(*18*). All these structures consist of long, amphipathic α helices that pack at a 20° angle, thus sharing many structural requirements with coiled-coil helices.

The search revealed several transcriptional activators that do not contain a basic DNA binding region or a leucine zipper but appear to have a zipperlike coiled-coil region. These include the homeobox-containing protein Ubx, the zinc finger protein Rpt1, and the prokaryotic transcriptional activator FixJ (Fig. 2). Ubx and Rpt1 have a topology similar to leucine zipper proteins, with the DNA binding domain followed at its COOH-terminus by the predicted coiled coil, whereas in FixJ the coiled coil precedes the DNA binding COOH-terminal domain (*19*).

In several proteins (Fig. 2), predicted coiled-coil segments lie in areas that are thought to play a functionally important role. For instance, in *Escherichia coli* alanyl tRNA synthetase, a predicted coiled coil lies in a segment that has been identified by deletion analysis as being responsible for oligomerization (*20*); in bacterial chemoreceptors, two predicted coiled-coil regions coincide with the methylated domains that have been implicated in sensory adaptation (*21*); and, in bacterial flagellins, the predicted coiled-coil domains are at the $NH_2$- and COOH-termini of the proteins in regions that are thought to mediate the polymerization of flagellin into the flagellar filament (*22*). The score profiles presented in Fig. 2 indicate that zipperlike coiled coils occur in proteins that mediate many different types of biological processes.


REFERENCES AND NOTES

1. C. Cohen and D. A. D. Parry, *Trends Biochem. Sci.* **11**, 245 (1986); *Proteins Struct. Funct. Genet.* **7**, 1, (1990).
2. D. A. D. Parry, *Biosci. Rep.* **2**, 1017 (1982).
3. The database was constructed from a translated version of GenBank (release 57, September to November 1988, 2,268,298 residues). Redundant entries were eliminated. Myosin sequences were aligned with nematode myosin [A. D. McLachlan and J. Karn, *J. Mol. Bio.* **164**, 605 (1983)], and the rod sequences were extracted (8444 residues). Keratin sequences were aligned with the 59-kD epidermal keratin from mouse [P. M. Steinert, R. H. Rice, D. R. Roop, B. L. Truss, A. C. Steven, *Nature* **302**, 794 (1983); P. M. Steinert and D. A. D. Parry, *Annu. Rev. Cell Biol.* **1**, 41 (1985)], and the four coiled-coil domains were extracted (5762 residues). All tropomyosin sequences (3293 residues) were extracted. These three groups of proteins were selected because their coiled-coil structure is well established and their sequences from several different organisms are known.
4. E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243**, 538 (1989); T. G. Oas, L. P. McIntosh, E. K. O'Shea, F. W. Dahlquist, P. S. Kim, *Biochemistry* **29**, 2891 (1990).
5. This procedure is an extension of a method proposed by Parry (*2*). We obtained the geometric average by multiplying the frequencies for the 28 residues in the window and taking the 28th root. This average was chosen because it gives high and low frequencies equal weight. In an arithmetic average, the score is dominated by high frequencies.
6. Each score is derived from a 28-residue sequence with a specified heptad repeat frame. For sequences with residues having low scores, corresponding frames change frequently. Within regions with high scores, a continuous frame is generally maintained. No frame changes were observed in tropomyosin, three changes were observed in the myosin rod, and one change was observed in keratin helix 2B. These changes corresponded to the three skip residues predicted in myosin and to the stutter predicted in keratin helix 2B (*3*). Although frame changes are generally accompanied by significant changes in score, this often does not happen when the frame continues unbroken after a skip residue or "stutter." In those cases, changes in frame are the only indicators of local discontinuities.
7. Using the coordinates obtained from the Protein Data Bank (Brookhaven National Laboratories, January 1989) we constructed the database of globular proteins. We eliminated all multiple entries or mutant forms of the same protein and the proteins tropomyosin and influenza hemagglutinin, which contain coiled coils. Our final database contained 150 proteins and 32,588 residues. We used the random number generator of a Phoenix computer to construct the database of random sequences. This database has the same overall amino acid composition as GenBank (Table 1) and contains 52,224 residues.
8. $G(x) = (\sigma \cdot 2\pi)^{-1} e^{-0.5((x-m)/\sigma)2}$ where $x$ = score, $m$ = mean, $\sigma$ = standard deviation. Justification for approximating score distributions with Gaussian curves is taken from the close fit of the score distribution for random sequences to its Gaussian curve (Fig. 1C).
9. A. McLachlan and J. Karn, *Nature* **299**, 226 (1982); W. F. Harrington and M. E. Rodgers, *Annu. Rev. Biochem.* **53**, 35 (1984).
10. We analyzed the leucine zipper proteins c-Myc, Jun, Fos, and C/EBP [W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988)], as well as CREB [J. P. Hoeffler, T. E. Meyer, Y. Yun, J. L. Jameson, J. F. Habene, *ibid.* **242**, 1430 (1988)], Cys3 [J. H. Fu, J. V. Paietta, D. G. Mannix, G. A. Marzluf, *Mol. Cell. Biol.* **9**, 1120 (1989)], Opaque2 [H. Hartings *et al.*, *EMBO J.* **8**, 2795 (1989)], Bsg25D (P. D. Boyer, P. A. Mahoney, J. A. Lengyel, *Nucleic Acids Res.* **15**, 2309 (1987)], v-Maf [M. Nishizawa, K. Kataoka, N. Goto, K. T. Fujiwara, S. Kawai, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7711 (1989)], and HBP1 [T. Tabata *et al.*, *Science* **245**, 965 (1989)]; their high scores ranged between 1.72 (v-Maf) and 1.88 (Opaque2) (in all cases, $P(S) > 0.99$). The region of high scores often included sequences outside the leucine repeat and did not always overlap the entire repeat.
11. F. D. Hong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5502 (1989).
12. R. G. Clerc, L. M. Corcoran, J. H. LeBowitz, D. Baltimore, P. A. Sharp, *Genes Dev.* **2**, 1570 (1988).
13. M. Yaneva, J. Wen, A. Ayala, R. Cook, *J. Biol. Chem.* **264**, 13407 (1989).
14. M. E. Maxon, J. Wigboldus, N. Brot, H. Weissbach, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7076 (1990).
15. To evaluate the reliability of coiled-coil prediction from a fourfold leucine heptad repeat, we extracted all the proteins from GenBank that contained this motif. After eliminating redundant entries we obtained 194 proteins, only 70 of which had $P(S)$ values larger than 0.5.
16. S. Cusack, C. Berthet-Colominas, M. Haertlein, N. Nassar, R. Leberman, *Nature* **347**, 249 (1990).
17. D. W. Speicher and V. T. Marchesi, *ibid.* **311**, 177 (1984).
18. D. Freymann *et al.*, *FASEB J.* **4**, A2166 (1990); D. Freymann, P. Metclaf, M. Turner, D. C. Wiley, *Nature* **311**, 167 (1984).
19. M. David *et al.*, *Cell* **54**, 671 (1988).
20. M. Jasin, L. Regan, P. Schimmel, *ibid.* **36**, 1089 (1984).
21. J. Stock, G. Lukat, A. Stock, *Annu. Rev. Biophys. Chem.* **20**, 109 (1991).
22. O. V. Fedorov, A. S. Kostyukova, M. G. Pyatibratov, *FEBS Lett.* **241**, 145 (1988); G. Kuwajima, *J. Bacteriol.* **170**, 3305 (1988).
23. K. Kornfeld *et al.*, *Genes Dev.* **3**, 243 (1989).
24. A VAX Pascal program implementing the described algorithm is available from the authors upon request.



