 kJ/mol which agrees with our prediction. Nelson et al 18 observed two classes of behaviour for ssDNA sliding in nanopores with dimensions from 1-2 nm: so-called “frictionless” and “stick-slip”. These behaviours were associated with forces of either 12-13 pN or 40-80 pN respectively. Comparison to our predicted ΔGsl values is complicated because calibrated cantilever spring constants for particular measurements are not reported, but for ssDNA the value of Φ falls from 3.51 to 0.38 as the pore diameter increases from 0.6 (β-CD) to 2 nm. Accordingly, the predicted force value we obtain for sliding ssDNA through a 2 nm pore using a probe with a spring constant of 5 pN/nm (within the range quoted by Nelson et al18) is 11 pN. Notwithstanding the difference in magnitude of the forces, the similarity of the values for the four nucleotides reflects the failure of the method to detect differences between nucleotides on the basis of their sliding forces. Both the smaller energy barrier in the simulation and the lack of differentiation between bases reflect the mobility of the base in the nucleotide, allowing it to fold close to the deoxyribophosphate backbone to pass through the CD pore presenting a much smaller cross-sectional area. Similarly, although the difference in predicted ΔGsl between glycine and tryptophan, for example, is large (35 vs. 139 kJ.mol-1), the differences between many amino acids are too small to resolve above thermal noise, and the same is true for methylation of nucleic acids. On the other hand, we have recently shown that the difference in force between sliding α-CD along a single alginate chain and using it to unzip a cross-linked junction zone between two such chains is between 68 and 87 pN, corresponding to 125 kJ/mol.9 Thus, assuming reasonable cantilever spring constants (20-100 pN/nm) and allowing for variation around the value of *f*eq due to thermal noise of 15 pN (this value also reflects ⋝2×SD for all the interactions studied here except for 1:β-CD (see Table 1), so assuming a normal distribution of the force values this range will encompass ⋝95% of events), we predict on the basis of Table S4 that SCFS will detect N- and O-glycosylation of amino acids and glycans. Figure 6 summarises the key results of Table S4 and shows predicted force values for short amino acid and glycan sequences highlighting the differences in predicted force signals for native and modified (phosphorylated and N- and O-glycosylated) amino acid sequences in sections of the MUC-1 protein and the monosaccharide decoration of a plant cell wall hemicellulose.

Examples where this new approach may yield new information include the study of micro- and macroheterogeneity in protein glycosylation40 and the pattern of monosaccharide decoration in polysaccharides, including hemicelluloses whose structure helps determine plant cell wall recalcitrance in bioenergy applications41,42. In both cases, the current state of the art method of analysis is mass spectrometry (MS). SCFS offers advantages over MS methods where the elucidation of sequence patterns over large distances are required. Therefore, SCFS offers the prospect of an alternative route to mapping critical post-translational modifications of proteins and a first method for mapping the pattern of sidechains in linear glycans that can be easily implemented in any standard AFM.

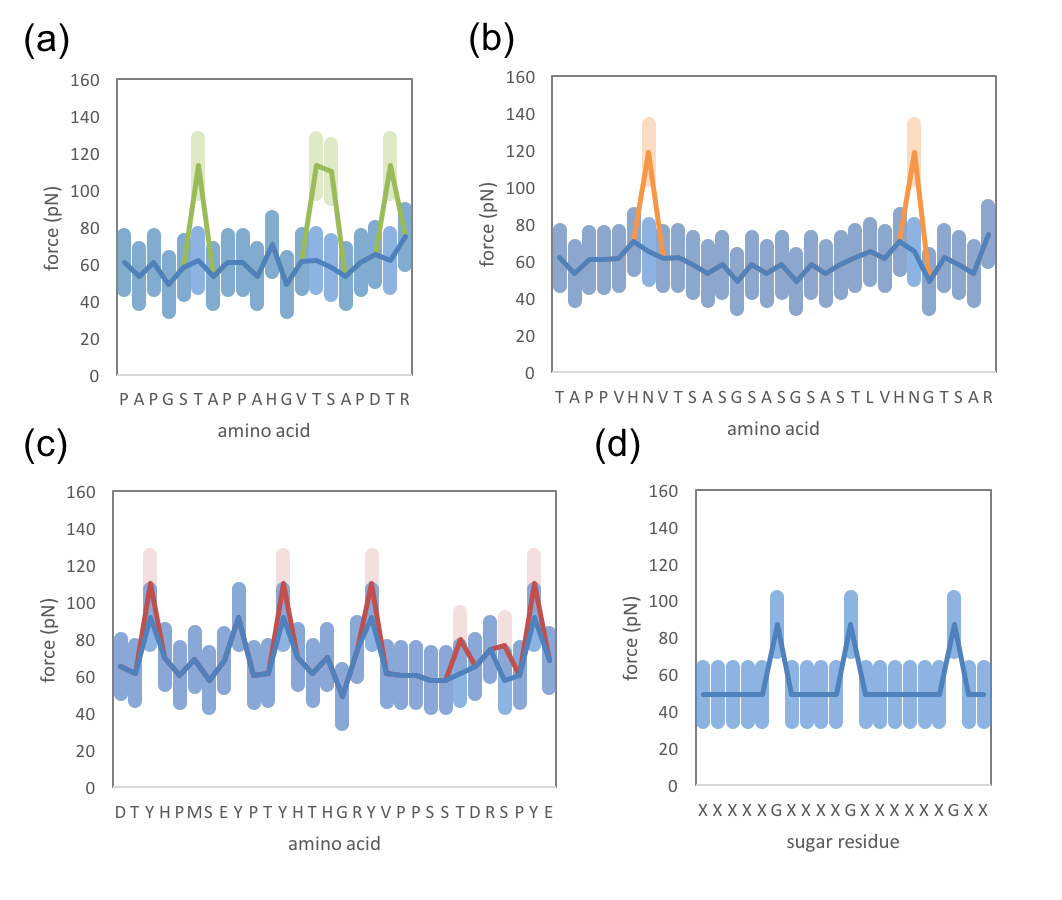


Figure 6. Plots of predicted force vs. sequence for (a) O-glycosylation of the 20 a.a. tandem repeat in the human protein MUC-1(glycosylated sequence in green, unmodified sequence blue); (b) N-glycosylation of the region 951-980 in MUC-1 (glycosylated sequence in orange, unmodified sequence blue); (c) phosphorylation of the region 1201-1230 in MUC-1(phosphorylated sequence in red, unmodified sequence blue); (d) glucuronic acid decoration of the xylan backbone in a hypothetical glucuronoxylan sequence. In all cases solid lines show mean predicted force for a cantilever with kc = 25 pN/nm, error bars represent ±15 pN.

**Experimental methods**

**Cyclodextrin functionalisation**

α- and β-cyclodextrins were modified with a bisamine-terminated PPG-PEG-PPG tether as described previously.1 Briefly, aldehyde groups were created on the cyclodextrins by treatment with Dess-martin periodinane and bis(2-aminopropyl) polypropylene oxide-polyethylene oxide block copolymer was coupled to the aldehyde in a Schiff base reaction.

**Polymer conjugation, pseudorotaxane formation and surface functionalisation of stations 1-3**

The polymers including stations 1, 2 and 3 that were investigated experimentally in this work were prepared as described elsewhere.1 Briefly, aminoaniline was coupled to a formyl-terminated PEG400 polymer by reductive amination, and a thiol group introduced at the distal, hydroxyl-terminated end of the PEG for coupling to a gold substrate.

Samples for AFM were prepared by depositing aqueous solutions of the polymers and rotaxanes on template-stripped gold as follows: 0.4% w/w of each polymer was mixed with a 1:1 mole equivalent of amino-functionalized α- or β-CD for 24 hours, and deposited onto template-stripped gold from water for 24 hours.

