IMAGING SOFT MATERIALS WITH SCANNING TUNNELING MICROSCOPY

J.T. Woodward IV1 and J.A. Zasadzinski2*

1Department of Physics and 2Department of Chemical Engineering,
University of California, Santa Barbara, CA 93106-5080

(Received for publication August 6, 1995 and in revised form March 21, 1996)

Abstract

By modifying freeze-fracture replication, a standard electron microscopy fixation technique, for use with the scanning tunneling microscope (STM), a variety of soft, non-conductive biomaterials can be imaged at high resolution in three dimensions. Metal replicas make near ideal samples for STM in comparison to the original biological materials. Modifications include a 0.1 μm backing layer of silver and mounting the replicas on a fine-mesh silver filters to enhance the rigidity of the metal replica. This is required unless STM imaging is carried out in vacuum; otherwise, a liquid film of contamination physically connects the STM tip with the sample. This mechanical coupling leads to exaggerated height measurements; the enhanced rigidity of the thicker replica eliminates much of the height amplification. Further improvement was obtained by imaging in a dry nitrogen atmosphere. Calibration and reproducibility were tested with replicas of well characterized bilayers of cadmium arachidate on mica that provide regular 5.5 nm steps. We have used the STM/replica technique to examine the ripple shape and amplitude in the P5 phase of dimyrstoylphosphatidylcholines (DMPC) in water. STM images were analyzed using a cross-correlation averaging program to eliminate the effects of noise and the finite size and shapes of the metal grains that make up the replica. The correlation averaging allowed us to develop a composite ripple profile averaged over hundreds of individual ripples and different samples. The STM/replica technique is sufficiently general that it can be used to examine a variety of hydrated lipid and protein samples at a lateral resolution of about 1 nm and a vertical resolution of about 0.3 nm.

*Address for correspondence:
Joseph A. Zasadzinski
Department of Chemical Engineering,
University of California,
Santa Barbara, CA 93106

Telephone number: 805-893-4769
FAX number: 805-893-4731
gorilla@squid.ucsb.edu

Introduction

The scanning tunneling microscope (STM) (Binnig et al., 1982; Quate, 1986) offers exciting new ways of imaging surfaces with resolution to the sub-molecular scale (Hansma and Tersoff, 1987). However, reproducible, three-dimensional STM images of biological or organic material are quite difficult to obtain as such materials are non-conductive and soft (Amrein et al., 1988; Baró, et al., 1985; Foster et al., 1988; Foster and Frommer, 1988; Hansma et al., 1988; Hörber et al., 1988; Lang et al., 1988; Travaglini et al., 1987). Imaging with any probe microscope invariably involves mounting the sample on a substrate, which is often accompanied by its own set of artifacts (Hörber et al., 1988; Lang et al., 1988; Clemmer and Beebe, 1991; Patrick and Beebe, 1993). As a result, little new biological information has been obtained via STM in spite of a significant investment of research time and money. Much of the initial promise in this area has proven false; artifacts abound in the literature because direct STM images of insulating materials are difficult to understand based on any reasonable model of electron tunneling (Lang et al., 1988; Lindsay and Barris, 1988; Spong et al., 1989; Clemmer and Beebe, 1991). However, recent work suggests that many images of biological materials are actually maps of thin layers of loosely bound water with sufficient conductivity to provide a current for imaging (Sonnenfeld and Hansma, 1987; Yuan et al., 1991; Guckenberger et al., 1994). A better explanation might be that images of biological materials are primarily due to electrochemical reactions at the STM tip, as in the scanning electrochemical microscope (Arca et al., 1994). However, such explanations are qualitative and it is difficult to directly relate image contrast to features in the image.
Imaging soft materials with STM

1994; Frisbie et al., 1994), and tapping mode (Hansma et al., 1994; Radmacher et al., 1994). There are many benefits to these techniques, the most important being (1) non-conductors are readily imaged; (2) imaging can be done under solvents or in near in vivo conditions and (3) specific chemical information can be obtained via tip-sample interactions. However, there are several drawbacks to each of these techniques including the relative cost of the equipment, the time it takes to acquire images, and interpretation of specific interactions and relating these interactions to structural features. Moreover, all these force microscopy techniques are still limited to imaging materials bound to a substrate.