We thank J. M. Lupas for help with the manuscript and D. Welsh for computer assistance. Supported by NIH grant AI20980.

8 August 1990; accepted 13 February 1991


# Self-Assembled Organic Monolayers: Model Systems for Studying Adsorption of Proteins at Surfaces


KEVIN L. PRIME AND GEORGE M. WHITESIDES*



**Self-assembled monolayers (SAMs) of ω-functionalized long-chain alkanethiolates on gold films are excellent model systems with which to study the interactions of proteins with organic surfaces. Monolayers containing mixtures of hydrophobic (methyl-terminated) and hydrophilic [hydroxyl-, maltose-, and hexa(ethylene glycol)–terminated] alkanethiols can be tailored to select specific degrees of adsorption: the amount of protein adsorbed varies monotonically with the composition of the monolayer. The hexa(ethylene glycol)–terminated SAMs are the most effective in resisting protein adsorption. The ability to create interfaces with similar structures and well-defined compositions should make it possible to test hypotheses concerning protein adsorption.**


UNDERSTANDING THE MECHANISM of protein adsorption at surfaces (*1, 2*) is an important element of research in protein chromatography (*3*), clinical diagnostics (*4*), biomedical materials (*5*), and cellular adhesion (*6*). No system is available that permits the structure and properties of the interface to be controlled in detail sufficient for the investigation of hypotheses concerning protein adsorption at the molecular level. We report a study of protein adsorption at interfaces between SAMs and aqueous buffer solutions. The results indicate that the organic interfaces prepared by the self-assembly of long-chain


Department of Chemistry, Harvard University, Cambridge, MA 02138.

*To whom correspondence should be addressed.

alkanethiols onto gold are suitable model systems for the study of protein adsorption at interfaces.