AFM probes (MLCT silicon nitride from Veeco Instruments, Santa Barbara, CA, USA) with nominal spring constants of 10 and 20 pN/nm were prepared by coating under vacuum with 1nm Cr and 10nm Au (both from Goodfellow Corp., Berwyn, PA, USA) before incubation with 1mM 11-11’-dithio-bis(succinimidyl undecanoate) in 1,4-dioxane for 10 minutes. Functionalised probes were used immediately or stored in an inert atmosphere.

**Polymer conjugation, pseudorotaxane formation and surface functionalisation of stations 4 and 5**

Alginate oligomers (stations 4 and 5) were fractionated from partially hydrolysed polyG by size exclusion chromatography and freeze dried as previously described.43 Size was assessed with HPAEC-PAD and compositional purity F(G) and degree of polymerisation (DP(n)) were calculated according to both of the methods described in a previous work44 from 1H NMR spectra recorded on a Bruker Avance 400 MHz spectrometer.45,46 HPAEC PAD chromatograms and NMR spectra of the oligoGs were presented previously47. Guluronic acid fractions with n = 6 and n = 16-18, and a mannuronic acid fraction with n = 10 were selected for conjugation to short PEG polymers using a reducing-end-selective method48 previously shown to link polysaccharides to AFM probes and substrates.49 For this conjugation, 0.5 mL of 5 M NaBH3CN (5.0 M solution of sodium cyanoborohydride in aqueous 1 M sodium hydroxide, Sigma-Aldrich), 0.1 mL of 0.5 mM oligosaccharide, 0.25 mL of 0.5 mM amino-PEG-Boc (3000 Da; polydispersity index 1.03) and 1.5 mL of MQ-water were mixed and incubated for 48-144 h. Gel Permeation Chromatography shows that the conjugate has a mass of ~3600 Da, close to the expected mass of ~4200 Da (for details of this analysis see **supporting information**). Prior to conjugation to the substrate surface *tert*-butoxycarbonyl (Boc) deprotection was carried out in a 50% TFA solution for 2 hrs on ice to limit acid hydrolysis.

Samples for AFM were prepared by an alternative method to that used for stations 1-3: freshly-cleaved mica was functionalised with 3-mercaptopropyl triethoxysilane (MTS) (Sigma-Aldrich) from a 2% solution in acetone (200 μL, 20 mins, washed 5x water). To crosslink the amine-terminated PEG-alginate polymer to the thiol-terminated substrate, a short PEG spacer with maleimide and succinimide end groups (SM(PEG)12, Thermo Fisher Scientific) was used (100 μL of 1:300 dilution in water deposited on to thiol-functionalised mica for 2h at RT or overnight at 4°C).

AFM probes for the alginate-pseudorotaxane experiments (MLCT silicon nitride from Veeco Instruments, Santa Barbara, CA, USA) with nominal spring constants of 10 and 20 pN/nm were silanised with thiol-terminated alkylsilane and then further functionalised with (α-maleimido-ω-N-hydroxysuccinimide)-propylene glycol as described above. Both probe treatments resulted in probes functionalized with succinimidyl groups for in-situ reaction with the amine groups on the cyclodextrins and gave comparable success rates.

**AFM force spectroscopy experiments**

Force spectroscopy experiments were carried out using a Multimode AFM with Nanoscope IIIa or V controllers (Veeco Instruments, Santa Barbara, CA, USA) and a JPK Nanowizard III (JPK, Berlin, Germany) in water. The spring constants, calibrated using the thermal tune principle50, ranged from 13.3 to 25.1 pN/nm. The force-distance data were recorded in contact mode, using a setpoint of 0.6 nN and a relative setpoint of 0.2 nN. The z-length varied between 150 nm and 1000nm and the approach speed was set at 0.5 microns per second. For the dynamic force spectroscopy study, retraction speeds were varied from 100 to 500 nm/s, and the resolution adjusted as required. Force spectra were collected in arrays of 100 × 100 data points over areas of 10 × 10 microns. Force spectra were exported and analysed using JPK’s data processing software (JPK instruments, DE, ver. 4.2.23). Observed events were fitted with an extended freely-jointed chain model and the compiled data was analysed using OriginProTM (OriginLab, ver. 8.0724).

**Calculation of Log P**

Chemicalize was used to calculate log P for each of the stations used in this work, April, 2017, <https://chemicalize.com/> developed by ChemAxon (http://www.chemaxon.com).

Acknowledgements

The authors acknowledge support from BBSRC (KB, ANR) and EPSRC (AD, ANR). We are grateful to Andrew Mayes of the School of Chemistry, UEA for his assistance with the GPC experiments.

The authors declare no competing interests.

**Supporting Information**

Details and results of SMFS data selection and analysis; UV data and method; calculation of Φ; table of predicted values of ΔGsl; experimental method and results of gel permeation chromatography of the PEG-uronic acid conjugate.

**References**

1 A. Varki, *Glycobiology*, 2017, **27**, 3–49.

2 O. Nørregaard Jensen, *Curr Opin Chem Biol*, 2004, **8**, 33–41.

3 A. Hirano, Y.-H. Fu and L. J. Ptáček, *Nat. Struct. Mol. Biol.*, 2016, **23**, 1053–1060.

4. M. J. Booth, E.-A. Raiber and S. Balasubramanian, *Chemical Reviews*, 2015, **115(6)**, 2240–2254.

5 F. A. Leibfarth, J. A. Johnson and T. F. Jamison, *Proceedings of the National Academy of Sciences*, 2015, **112**, 10617–10622.

6 J. C. Barnes, D. J. C. Ehrlich, A. X. Gao, F. A. Leibfarth, Y. Jiang, E. Zhou, T. F. Jamison and J. A. Johnson, *Nat Chem*, 2015, **7**, 810–815.

7 J.-F. Lutz, J.-M. Lehn, E. W. Meijer and K. Matyjaszewski, *Nature Reviews Materials*, 2016, **1**, 16024.

8 A. Dunlop, J. Wattoom, E. A. Hasan, T. Cosgrove and A. N. Round, *Nanotechnology*, 2008, **19**, 345706.

9 K. A. Bowman, O. A. Aarstad, B. T. Stokke, G. Skjåk-Braek and A. N. Round, *Langmuir*, 2016, **32**, 12814-12822.

10 B. A. Ashcroft, Q. Spadola, S. Qamar, P. Zhang, G. Kada, R. Bension and S. M. Lindsay, *Small*, 2008, **4**, 1468–1475.

11 A. Harada and M. Kamachi, *Macromolecules*, 1990, **23**, 2821–2823.

12 C. Cheng and J. F. Stoddart, *Chemphyschem*, 2016, **17**, 1780–1793.

13 M. von Delius and D. A. Leigh, *Chem. Soc. Rev.*, 2011, **40**, 3656–3676.

14 P. Lussis, T. Svaldo-Lanero, A. Bertocco, C.-A. Fustin, D. A. Leigh and A.-S. Duwez, *Nature Nanotechnology*, 2011, **6**, 553–557.

15 M. Wanunu, *Physics of Life Reviews*, 2012, **9**, 125–158.

16 A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure and J. H. Gundlach, *Nat Biotechnol*, 2014, **32**, 829–833.

17 G. Baaken, I. Halimeh, L. Bacri, J. Pelta, A. Oukhaled and J. C. Behrends, *Acs Nano*, 2015, **9**, 6443–6449.

18 E. M. Nelson, H. Li and G. Timp, *Acs Nano*, 2014, **8**, 5484–5493.

19 G. Wenz, B. Han and A. Muller, *Chem Rev*, 2006, **106**, 782–817.

20 H. Shigekawa, K. Miyake, J. Sumaoka, A. Harada and M. Komiyama, *J Am Chem Soc*, 2000, **122**, 5411–5412.

21 B. Brough, B. H. Northrop, J. J. Schmidt, H.-R. Tseng, K. N. Houk, J. F. Stoddart and C.-M. Ho, *Proceedings of the National Academy of Sciences*, 2006, **103**, 8583–8588.