An alternative to these various AFM techniques that avoids much of the confusion of direct STM imaging of biomaterials is coating non-conductive surfaces with metal layers so as to make them conductive, then imaging with the STM (Travaglini et al., 1987; Amrein et al., 1988; Zasadzinski et al., 1988; Obcemea and Vidic, 1992; Woodward and Zasadzinski, 1994). We have found that conventional freeze-fracture, which is extremely useful for imaging bulk organic materials with transmission electron microscopy (TEM), can be modified for reliable, high resolution, quantitative three dimensional imaging via the STM. In principle, a platinum replica of a fracture surface is an ideal sample for STM imaging - it is highly conductive, chemically homogeneous, inert, and easily manipulated. However, imaging with the STM is fundamentally different than with the TEM and has led us to modify the freeze-fracture technique to match the different requirements of the STM. The freeze-fracture STM techniques detailed in this paper are an important alternative to conventional AFM imaging of soft or insulating surfaces, and can be used to prepare substrate free samples for high resolution imaging. The technique works best for, and is illustrated by periodic membrane surfaces with various three dimensional contours. As we show here, on repetitive surfaces such as the ripples of the P$\gamma$ phase of dimyristoylphosphatidylcholine (DMPC) in water, with the combination of freeze-fracture replication, careful STM imaging in controlled environments, and correlation averaging image analysis, we can achieve a 1 nm lateral and 0.3 nm vertical resolution. Such a high resolution, quantitative description of soft materials in solution is impossible by any other technique.

Sample Preparation for STM-Freeze-Fracture

Freeze-fracture replication is a four step thermal fixation technique originally developed by Steere (1957) to image thermally fixed as opposed to chemically fixed cells and suspensions. The basic fracture tables and evaporation equipment for the technique was first developed by Moor et al. (1961) (now commercially available from Balzers and others). Branton (1966) was the first to describe in detail the progression of the fracture through various materials including cell membranes. In the modern version of the technique, a thin layer of sample is trapped between two copper plates, quickly frozen, and then fractured under vacuum (Zasadzinski and Bailey, 1989). The fracture surface is replicated by evaporating a thin metal layer obliquely to the fracture surface; the metal film is followed by a thicker carbon layer for strength. Any remaining portions of the original sample is carefully cleaned away, and the replica is then examined under ambient conditions in the transmission electron microscope. In any individual replica of a non-repetitive surface examined in the TEM, the absolute resolution in a particular image is limited by the size of the metal film grains to about 2 nm (Chiruvolu et al., 1994). Image analysis techniques can remove some of the influence of the granularity of the replica, especially for periodic surfaces.

Rapid Freezing

The goal of rapid freezing is to remove heat at such a rate that (1) the details of the individual cell, liposome or vesicle "particles" are retained and (2) the distribution and orientation of the particles are not disturbed. For many systems, the second criterion is much more difficult to achieve, especially for particles less than 0.1 μm in size and usually requires that the solvent or continuous phase be vitrified. In freeze-fracture, vitrification generally is taken to mean that the solvent phase contains no recognizable crystals larger than the typical resolution of the images, although this is far from proving that the solvent phase is truly amorphous. Simple plunge freezing into liquid propane or ethane has been sufficient to vitrify water films up to about 50-100 μm in thickness for freeze fracture. Direct imaging of thin aqueous films at cryogenic temperatures show that the water can indeed be frozen as an amorphous solid without any crystallization and held in this phase indefinitely at low temperatures (Adrian et al., 1984; Dubochet and McDowall, 1981; Dubochet et al., 1982; Bellare et al., 1988; Chiruvolu et al., 1994).

However, if the solvent phase is crystallized, dispersed particles and solutes are swept to the crystalline grain boundaries and information on the original orientation and distribution of the particles in the solvent is lost. The distribution of the particles reflects the crystallization behavior of the solvent rather than any of the original properties of the system. Theoretical estimates of the cooling rates necessary to vitrify water range from $10^4$ to $10^{10}$ K/sec (Bruegeller and Mayer, 1980), although experimentally, the minimum cooling rate to create vitrified water appears to be in the range
of 10^5 K/sec (Dubochet and McDowall, 1981; Dubochet et al., 1982; Chiruvolu et al., 1994). These cooling rates are readily accessible in the laboratory with a minimum investment via a variety of plunge and jet freezing equipment (Bellare et al., 1988; Bailey et al., 1991; Gilkey and Staehelin, 1986; Jahn and Strey, 1988; Müller et al., 1980).

As treatment with chemical cryoprotectants must be avoided, the physical parameters such as the heat capacity, thermal conductivity and crystallization behavior of the biomaterial system to be studied are fixed and usually cannot be modified without inducing chemical or ultrastructural changes in the sample. For freeze-fracturing a broad variety of liquid samples of any viscosity such as suspensions, solutions, and emulsions, the most practical method of rapid freezing is "sandwiching" the sample in a rigid container that is then contacted with a liquid cryogen. A wide variety of sample holders have appeared in the literature (Costello and Corless, 1978); however, the most commonly used are variations on the "copper sandwich" holders developed by Gulik-Krzywicki and Costello (1978) and commercialized by Balzers (BUO-12-056T and variations; Hudson, New Hampshire) to produce complementary replicas. Typically, from 0.1-0.5 ml of sample liquid is pipetted onto one of the planchettes, then a second planchette is used to spread the liquid to form a thin, 10-50 μm thick film. A variant of this sample holder with an annular opening in the top planchette is used by Jahn and Strey (1988). By far the most common method of rapid freezing is immersing the sample sandwich into a liquid cryogen, typically with a spring-loaded device to increase the relative specimen-cryogen velocity (Bailey et al., 1991; Bellare et al., 1988; Jahn and Strey, 1988).