We prepared the SAMs by the chemisorption of alkanethiols from 0.25 mM solutions in ethanol or methanol onto thin (200 ± 20 nm) gold films supported on silicon wafers (*7*). In SAMs derived from ω-substituted alkane-1-thiols [$R(CH_2)_nSH$, $n \geq 10$, where R is a small functional group], the molecules pack densely on the gold surface in a predominantly *trans*-extended conformation, with the axes of the polymethylene chains at an average cant of ≈30° from the surface normal (*8*). The internal domains of these monolayers are pseudo-crystalline; the chain termini are less ordered (*9*). One can control the interfacial properties of these monolayers by changing the tail group, R. SAMs comprising mixtures of two or more components can be prepared by adsorption from solutions containing mixtures of these components: the components of such "mixed SAMs" are not segregated into macroscopic islands (*10*). This combination of a uniform substrate and the ability to control the composition—and to some degree the structure—of the interface at the molecular scale have made SAMs excellent systems with which to study the physical-organic chemistry of organic interfaces.

We used five alkanethiols, $R(CH_2)_{10}SH$: R = $HOCH_2$–, **1** (*10*); R = Glc–α(1,4)–Glc–β(1)–O–, Glc = glucose, **2** (*11*); R = $HO(CH_2CH_2O)_6CH_2$–, **3** (*12*); R = H–, **4** (*13*); and R = $CH_3$–, **5** (*10*). The SAMs derived from **1**, **2**, and **3** model three materials that resist the adsorption of proteins: hydroxylated polymers such as poly(hydroxyethyl methacrylate) (*14*), agarose (*15*), and polymers containing poly(ethylene oxide) (*16*), respectively. For each model system, we prepared a series of mixed SAMs (*10*) from a hydrophilic alkanethiol (*17*) (**1**, **2**, or **3**) and a hydrophobic alkanethiol (**5** with **1** and **3**; **4** with **2**). The structures of these mixed SAMs are shown schematically in Fig. 1. We calculated the mole fraction of hydrophilic alkanethiolate in each mixed SAM, χ, by normalizing the intensity of the O(1*s*) x-ray photoelectron peak obtained from the mixed SAM to that of a SAM containing only the hydrophilic component and by assuming that this normalized intensity is directly proportional to the number of oxygen atoms in the SAM. In the case of SAMs formed from mixtures of **3** and **5**, the intensity of the O(1*s*) peak is linearly proportional to the ellipsometric thickness of the SAM (*12*); this observation is strong evidence that our assumption is valid for the other two cases.

We examined the adsorption of five well-characterized proteins, ribonuclease A (RNase A), pyruvate kinase, fibrinogen, lysozyme, and chymotrypsinogen (*18*), on

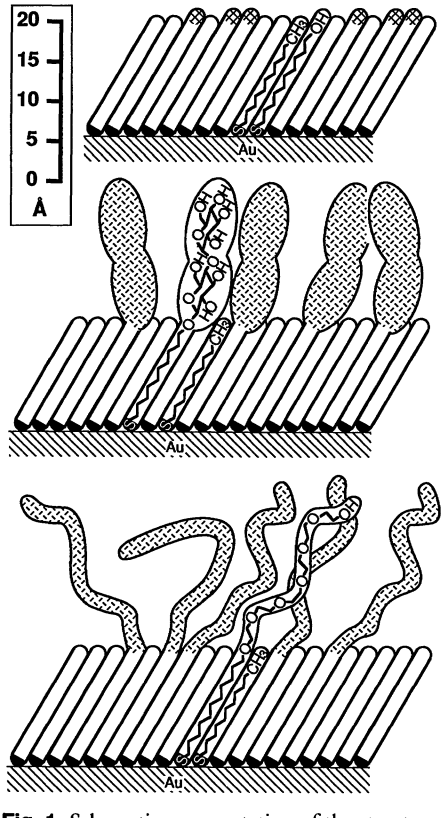


**Fig. 1.** Schematic representation of the structures of mixed monolayers of $HO(CH_2)_{11}SH$ and $CH_3(CH_2)_{10}SH$ (top), of Glc–α(1,4)–Glc–β(1)–$O(CH_2)_{10}SH$ and $CH_3(CH_2)_9SH$ (middle), and of $HO(CH_2CH_2O)_6(CH_2)_{11}SH$ and $CH_3(CH_2)_{10}SH$ (bottom). The ethylene glycol chains in the lower structure are flexible but probably prefer a helical conformation when in contact with water (*32*). The areas of the hatched regions are roughly proportional to the cross-sectional areas of the polar tail groups. The scale bar is approximate and applies to all three illustrations.