22 N. K. Voulgarakis, A. Redondo, A. R. Bishop and K. O. Rasmussen, *Nano Lett*, 2006, **6**, 1483–1486.

23 S. Qamar, P. M. Williams and S. M. Lindsay, *Biophys J*, 2008, **94**, 1233–1240.

24 R. W. Friddle, A. Noy and J. J. De Yoreo, *Proceedings of the National Academy of Sciences*, 2012, **109**, 13573–13578.

25 A. Noy and R. W. Friddle, *Methods*, 2013, **60**, 142–150.

26 E. Evans, *Faraday Discuss*, 1999, **111**, 1–16.

27 S. Guo, N. Li, N. Lad, S. Desai and B. B. Akhremitchev, *The Journal of Physical Chemistry C*, 2010, **114**, 8755–8765.

28 B. D. Rackham, L. A. Howell, A. N. Round and M. Searcey, *Org Biomol Chem*, 2013, **11**, 8340–8347.

29 S. J. Suresh and V. M. Naik, *The Journal of Chemical Physics*, 2000, **113**, 9727–9732.

30 I. F. Hakem, A. Boussaid and H. Benchouk-Taleb, *The Journal of …*, 2007.

31 T. Auletta, M. R. de Jong, A. Mulder, F. C. J. M. van Veggel, J. Huskens, D. N. Reinhoudt, S. Zou, S. Zapotoczny, H. Schonherr, G. J. Vancso and L. Kuipers, *J Am Chem Soc*, 2004, **126**, 1577–1584.

32 K. A. Connors, *Journal of pharmaceutical sciences*, 1995, **84**, 843–848.

33 H. J. Buschmann and E. Schollmeyer, *J Incl Phenom Macro*, 1997, **29**, 167–174.

34 H. J. Buschmann, K. Jansen and E. Schollmeyer, *J Incl Phenom Macro*, 2000, **37**, 231–236.

35 A. F. D. de Namor, P. M. Blackett, M. C. Cabaleiro and J. M. A. Al Rawi, *J Chem Soc Faraday T*, 1994, **90**, 845.

36 V. Lulevich, S. Kim, C. P. Grigoropoulos and A. Noy, *Nano Lett*, 2011, **11**, 1171–1176.

37 A. Scherer, C. Zhou, J. Michaelis, C. Brauchle and A. Zumbusch, *Macromolecules*, 2005, **38**, 9821–9825.

38 V. Privalko, *Macromolecules*, 1980, **13**, 370–372.

39 A. Müller and G. Wenz, *Chemistry*, 2007, **13**, 2218–2223.

40 K. Mariño, J. Bones, J. J. Kattla and P. M. Rudd, *Nat Chem Biol*, 2010, **6**, 713–723.

41 M. Busse-Wicher, N. J. Grantham, J. J. Lyczakowski, N. Nikolovski and P. Dupree, *Biochem Soc T*, 2016, **44**, 74–78.

42 S. M. Cragg, G. T. Beckham, N. C. Bruce, T. D. Bugg, D. L. Distel, P. Dupree, A. G. Etxabe, B. S. Goodell, J. Jellison, J. E. McGeehan, S. J. McQueen-Mason, K. Schnorr, P. H. Walton, J. E. Watts and M. Zimmer, *Curr Opin Chem Biol*, 2015, **29**, 108–119.

43 S. Ballance, S. Holtan, O. A. Aarstad, P. Sikorski, G. Skjåk-Braek and B. E. Christensen, *Journal of Chromatography A*, 2005, **1093**, 59–68.

44 H. Grasdalen, B. Larsen and O. Smidsrød, *Carbohyd Res*, 1979, **68**, 23–31.

45 C. Campa, S. Holtan, N. Nilsen, T. M. Bjerkan, B. T. Stokke and G. Skjåk-Braek, *Biochem J*, 2004, **381**, 155–164.

46 C. Campa, A. Oust, G. Skjåk-Braek, B. S. Paulsen, S. Paoletti, B. E. Christensen and S. Ballance, *Journal of Chromatography A*, 2004, **1026**, 271–281.

47 K. A. Bowman, O. A. Aarstad, M. Nakamura, B. T. Stokke, G. Skjåk-Braek and A. N. Round, *Carbohyd Polym*, 2016, **148**, 52–60.

48 G. Gray, *Meth Enzymol*, 1978, **50**, 155–160.

49 M. Takemasa, M. Sletmoen and B. T. Stokke, *Langmuir*, 2009, **25**, 10174–10182.

50 J. L. Hutter and J. Bechhoefer, *Rev Sci Instrum*, 1993, **64**, 1868–1873.

**Electronic Supporting Information**

**Polymer sequencing by molecular machines: A framework for predicting the resolving power of a sliding contact force spectroscopy sequencing method.**

*Alex Dunlopa, Kate Bowmanb, Olav Aarstadc, Gudmund Skjåk-Brækc, Bjørn T. Stokked, Andrew N. Roundb\**

(a) Dr. A. Dunlop

HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

(b) Dr. K. Bowman, Dr. A. N. Round (\*corresponding author; e-mail: a.round@uea.ac.uk)

School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

(c) Dr. O. A. Aarstad, Prof. G. Skjåk-Bræk

Department of Biotechnology, The Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

(d) Prof. B. T. Stokke

Department of Physics, The Norwegian University of Science and Technology, NTNU, NO-7491 Trondheim, Norway

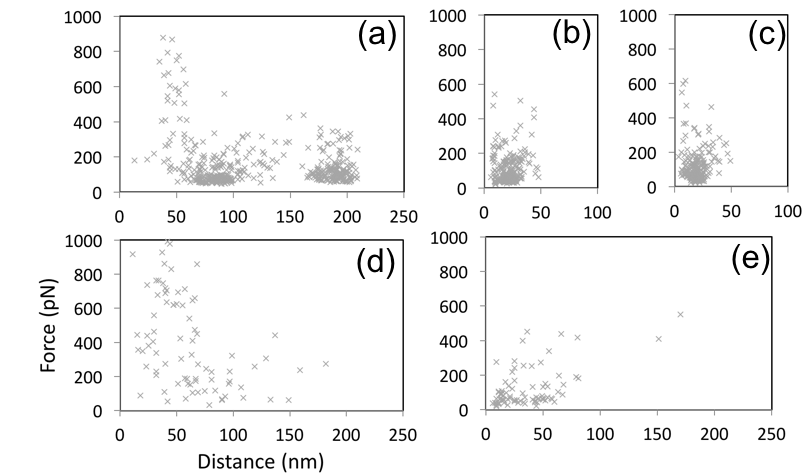
**1. Single Molecule Force Spectroscopy Analysis**

**1.1 Data Selection**

Force spectroscopy data was first assessed within the JPK processing software. All force-distance profiles to which the extended freely jointed chain model (eFJC) [S1] could be fit were selected for further analysis. Fits to the eFJC model were performed within the JPK processing software, and values for the free parameters fitted to the model (contour length, breaking force and Kuhn length) were extracted.

Within the data collected, three types of force-distance profiles could be discerned (Figure 3a, main text). In the first, found when the CD is pulled along the rotaxane containing stations 1 and 2, there are two stretches with rupture forces of around 50-250 pN separated by a distance of approximately 100nm, between which the force reaches a plateau value approximately 40 pN above the baseline. In the second, found when the CD is pulled along rotaxanes containing stations 4 or 5, there is a single stretching event terminating at a force of approximately 50-150 pN which may sometimes be preceded by a discernible plateau at around 30-40 pN. The third type of event is found in all experiments and consists of a single stretching event, terminating at a high rupture force (these are the only events to exceed 300 pN). When polymers lacking threaded CDs are used instead of the corresponding rotaxane, and when an AFM probe with no tether is used instead of the NHS-terminated tether, then few events occur and when they do, force-distance profiles of the third type are exclusively observed. **Figure 3(a)** in the main text shows a selection of force-distance profiles from the systems explored here, with examples of each of first two types of force-distance profiles discussed above.

Plots of breaking force *f* vs. contour length L (**Figure S1**) were inspected in order to identify the force-distance profiles corresponding to sliding over the incorporated stations. In the rotaxane containing stations 1 and 2, a high density of breaking events occur around lengths of 90 ± 20 nm and 200 ± 20 nm, in agreement with the expected positions of stations 2 and 1 respectively as previously shown [S2]. Breaking events that fell within these two length ranges, and which possess a plateau in force preceding the stretching event, were identified as station sliding events and were selected for further analysis. A similar procedure was conducted for the rotaxanes containing stations 4 and 5. Here the oligouronates are attached to shorter PEGs with contour lengths of about 20 nm and a high density of rupture events occurs at the corresponding distance from the surface.

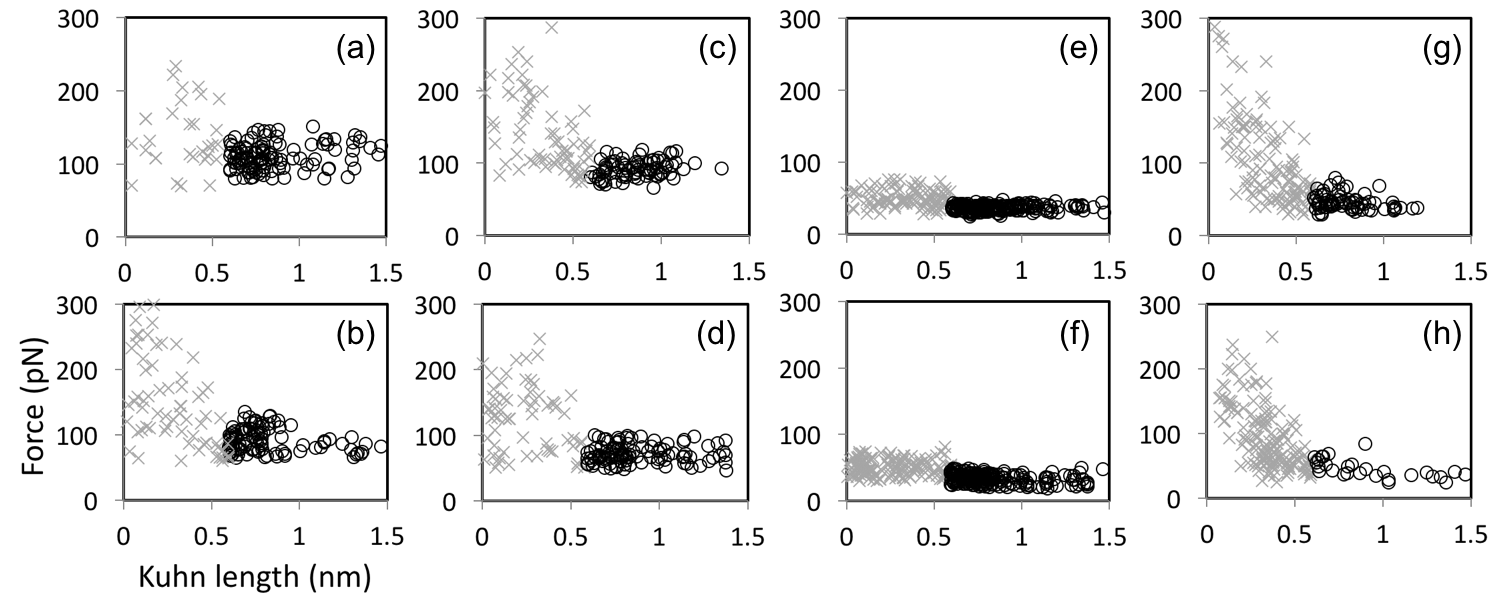


**Figure S1.** Plots of breaking force (pN) vs distance (nm) for rupture events in the following systems: (a,e) rotaxanes containing α-CD and stations 1, 2 and 3; (b) rotaxanes containing α-CD and stations 3 and 4; (c) rotaxanes containing α-CD and stations 3 and 5; (d) polymers containing stations 1, 2 and 3 but no CD; (e) same as (a) but with a bare tip.

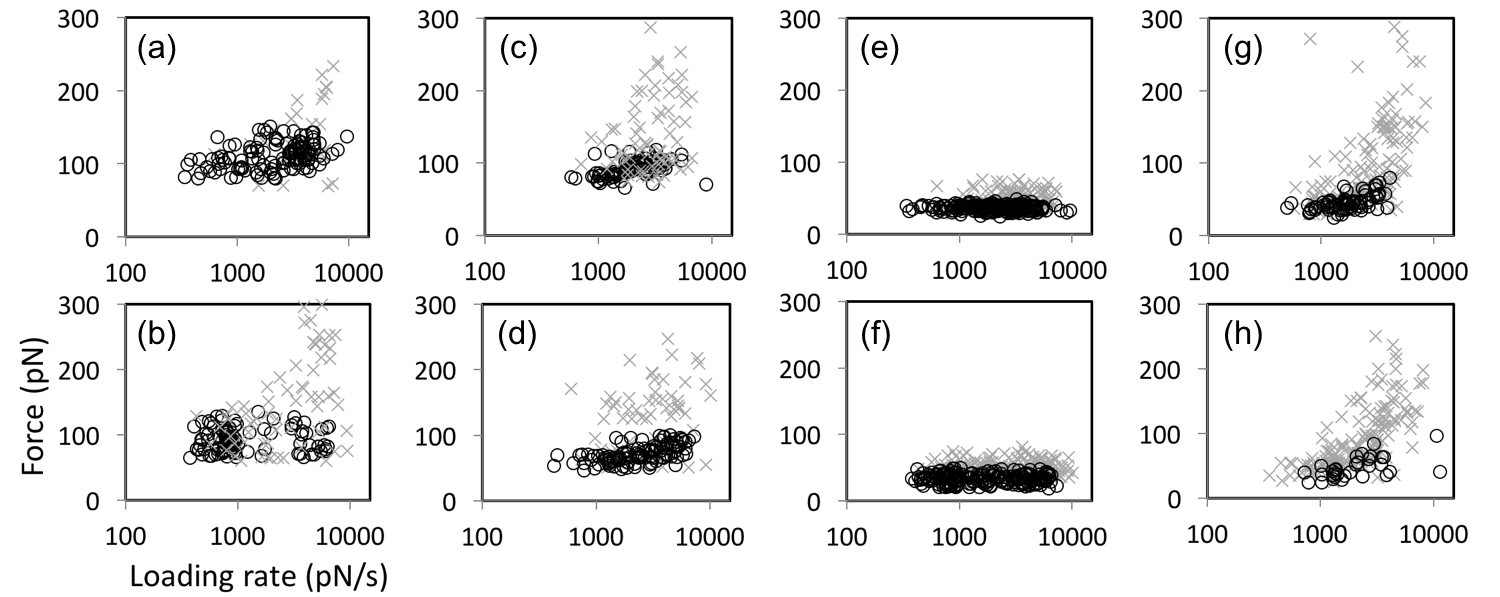
The Kuhn length for a single polyethylene glycol (PEG) polymer (as we use here as a flexible spacer) has been estimated to be 0.7 nm [S3]. Akhremitchev has shown [S4] that the Kuhn length is diagnostic for the presence of a single polymer stretch, since the value of the apparent Kuhn length varies as 1/n, where n = number of simultaneous parallel polymer stretches. **Figure S2** shows plots of rupture force vs. Kuhn length for the rupture events in the rotaxane systems described in this work. In the case of stations 1, 2, 4 and 5, it can be seen that below a Kuhn length value of 0.6 nm there is an increase in the number of high force events, while there are also substantial populations of events with similar forces to the events where the Kuhn length exceeds 0.6 nm. For sliding events involving station 3 the Kuhn length cannot be determined directly so the designated Kuhn length is that found for the sliding event involving one of the other stations that immediately follows the sliding plateau over station 3. **Figure S3** shows that, at comparable loading rates, the stretches fit by Kuhn lengths of 0.6 nm or greater ruptured at lower forces than the stretches fit by Kuhn lengths lower than 0.6 nm. This result is in accordance with the predictions of Akhremitchev and suggests that events with lower Kuhn lengths represent multiple simultaneous (or nearly so) bond ruptures. We therefore selected for further analysis only those events with a Kuhn length of 0.6 nm or greater.