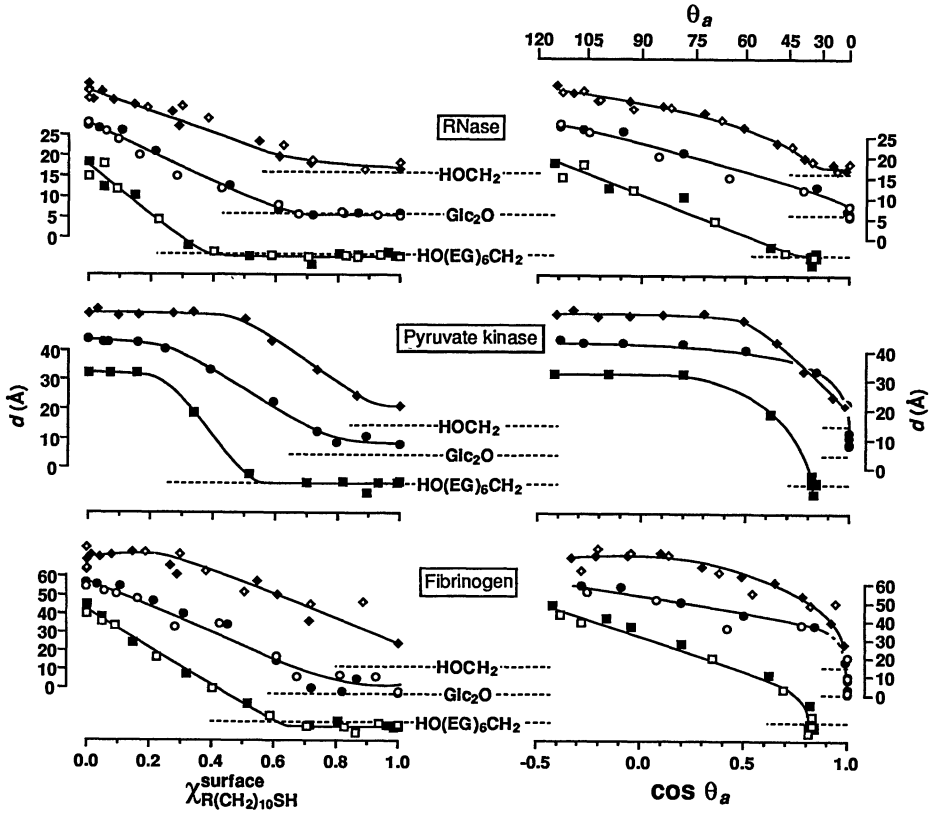


**Fig. 2.** Adsorption of proteins to mixed SAMs varies monotonically with the composition of the SAM. The thickness, *d*, of the adsorbed film of RNase A (top), pyruvate kinase (middle), and fibrinogen (bottom) on mixed SAMs containing $HO(CH_2)_{11}SH$ and $CH_3(CH_2)_{10}SH$ (diamonds, R = $HOCH_2$), Glc–α(1,4)–Glc–β(1)–$O(CH_2)_{10}SH$ and $CH_3(CH_2)_9SH$ (circles, R = $Glc_2O$, Glc = glucose), or $HO(CH_2CH_2O)_6(CH_2)_{11}SH$ and $CH_3(CH_2)_{10}SH$ (squares, R = $HO(EG)_6CH_2$, EG = ethylene glycol, $-OCH_2CH_2-$) is plotted as a function of the composition (left) and wettability (right) of the SAM. The filled and hollow symbols represent data derived from two independent experiments. The values of *d* were determined by ellipsometry and represent the average of three measurements made at different positions on a single sample. The standard deviations of the observed values of *d* are no larger than the symbols representing the data. The values of χ, the mole fraction of $R(CH_2)_{10}S$ on the surface, were measured before protein adsorption. Each value is the intensity of the O(1*s*) x-ray photoelectron peak of the SAM, normalized to $\chi^{surface}_{R(CH_2)_{10}SH} = 1$ for a SAM containing only $R(CH_2)_{10}S$. The values of $\theta_a$ are the maximum advancing contact angles of water (*10, 30*) on the SAM before protein adsorption. The data are offset vertically for clarity; the dashed lines show the location of $d = 0$ Å (no adsorbed protein) for each series of mixed SAMs. The solid curves organize the data visually but do not represent an attempt to model the data.

these mixed SAMs (*19*). The results for RNase, fibrinogen, and pyruvate kinase are summarized in Fig. 2 (*20*). We measured the thickness, *d*, of the adsorbed protein film on each SAM by ellipsometry, treating the film as a homogeneous layer of uniform thickness with a refractive index of 1.45 (*21*). Any difference between the real refractive index of the adsorbed protein and 1.45 results in a systematic error in the calculated thickness but does not change the relative values or the conclusions. The calculated values of thickness are accurate to within ≈25% (*22*).