**Figure S2** (following page). Plots of Force (pN) vs. Kuhn length (nm) for sliding events for (a) α-CD and (b) β-CD over station 1; (c) α-CD and (d) β-CD over station 2; (e) α-CD and (f) β-CD over station 3; (g) α-CD over station 4 and (h) α-CD over station 5. Crosses represent events fit by a Kuhn length less than 0.6 nm (representing multiple interactions), open circles represent events fit by Kuhn lengths of 0.6 nm or greater (representing single interactions).

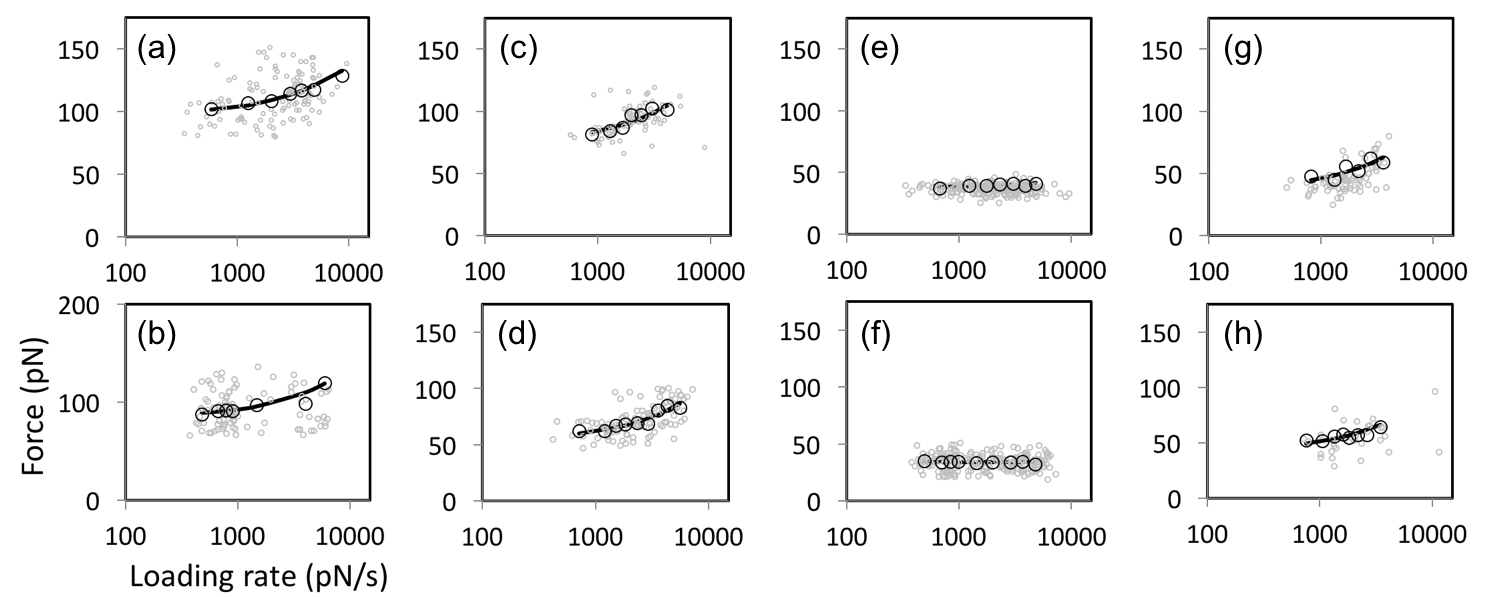
**Figure S3** (following page)**.** Plots of Force (pN) vs. Loading rate (pN/s) for sliding events for (a) α-CD and (b) β-CD over station 1; (c) α-CD and (d) β-CD over station 2; (e) α-CD and (f) β-CD over station 3; (g) α-CD over station 4 and (h) α-CD over station 5. Note the logarithmic scale of the *x*-axis. Crosses represent events fit by a Kuhn length less than 0.6 nm (representing multiple interactions), open circles represent events fit by Kuhn lengths of 0.6 nm or greater (representing single interactions).



**Figure S2.**



**Figure S3.**

****

**Figure S4.**

**Figure S4.** Dynamic Force Spectra (DFS) plots of Force (pN) vs. Loading rate (pN/s) for sliding events for (a) α-CD and (b) β-CD over station 1; (c) α-CD and (d) β-CD over station 2; (e) α-CD and (f) β-CD over station 3; (g) α-CD over station 4 and (h) α-CD over station 5. Note the logarithmic scale of the *x*-axis. Open circles represent most probable rupture forces in intervals of loading rate, solid lines represent fits of equation S1 to the data.

**1.2 Dynamic Force Spectroscopy Analysis**

The data selected by the above criteria was ordered by loading rate into bins of equal width and histograms of the rupture forces in each loading rate bin were fit by Gaussian distributions. The most probable rupture forces predicted by these Gaussian fits were then used to construct a dynamic force spectrum, as presented in **Figure S4**. This spectrum was then fit by the Friddle-Noy method [S5, S6] in order to extract values for the free parameters koff and xt (the intrinsic unbinding rate of the bond and the distance to the transition state, respectively) using equations S1 and S2, as well as a third parameter, the equilibrium force *f*eq (the minimum force required to move the binding pair apart by the distance xt, beyond which they can no longer instantaneously rebind). From *f*eq may be obtained ΔGbu, the free energy of unbinding, for the bond using equation S3 [S5, S6]. All fits to the data were carried out using OriginPro™ (OriginLab, ver. 8.0724).

(S1)

Here <*f*> is the mean rupture force, *f*β = kBT/xt is the thermal force scale, kBT is the Boltzmann constant multiplied by the temperature, γ = 0.577… is Euler’s constant and r is the loading rate. Values of the parameters xt, ku(f) (the unbinding constant at *f*) and *f*eq were obtained from the fit of equation S1 to the data, while koff was obtained from the value of ku(f) and equation S2:

(S2)

where keff is the effective spring constant of the system comprising the cantilever stiffness kc and the linker stiffness kl; keff = kc.kl/(kc+kl). A value of 0.067 Nm-1 was used for the linker stiffness [S3]. From the values of *f*eq and keff the value of ΔGbu can be calculated using equation S3, using the terms already defined above:

(S3)

Values of xt, koff, *f*eq and ΔGbu were found from the fits of equations S1-3 to the data and presented in **Table S1**. For station 3, no dependence of <*f*> on r is observed so xβ and koff cannot be calculated. **Figure S4** shows the dynamic force spectra and fits to equation S1. In the cases where we are analysing SCFS data, the sliding force term ΔGsl is identical to the term ΔGbu used above.