The data in Fig. 2 point to several conclusions. (i) The system comprising proteins adsorbed on SAMs of alkanethiolates on gold generates reproducible data concerning the extent of protein adsorption. The standard deviations of measurements of *d* taken on several independently prepared samples are within the range of 1 to 4 Å, near the 1 to 2 Å limit of ellipsometry. The N(1*s*) photoelectron signals from adsorbed films of chymotrypsinogen correlate well with the values of *d* determined by ellipsometry (*23*). This observation suggests that variability in the refractive indices of the adsorbed proteins, which would cause nonuniform errors in the calculation of *d*, are not important in this system. (ii) SAMs containing high concentrations of **3** prevent adsorption of the five proteins examined, including fibrinogen. SAMs containing high concentrations of **2** nearly eliminate the adsorption of fibrinogen and pyruvate kinase and prevent adsorption of the other proteins examined. (iii) The observed value of the thickness of the adsorbed protein layer on the hydrophobic, methyl-terminated surface (**4** or **5**; $\chi = 0$ in Fig. 2) corresponds approximately to that expected for a monolayer of native protein (*24–27*). Consistent with others' observations (*28*), multilayers of protein appear not to form. (iv) There is only a general correlation between the interfacial free energy of the SAM [as measured by $\cos \theta_a$, the cosine of the maximum advancing contact angle of water on the SAM (*29*)] and *d*. Although within a set of SAMs derived from the same components more hydrophobic surfaces adsorb greater quantities of protein, the thickness of the adsorbed protein film at any given interfacial free energy differs for each hydrophilic component. For example, when $\theta_a = 34°$, proteins do not adsorb to SAMs containing $HO(CH_2CH_2O)_6$– groups but do adsorb to SAMs containing Glc–α(1,4)–Glc–β(1)–O– or $HOCH_2$– groups. The same effect is observed when the values of *d* for different proteins on SAMs of equal receding contact angle, $\theta_r$, are compared.

From this limited set of data, it is premature to infer mechanisms of adsorption of proteins at interfaces. The observation that adsorption increases as hydrophobicity increases (for a given set of components) is expected and consistent with the idea that hydrophobic interactions are important in protein adsorption. The observation that $HO(CH_2CH_2O)_6$– groups are especially effective in preventing protein adsorption suggests that steric stabilization—a phenomenon commonly used to explain the stability of colloidal suspensions in the presence of polymers (*30*)—is important in preventing protein adsorption (*31*). The extent to which entropic repulsion (*30*) contributes to the steric stabilization is not clear and may vary with χ: the steric requirements of packing in the SAM should reduce the conformational entropy of the $HO(CH_2CH_2O)_6$– groups as their concentration in the SAM increases. We believe that SAMs are the best defined systems now available for examining the interactions of proteins and surfaces and that they will provide the means to test many of the current hypotheses regarding the mechanisms of these interactions.

## REFERENCES AND NOTES


1. J. D. Andrade and V. Hlady, *Adv. Polym. Sci.* **79**, 1 (1986).
2. J. L. Brash and T. A. Horbett, Eds., *Proteins at Interfaces* (ACS Symposium Series 343, American Chemical Society, Washington, DC, 1987).
3. H. P. Jennissen, *Ber. Bunsenges. Phys. Chem.* **93**, 948 (1989); J. Porath, *Biotechnol. Prog.* **3**, 14 (1987); T. Mizutani, *J. Liq. Chromatogr.* **8**, 925 (1985).
4. J. W. Boretos, in *Synthetic Biomedical Polymers: Concepts and Applications*, M. Szycher and W. J. Robinson, Eds. (Technomic, Westport, CT, 1980), pp. 187–200.
5. V. I. Sevastianov, *Crit. Rev. Biocompat.* **4**, 109 (1988); T. A. Horbett, in *Hydrogels in Medicine and Pharmacy*, N. A. Peppas, Ed. (CRC, Boca Raton, FL, 1986), vol. 1, pp. 127–172; B. Ivarsson and I. Lundström, *Crit. Rev. Biocompat.* **2**, 1 (1986); J. L. Brash, *Makromol. Chem. Suppl.* **9**, 69 (1985).
6. J. M. Schakenraad and H. J. Busscher, *Colloids Surf.* **42**, 331 (1989).
7. For a review, see G. M. Whitesides and P. E. Laibinis, *Langmuir* **6**, 87 (1990).
8. R. G. Nuzzo, L. H. Dubois, D. L. Allara, *J. Am. Chem. Soc.* **112**, 558 (1990).
9. R. G. Nuzzo, E. M. Korenic, L. H. Dubois, *J. Chem. Phys.* **93**, 767 (1990); C. E. Chidsey, G. Y. Liu, P. Rowntree, G. Scoles, *ibid.* **91**, 4421 (1989).
10. We use "macroscopic" to mean sufficiently large, perhaps more than 10 nm in diameter, that the properties of the monolayer are determined by molecules of each component in environments indistinguishable from the environment in a pure monolayer of that component. A number of experimental results suggest that macroscopic islands do not form [C. D. Bain, J. Evall, G. M. Whitesides, *J. Am. Chem. Soc.* **111**, 7155 (1989); C. D. Bain and G. M. Whitesides, *ibid.*, p. 7164]. The advancing contact angles of water on SAMs containing mixtures of long and short hydroxyl-terminated chains [$HO(CH_2)_{19}SH$ and $HO(CH_2)_{11}SH$, respectively] reach a pronounced maximum ($\theta_{adv} = 40°$) near $\chi^{surface}_{HO(CH_2)_{11}SH} = 0.5$ [C. D. Bain and G. M. Whitesides, *Science* **240**, 62 (1988)]. Because pure SAMs of either hydroxyl-terminated chain are wet by water ($\theta_{adv} < 15°$), this increase in the contact angle is inconsistent with the presence of macroscopic islands, which would also wet, and supports the hypothesis that the chains in two-component SAMs are well mixed at length scales near molecular dimensions.
11. β-(9-Decenyl)maltoside heptaacetate was prepared from commercial maltose octaacetate and 9-decen-1-ol [P. Rosevar, T. VanAken, J. Baxter, S. Ferguson-Miller, *Biochemistry* **19**, 4108 (1980)]. β-(10-Mercaptodecyl)maltoside octaacetate was prepared by the photochemical addition of thiolacetic acid to the olefin (*12*). We achieved deacetylation by allowing the acetate to stand in a 2:1:1 mixture of methanol, water, and triethylamine. The products were purified by flash chromatography (silica gel, methanol-chloroform) after each step. The overall yield was ≈60% from maltose octaacetate.
12. C. Pale-Grosdemange, E. S. Simon, K. L. Prime, G. M. Whitesides, *J. Am. Chem. Soc.* **113**, 12 (1991).
13. Decane-1-thiol (Aldrich Chemical Company) was purified by flash chromatography (silica gel, hexane) before use.
14. J. L. Bohnert, T. A. Horbett, B. D. Ratner, F. H. Royce, *Invest. Ophthalmol. Vis. Sci.* **29**, 363 (1988).
15. E. Sada, S. Katoh, T. Inoue, M. Shiozawa, *Biotechnol. Bioeng.* **27**, 514 (1985); T. P. West, I. Maldonado, E. R. Kantrowitz, *Biochim. Biophys. Acta* **839**, 32 (1985); G. Halperin, M. Tauber-Finkelstein, S. Shaltiel, *J. Chromatogr.* **317**, 103 (1984).
16. G. Fleminger, B. Solomon, T. Wolf, E. Hadas, *J. Chromatogr.* **510**, 271 (1990); K. Allmer, J. Hilborn, P. H. Larsson, A. Hult, B. Raanby, *J. Polym. Sci. Part A* **28**, 173 (1990); D. W. Grainger, T. Okano, S. W. Kim, *J. Colloid Interface Sci.* **132**, 161 (1989).
17. We use the term "hydrophilic/hydrophobic alkanethiol" as a convenient shorthand for "alkanethiol that forms a hydrophilic/hydrophobic interface when adsorbed on gold."
18. α-Chymotrypsinogen A (type II from bovine pancreas, Sigma), lysozyme (E.C. 3.2.1.17, grade III from chicken egg white, Sigma), RNase A (E.C. 3.1.27.5, type III-A from bovine pancreas, Sigma), pyruvate kinase (E.C. 2.7.1.40, type PK-3 from rabbit muscle, Biozyme), and fibrinogen (fraction I from human plasma, Sigma) were used as received.
19. The following protocol was used in these experiments: the SAM (7) was rinsed with ethanol and blown dry with nitrogen. It was immersed in an unstirred solution of protein (1 mg/ml in 10 mM aqueous phosphate buffer, pH 7.5, 23°C, 1 hour) under air. The SAM was then removed from solution, rinsed six times with 2-ml aliquots of distilled, deionized water applied as a stream from a Pasteur pipette, and blown dry with a stream of nitrogen passed through another Pasteur pipette at a pressure of 10 psi (70 kPa) above atmospheric pressure. Using this protocol, we found that our results were insensitive to small variations in the rinsing procedure and in the time of immersion.
20. Our choice of proteins for Fig. 2 was made to illustrate representative adsorption isotherms. A suitably scaled plot of the data obtained from lysozyme would superimpose upon the data from fibrinogen. Likewise, the data for chymotrypsinogen on the series prepared from **2** and from **3** are nearly indistinguishable from the data for RNase; chymotrypsinogen and pyruvate kinase show behavior nearly indistinguishable from each other on mixed SAMs prepared from **1**.
21. The refractive index of an adsorbed protein film is effectively constant once the adsorption plateau is reached [P. A. Cuypers *et al.*, *J. Biol. Chem.* **258**, 2426 (1983)]. We use $n = 1.45$ because it is near the average of the reported values of *n* for a number of proteins [range, 1.34 to 1.71 (*33*)] and it corresponds to the value of *n* used in earlier studies of SAMs (7).
22. Choosing $n = 1.33$ or $n = 1.71$ causes a 25% increase or decrease, respectively, in the calculated values of *d*, relative to the values obtained for $n = 1.45$.
23. The photoelectron signal observed from an atom in a thin film is proportional to $1 - e^{(-d/\lambda \sin \theta)}$, where *d* is the thickness of the film, λ is the escape depth of the photoelectron through the film, and $\sin \theta$ is the angle between the surface normal and the analyzer [C. D. Bain and G. M. Whitesides, *J. Phys. Chem.* **93**, 1670 (1989)]. The value of λ for protein films is