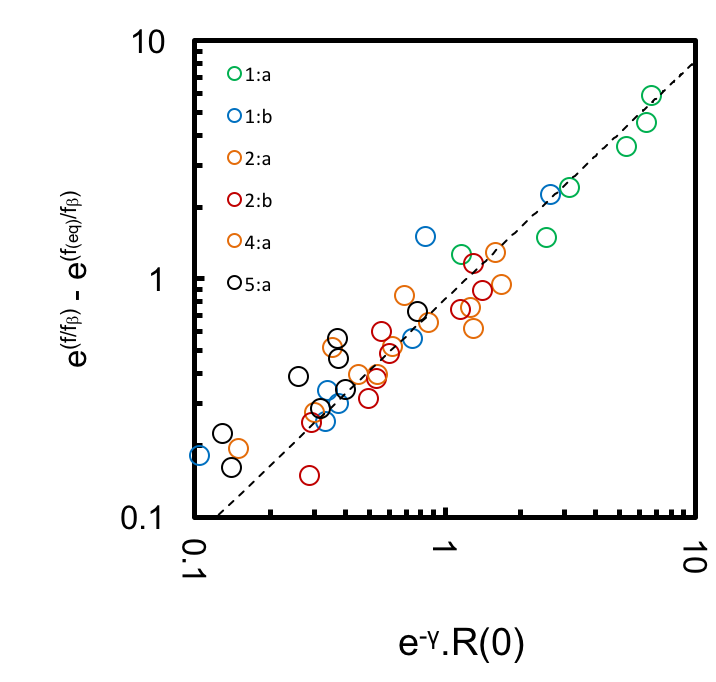
**Table S1.** Fitted parameter values for the F-N-Y fit to the dynamic force spectra for stations 1, 2, 4 and 5.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Station** | **Bead** | **koff (/s)** | **xt (nm)** | ***f*eq ± sd (pN)** | **kc (pN/nm)** | **ΔGbu ± sd**  **(kJ/mol)** |
| 1 | α-CD | 8.4 | 0.10 | 98 ± 8 | 25 | 159 ± 17 |
| β-CD | 14.4 | 0.08 | 86 ± 23 | 22 | 134 ± 26 |
| 2 | α-CD | 14.6 | 0.08 | 76 ± 6 | 25 | 95 ± 13 |
| β-CD | 26.1 | 0.08 | 63 ± 12 | 22 | 72 ± 14 |
| 3 | α-CD | - | - | 29 ± 4 - 37 ± 3 | 14-25 | 22 ± 4 - 23 ± 4 |
| β-CD | - | - | 34 ± 5 | 22 | 21 ± 6 |
| 4 [e] | α-CD | 28.9 | 0.09 | 45 ± 5 | 14 | 54 ± 12 |
| 5 [e] | α-CD | 23.1 | 0.10 | 46 ± 7 | 16 | 52 ± 15 |

Equation S4 is the result of rearranging equation S1 [S5], so that we can conveniently test the validity of S1 graphically by plotting the terms in the equation against each other; a linear fit with a gradient close to 1 implies the validity of equation S1. **Figure S5** shows this plot for stations 1, 2, 4 and 5.

, (S4)

(S4a)



**Figure S5.** Plot of the same data shown in Figure S4 (a-d and g-h), plotted in the natural coordinates of equation S1, showing the collapse of the data onto a straight line. The dashed line is a linear fit to the data, with slope = 0.82 (R2 = 0.97).

**1.3. Calculating ΔGbu values for the data published by Auletta et al [S7]**

Auletta et al. (reference 31 in the main text) provide data on the observed rupture forces (and their lack of dependence on loading rate, implying that these are equilibrium forces) and a range of values for the spring constants of the probes used for their force measurements, so we are able to estimate values of ΔGbu for the rupture of complexes between the guest molecules they studied and β-CD using equation S3. **Table S2** reports these values, together with the same calculations based on the estimated forces simulated by Qamar et al [S8] for the sliding of β-CD over purine and pyrimidine DNA nucleotides.

**Table S2.** Calculated values of ΔGbu using the reported values of *f*eq and kc (0.05-0.12 N/m) in Auletta et al [S7], and in Qamar et al (kc = 0.3 N/m) [S8].

|  |  |  |
| --- | --- | --- |
| **Guest** | ***f*eq ± sd (pN)** | **ΔGbu ± sd (kJ/mol)** |
| anilyl | 39 ± 15 | 12.7 ± 9.5 |
| toluidyl | 45 ± 15 | 16.9 ± 10.9 |
| ferrocenyl | 55 ± 10 | 25.3 ± 8.9 |
| tert-butylphenyl | 89 ± 15 | 66.1 ± 30.9 |
| adamantyl | 102 ± 15 | 86.8 ± 24.7 |
| purines | 75 | 31 |
| pyrimidines | 78 | 33 |

**2. UV Measurements of Equilibrium constant K**

In order to explore the nature of the relationship between the observed free energy of sliding and the equilibrium constant K for host:guest complexation we collected values of K for the host:guest complexes studied. Literature values are reported for stations 2, 3, 4 and 5 with α- and β-CD (references 22-25 in the main text). To the best of our knowledge the equilibrium constants for the host:guest complex between station 1 and α- and β-CDs have not previously been reported, so we measured K for stations 1 and 2 using the UV spectroscopic method presented below. All values of log K are presented in **Table 1** in the main text.

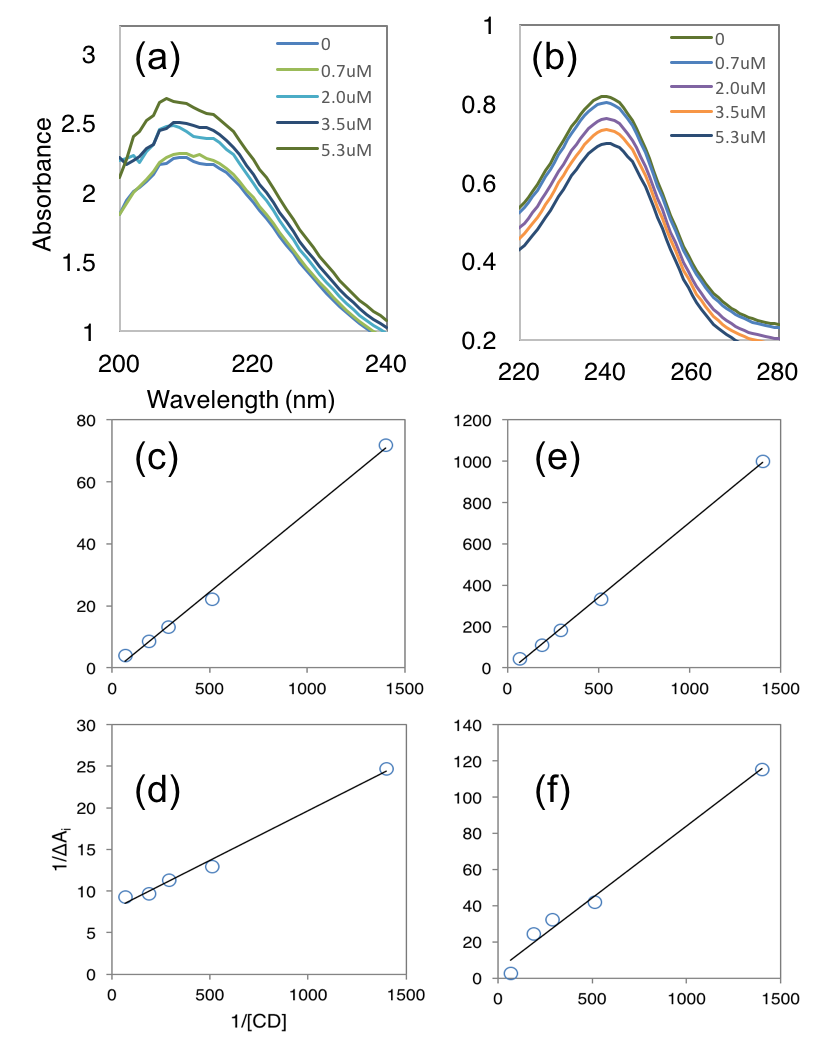
We employed a spectroscopic method that takes advantage of the red shift in the wavelength of maximum absorption (λmax) and corresponding increase in molar absorption coefficient (ε) that is observed when an aromatic group is included in a CD host in aqueous conditions [S9]. UV spectra are obtained for solutions with a fixed guest concentration [G] and a varied host concentration [CD], and the absorbances at λmax of the uncomplexed guest and at a background wavelength away from the peak in absorbance are recorded. Initially, in order to most closely replicate the situation addressed by the AFM experiments, we sought to measure the complexation of CD with PEG-tethered stations 1 and 2. In each case we observed no, or very little, shift in the UV spectra, suggesting that no complexation between the CDs and stations 1 or 2 occurred. This may be because the CDs preferentially complexed with the PEG, and since PEG lacks chromophores detectable by UV-vis spectroscopy no signal was observed. When, instead, stations 1 and 2 were introduced to CD alone, we observed shifts in the spectra as shown in **Figure S6**. Where such a system exhibits a 1:1 stoichiometry with the host CD, the equilibrium constant K can be evaluated using equation S5:

(S5)

Using the well-known Benesi-Hildebrand method [S10], we see that equation S6 describes a linear relationship between 1/ΔAi and 1/[CD], with a slope of 1/(ΔA.K) and an intercept of 1/ΔA, so that the ratio of intercept to slope yields K. Here, ΔAi is the difference in absorption at λmax between the guest alone and the guest:host complex and ΔA is Δε.b.[G], where Δε = ε∞ – ε0 (difference in molar absorption coefficients for the complex and the guest alone) and b = path length of the UV cell.

(S6)

The Benesi-Hildebrand method has been shown to produce inaccurate results under some conditions but when K is in the region 50-1000 M-1 and Δε is greater than 50 m2/mol then reliable results are produced [S10]. In the work presented here K varied from 25 to 646 M-1 and Δε varied between 13 and 150 m2/mol, so the method used may have introduced some errors. Where literature data (derived from calorimetric titration) is available for the aminoaniline (station 2) complexation with α- and β-CD [S11], our result is in broad agreement (log K = 1.47 and 1.41 respectively for our experiments and the literature value for the complexation of aminoaniline and α-CD, and for aminoaniline:β-CD log K = 1.74 and 2.05 respectively for our experiments and the literature value).



**Figure S6.** UV spectra for (a) station 1 (aminoaniline) and (b) station 2 (pyromellitic acid) complexing with CD at concentrations ranging from 0 to 5.3 μM. Benesi-Hildebrand plots for the complexation of (c,d) station 1 with α- and β-CD; (e,f) station 2 with α- and β-CD.

**3. Calculating the dimensionless space-filling parameter Φ**

The dimensionless space-filling parameter Φ is the ratio of the cross-sectional area of the monomer, A, and the cavity area of the CD used (20.18 Å2 for α-CD and 32.95 Å2 for β-CD [S12]). Privalko [S13] proposed a set of equations to allow the calculation of A from m0, the monomer molecular mass per main chain bond, depending on its expected crystalline conformation (extended chain, EC, equation S7a; helical chain, HC1, equation S7b; helical chain with bulky sidegroups, HC2, equation S7c; constants of proportionality k and k’ have numerical values of 1 but dimensions of L2/M0.5 (S7a) and L2/M (S7b and c)):

(S7a)

(S7b)

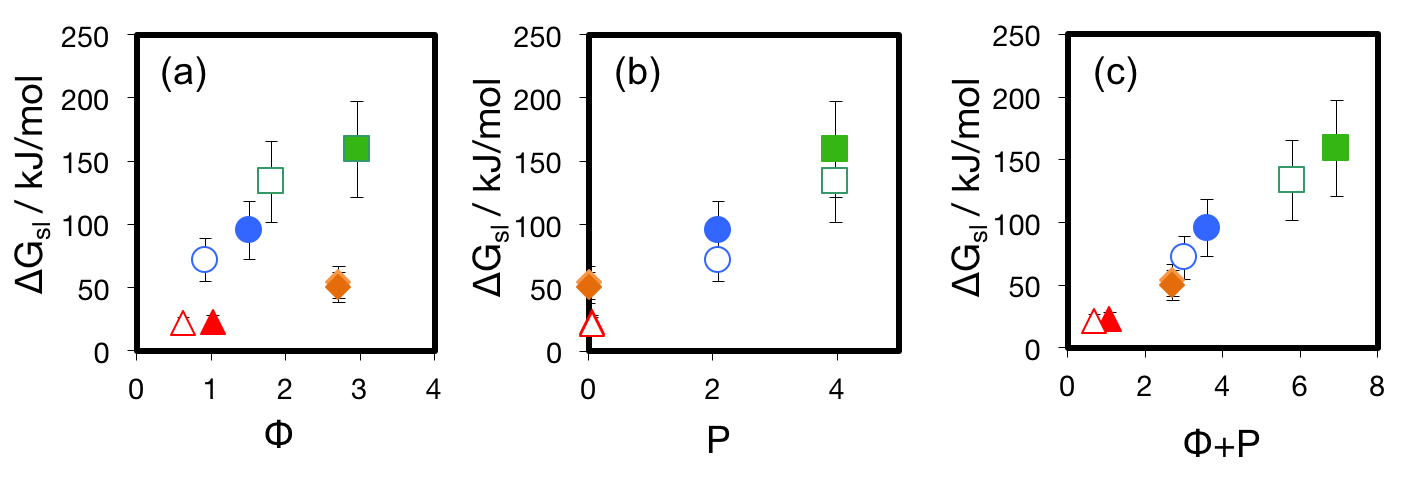
(S7c)

Following the assignments made in the literature [S13], all the stations studied here would be expected to adopt the HC1 conformation, so equation S7b is appropriate. Table S3 presents the calculated values of A and Φ for the stations and CDs used in this work.

**Table S3.** Calculated values of A and Φ for stations 1-5 and α- and β-CD.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Station** | **Monomer MW** | **No. main chain bonds** | **A (Å2)** | **Φ (α-CD)** | **Φ (β-CD)** |
| 1 | 254.15 | 6 | 59.73 | 2.96 | 1.81 |
| 2 | 108.144 | 5 | 30.50 | 1.51 | 0.93 |
| 3 | 44.05 | 3 | 20.70 | 1.03 | 0.63 |
| 4 | 194.13 | 5 | 54.74 | 2.71 | n.d. |
| 5 | 194.13 | 5 | 54.74 | 2.71 | n.d. |

**Dependences of ΔGsl on Φ and P**

****

**Figure S7.** Plots of the relationships between ΔGsl and (a) Φ; (b) P; and (c) Φ+P. Colours of datapoints follow the pattern used in figure 3 in the main text.