not known; determination of an accurate value of λ requires continuous films of different thicknesses. This criterion is not met by the probably discontinuous films formed in this study. Thus, only a qualitative comparison of results from ellipsometry and x-ray photoelectron spectroscopy (XPS) is possible at this time. Using the values of $d$ obtained by ellipsometry and the intensities of the N(1$s$) peaks, we found that least-squares fitting to the above equation yielded a fit with $r^2 = 0.98$, $n = 11$.

24. The averages and standard deviations of the observed thicknesses of adsorbed protein films on methyl-terminated SAMs were 21 ± 1 Å (RNase), 58 ± 3 Å (fibrinogen), and 38 ± 1 Å (pyruvate kinase). Each value represents an average of 18 measurements. Three measurements were made from different positions on each of six independently prepared samples to derive these values.
25. RNase A (molecular weight ≈13,700) forms monoclinic crystals with one molecule per unit cell. The parameters of the unit cell are $a$ = 30 Å, $b$ = 38 Å, $c$ = 53 Å, β = 106° [A. Wlodawer, L. A. Svensson, L. Sjoelin, G. L. Gilliland, *Biochemistry* **27**, 2705 (1988)].
26. Pyruvate kinase is a tetrameric enzyme of identical subunits; the dimensions of the tetrameric molecule are 75, 95, and 125 Å. The molecular weight of each subunit is 54,600 in yeast and 57,900 in cat muscle. (In this study we used enzyme isolated from rabbit muscle.) The shape of each subunit is approximated by an ellipsoid 75 Å long with a maximum transverse diameter of 45 Å [H. Muirhead, *Biol. Macromol. Assem.* **3**, 143 (1987)]. We do not know whether the molecule retains its tetrameric structure upon adsorption.
27. Human fibrinogen (molecular weight ≈340,000) is a structurally complex protein containing several domains connected by more flexible segments [J. A. Shafer and D. L. Higgins, *Crit. Rev. Clin. Lab. Sci.* **26**, 1 (1988)]. It is, therefore, difficult to provide specific molecular parameters for comparison with our experimental results.
28. F. MacRitchie, *Adv. Protein Chem.* **32**, 283 (1978).
29. R. H. Dettre and R. E. Johnson, *J. Phys. Chem.* **69**, 1507 (1965).
30. Steric stabilization of hydrophobic colloids in aqueous solution occurs when a hydrophilic polymer is adsorbed at the colloid-water interface. This polymer prevents flocculation of two colloid particles in two ways. First, approach of the particles to a distance such that the strength of the attractive hydrophobic interaction rises above $kT$ (where $k$ is the Boltzmann constant and $T$ is temperature) is inhibited enthalpically by changes in the configuration of the polymer and perhaps by its desolvation as it is compressed. Second, for coiled polymers, loss of conformational entropy due to the approach of the two particles to one another creates a repulsive force that helps to oppose the attractive hydrophobic interaction. We refer to the second effect as "entropic repulsion," consistent with D. H. Everett [*Basic Principles of Colloid Science* (Royal Society of Chemistry, London, 1988), pp. 45–50].
31. For a computational treatment of surface-grafted oligo(ethylene glycol) chains, see M. Björling, P. Linse, G. Karlström, *J. Chem. Phys.* **94**, 471 (1990); G. Karlström, *ibid.* **89**, 4962 (1985). For a theoretical consideration of the protein resistance of surface-grafted poly(ethylene glycol) surfaces, see S. I. Jeon, J. H. Lee, J. D. Andrade, P. G. de Gennes, *J. Colloid Interface Sci.* **142**, 149 (1991); S. I. Jeon and J. D. Andrade, *ibid.*, p. 159.
32. H. Matsuura, K. Fukuhara, S. Masatoki, M. Sakakibara, *J. Am. Chem. Soc.* **113**, 1193 (1991); H. Matsuura and K. Fukuhara, *J. Mol. Struct.* **126**, 251 (1985).
33. For several examples, see T. Arnebrant, B. Ivarsson, K. Larsson, I. Lundström, T. Nylander, *Prog. Colloid Polym. Sci.* **70**, 62 (1985); P. A. Cuypers, W. T. Hermens, H. C. Hemker, *Anal. Biochem.* **84**, 56 (1978).
34. Supported by the National Science Foundation (NSF) under the Engineering Research Center Initiative to the Massachusetts Institute of Technology Biotechnology Process Engineering Center (cooperative agreement CDR-88-03014), by the Office of Naval Research, and by the Defense Advanced Research Projects Agency (DARPA). The x-ray photoelectron spectrometer was provided by DARPA through the University Research Initiative and is housed in the Harvard University Materials Research Laboratory, an NSF-funded facility. Initial research in this project was carried out by C. Palegrosdemange. We thank H. Biebuyck, J. Folkers, A. Jain, P. Laibinis, and T. R. Lee for critical readings of the manuscript. K.L.P. was an NSF predoctoral fellow, 1986 to 1989.