**Table S4.** Calculated values of P, Φ, ΔG*sh* and *feq* for selected monomers and their interactions with α- and β-CD.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | P  (×10-4) | Φ | ΔGsh  (kJ/mol) | feq  (pN) |
| nucleotides and modifications  (β-CD)\* |  |  |  |  |
| adenosine monophosphate | 10.2 | 3.6 | 71.6 | 65.8 |
| thymidine monophosphate | 575.4 | 3.3 | 68.2 | 64.2 |
| cytidine monophosphate | 7.2 | 3.4 | 66.8 | 63.5 |
| methylcytidine monophosphate | 17.0 | 3.5 | 68.9 | 64.6 |
| guanosine monophosphate | 7.6 | 3.8 | 75.6 | 67.6 |
|  |  |  |  |  |
| Glycans (α-CD) |  |  |  |  |
| xylose | 50.1 | 1.8 | 34.8 | 45.9 |
| glucuronoxylose disaccharide\* | 1.8 | 5.5 | 109.6 | 81.6 |
|  |  |  |  |  |
| polyhydroxyalkanoates  (α-CD) |  |  |  |  |
| polylactic acid | 3388 | 2.1 | 50.3 | 55.1 |
| polyglycolic acid | 912 | 1.8 | 37.4 | 47.6 |
| Polyhydroxybutyric acid | 4074 | 1.8 | 46.5 | 53.0 |
| Polyhydroxyvaleric acid | 13804 | 2.1 | 76.3 | 67.9 |
| Polycaprolactone | 23442 | 1.3 | 86.3 | 72.2 |
| Polyhydroxypropylmethacrylamide | 13490 | 5.1 | 135.1 | 90.4 |
|  |  |  |  |  |
| amino acids and modifications  (α-CD) |  |  |  |  |
| arginine | 6.9 | 4.1 | 80.4 | 69.7 |
| histidine | 2.4 | 3.6 | 71.6 | 65.8 |
| lysine | 6.2 | 3.4 | 67.5 | 63.9 |
| aspartic acid | 3.2 | 3.1 | 61.4 | 60.9 |
| glutamic acid | 5.8 | 3.4 | 67.9 | 64.1 |
| serine | 1.3 | 2.4 | 48.5 | 54.2 |
| phosphoserine | 6.6 | 4.3 | 85.5 | 71.9 |
| O-glycosylserine\* | 2.0 | 8.8 | 174.6 | 102.8 |
| threonine | 3.4 | 2.8 | 55.0 | 57.7 |
| phosphothreonine | 20.0 | 4.6 | 92.0 | 92.0 |
| O-glycosylthreonine\* | 2.0 | 9.3 | 184.0 | 105.5 |
| asparagine | 0.5 | 3.1 | 61.0 | 60.7 |
| N-acetylglucosylaminoasparagine\* | 2.0 | 10.3 | 203.5 | 110.9 |
| glutamine | 1.0 | 3.4 | 67.4 | 63.8 |
| cysteine | 16.2 | 2.8 | 55.9 | 58.1 |
| glycine | 3.8 | 1.7 | 34.7 | 45.8 |
| proline | 26.9 | 2.7 | 53.2 | 56.7 |
| alanine | 14.5 | 2.1 | 41.2 | 49.9 |
| valine | 109.6 | 2.7 | 54.3 | 57.3 |
| isoleucine | 309.0 | 3.1 | 61.3 | 60.9 |
| leucine | 257.0 | 3.1 | 61.2 | 60.8 |
| methionine | 64.6 | 3.5 | 69.0 | 64.6 |
| phenylalanine | 645.7 | 5.6 | 112.5 | 82.5 |
| tyrosine | 316.2 | 6.1 | 122.4 | 86.0 |
| phosphotyrosine | 309.0 | 8.8 | 176.1 | 103.2 |
| tryptophan | 812.8 | 6.9 | 139.1 | 91.7 |

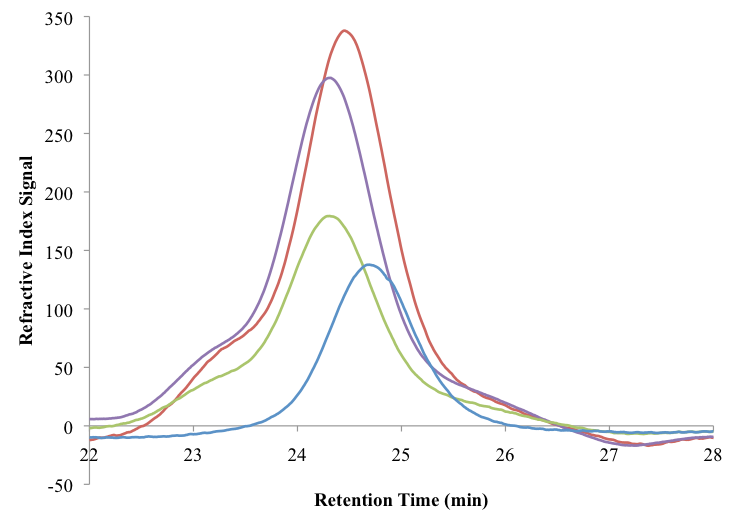
\* Φ calculated using equation S7c

**4. Gel Permeation Chromatography (GPC)**

Gel permeation chromatography (GPC) with refractive index detection (PL aquagel-OH 30 (Agilent Technologies, Santa Clara, CA, USA) column with a low flow rate of 0.3 ml/min and a run time of 40 min) was used to determine the increase in molecular weight and so establish the success of the conjugation between oligomers of stations 4 and 5 (guluronic and mannuronic acids) and PEG. Calibration of the GPC retention time with respect to molecular weight was carried out using two polymer series: PEGs coupled to maltose and a series of guluronate oligomers. Different relationships between molecular weight and retention time were found for the two series (**figure S8**). Calculation of the molecular weight of the guluronate -PEG conjugate used in the AFM experiments was carried out using the linear fits to the two polymer series in proportion to the contribution of each polymer to the overall molecular weight of the conjugate. The PEG alone was retained at t=25.82 mins, corresponding to a mass of 2500 Da and the guluronates were retained at t = 24.68 mins, corresponding to a mass of 1750 Da.

**Figure S8**. Plot of retention time vs. log MW for a series of PEG-maltose conjugates (red triangles) and a series of guluronate oligomers (orange diamonds).

Guluronate-PEG conjugates were then run on the column, giving rise to the retention profiles presented in **figure S9**. Retention times for the PEG-alginate conjugate are shorter than for PEG alone, and reach a maximum after 48 hours. The retention time reached after 72 hours is 24.2 minutes, which for a conjugate consisting of 60% w/w PEG and 40% w/w alginate oligomer corresponds to a mass of approximately 3600 Da. This value is in reasonable agreement (85%) with the predicted mass of the conjugate.



**Figure S9.** Retention times of a mannuronic acid decamer and its conjugates with a 2500 MW diamino-functionalised PEG. Blue trace is mannuronate alone, red, purple and green traces are the conjugate after 24, 48 and 72 hours reaction time.

**References**

[S1] A. Janshoff, M. Neitzert, Y. Obersörfer, and H. Fuchs, *Angew. Chem. Int. Ed.,* 2000, **39**, 3212.

[S2] A. Dunlop, J. Wattoom, E. A. Hasan, T. Cosgrove and A. N. Round, *Nanotechnology*, 2008, **19**, 345706

[S3] F. Oesterhelt, M. Rief, and H. E. Gaub, *New Journal of Physics* 1999, **1**, 6.

[S4] S. Guo, N. Li, N. Lad, S. Desai, and B. B. Akhremitchev, *Journal of Physical Chemistry C*, 2010, **114**, 8755.

[S5] R. W. Friddle, A. Noy, and J. J. De Yoreo, *PNAS*, 2012, **109**, 13573.

[S6] A. Noy, and R. W. Friddle, *Methods*, 2013, **60**, 142.

[S7] T. Auletta, M. R. de Jong, A. Mulder, F. C. J. M. van Veggel, J. Huskens, D. N. Reinhoudt, S. Zou, S. Zapotoczny, H. Schonherr, G. J. Vancso and L. Kuipers, *J Am Chem Soc*, 2004, **126**, 1577–1584.

[S8] S. Qamar, P. M. Williams and S. M. Lindsay, *Biophys J*, 2008, **94**, 1233–1240.

[S9] M. V. Rekharsky and Y. Inoue, *Chemical Reviews*, 1998 **98(5)**, 1875–1918.

[S10] C. Yang, C., L. Liu, L., T.-W. Mu and Q.-X. Guo, *Analytical Sciences*, 2000 **16(5),** 537–539.

[S11] H. J. Buschmann, K. Jansen and E. Schollmeyer, *Journal of inclusion …*, 2000.

[S12] G. Wenz, B. Han and A. Muller, *Chem Rev*, 2006, **106**, 782–817

[S13] V. Privalko, *Macromolecules*, 1980, **13**, 370–372