17 December 1990; accepted 13 March 1991

# Arginine-Mediated RNA Recognition: The Arginine Fork

BARBARA J. CALNAN, BRUCE TIDOR, SARA BIANCALANA, DEREK HUDSON, ALAN D. FRANKEL*

**Short peptides that contain the basic region of the HIV-1 Tat protein bind specifically to a bulged region in TAR RNA. A peptide that contained nine arginines ($R_9$) also bound specifically to TAR, and a mutant Tat protein that contained $R_9$ was fully active for transactivation. In contrast, a peptide that contained nine lysines ($K_9$) bound TAR poorly and the corresponding protein gave only marginal activity. By starting with the $K_9$ mutant and replacing lysine residues with arginines, a single arginine was identified that is required for specific binding and transactivation. Ethylation interference experiments suggest that this arginine contacts two adjacent phosphates at the RNA bulge. Model building suggests that the arginine η nitrogens and the ε nitrogen can form specific networks of hydrogen bonds with adjacent pairs of phosphates and that these arrangements are likely to occur near RNA loops and bulges and not within double-stranded A-form RNA. Thus, arginine side chains may be commonly used to recognize specific RNA structures.**

RNA-PROTEIN INTERACTIONS ARE important for many regulatory processes, but little is known about the details of sequence-specific recognition. From what is known, it appears that both RNA structure and nucleotide sequence function in recognition. The crystal structure of the glutaminyl tRNA synthetase–tRNA complex (*1*) has shown that specific contacts are made between amino acid side chains and bases in non–base paired regions of the RNA, while studies of the R17 coat protein (*2*) have suggested that the overall three-dimensional RNA conformation contributes substantially to recognition. Recently, an arginine-rich RNA-binding motif has been identified in several RNA-binding proteins (*3*), including the human immunodeficiency virus (HIV) Tat protein. Peptides that contain this region of Tat bind specifically to an RNA stem-loop structure named TAR (*4, 5*), which is located in the HIV long terminal repeat, and RNA binding is essential for Tat-dependent transcriptional activation (*5*). The overall charge density of the Tat peptides is important for binding, however, the amino acid sequence requirements are flexible; the sequence can be scrambled and still bind specifically to TAR (*5*).

The basic RNA-binding region of Tat, RKKRRQRRR (residues 49 to 57), is nine amino acids long and contains a glutamine at position 54 that is not essential for binding or activity (*5*). Because it is known that a high positive charge density is important for RNA binding, we synthesized (*6*) two peptides, $R_9$, which contains a stretch of nine adjacent arginines (with a tyrosine at the $NH_2$-terminus and an alanine at the COOH-terminus), and $K_9$, which contains a stretch of nine lysines (and a surrounding tyrosine and alanine), and measured their binding to TAR RNA (*7*). The $R_9$ peptide bound to TAR RNA with the same affinity as the wild-type Tat peptide and with tenfold higher affinity than $K_9$ (Fig. 1). The specificity of $R_9$ binding to TAR was identical to the wild-type peptide, whereas $K_9$ binding was nonspecific (*7*). Because RNA binding of Tat peptides correlates with Tat's function as a transcriptional activator (*5*), we asked whether $R_9$ or $K_9$ could function in the context of the intact protein. The nine– amino acid basic region of Tat was replaced by $R_9$ or $K_9$ in a Tat expression vector, and activation of HIV-1 transcription by the chimeric Tat proteins was tested in transient transfection assays (*8*). The $R_9$-containing protein gave wild-type transactivation activity and was 100-fold more active

B. J. Calnan, B. Tidor, A. D. Frankel, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142.
S. Biancalana and D. Hudson, MilliGen/Biosearch, Division of Millipore, 81 Digital Drive, Novato, CA 94949.

*To whom correspondence should be addressed.