Freezing a multicomponent, structured biomaterial sample sandwiched between metal sheets in a liquid cryogen is a complicated heat transfer process that involves convection at the cryogen-sandwich boundary, conduction through the sandwich, and conduction and possible phase changes, crystallization, etc. within the fluid specimen (Talmon et al., 1981). A simplified model of cooling (Zasadzinski, 1988a) that has been experimentally validated (Bailey and Zasadzinski, 1991) shows that the cooling rate of samples is limited by convection from the cryogen to the specimen surface rather than by conduction through the specimen. The important criterion is that the Biot modulus of the specimen, hV/kA is << 1; (see Zasadzinski, 1988a, for discussion) h is the heat transfer coefficient from the cryogen to the sample, V is the sample volume, A is the surface area of the sample in contact with the cryogen, and k is the average sample thermal conductivity. For a typical freeze-fracture sample of thickness 100-200 μm, the Biot number, Bi, is 0.05 to 0.5. An important physical consequence of convection limited cooling (Bi < 1) is that the temperature is spatially uniform within the sample during freezing (Bailey and Zasadzinski, 1991). This means that there are no spatial temperature gradients in the sample, and hence no driving force for spatial reorganization of the solutes or particles in the specimen, or localized phase separation. In this approximation, the cooling rate of a sample of heat capacity C_p, and density ρ, is:

\[ \frac{dT}{dt} = \frac{-A}{V} k(T - T_c) \frac{1}{\rho C_p} \]  

Equation 1

T is the sample temperature and T_c is the liquid cryogen temperature. Experimental measurements of the average cooling rate of a wide variety of specimens using various cryogens can be correlated using this simplified model (Zasadzinski, 1988a). The cooling rate is proportional to the ratio of specimen surface area to volume; this defines a characteristic sample dimension that governs the cooling rate. The only real dependence of the cooling rate on the sample composition is the inverse dependence on the thermal density, ρC_p, of the sample. The thermal density varies little between the typical materials encountered in freeze-fracture experiments. Surprisingly, the cooling rate is independent of the sample thermal conductivity, and hence, virtually all samples of the same characteristic dimensions freeze at the same rate (Bailey and Zasadzinski, 1991).

Optimization of the freezing process can be achieved by maximizing h, the heat transfer coefficient, while minimizing T_c, the cryogen melting point, by proper choice of cryogen and the velocity at which the cryogen contacts the sample. A limited amount of cryogen boiling enhances the heat transfer coefficient; too much boiling results in the formation of a vapor film around the sample that drastically reduces the rate of heat transfer (Bailey and Zasadzinski, 1991). Liquid nitrogen, and other cryogens at their boiling points should be avoided for this reason. The best practical cryogen appears to be ethane or propane cooled to near its freezing point by liquid nitrogen (Costello and Corless, 1978; Bailey and Zasadzinski, 1991). The distribution and orientation of dispersed tobacco mosaic virus in water has been preserved using both controlled plunge freezing in liquid propane and propane jet freezing (Zasadzinski and Meyer, 1986), although it is impossible to say if the water surrounding the colloidal particles is amorphous or microcrystalline.

Many interesting biological materials exist naturally either above or below room temperature. This requires that a temperature and environment controlled chamber be coupled to the freezing apparatus. Sealed, two stage ovens have been used to observe temperature dependent...
phase transitions in lyotropic nematic micellar phases (Sammon et al., 1987) and to study phospholipid phase transitions (Zasadzinski and Schneider, 1987). Bellare et al. (1988) have constructed a temperature and humidity controlled cell in which samples can be equilibrated prior to plunge freezing and has used it to demonstrate a number of temperature dependent phases. Jahn and Strey (1988) have used a simple spring-loaded sample making device to look at composition and temperature sensitive materials. Thermotropic liquid crystalline phases stable over less than 1°C have been resolved using quick freezing techniques followed by freeze-fracture (Costello et al., 1984; Zasadzinski et al., 1986), provided that the equilibrium phase transition occurs by nucleation and growth. Thermotropic smectic phases have been successfully quenched from above 100°C for freeze-fracture investigation (Ihn et al., 1992).

Because the chemical and physical properties of a biomaterial cannot be optimized for rapid freezing by chemical or physical cryoprotectants without changing the structure, a judicious choice of systems to investigate usually is the difference between success and failure. As avoiding solvent crystallization is of primary importance to successful images, it is useful to understand the solvent properties that affect crystallization. Crystallization consists of two steps, nucleation of small crystals of a critical size, then the growth of these crystals. Except for extremely pure liquids, nucleation occurs heterogeneously at insoluble impurities; such impurities are often the colloidal particles we wish to study, and hence are unavoidably present. The crystal growth velocity, \( u \), is therefore the more important quantity to minimize by appropriate choice of sample properties. The crystal growth velocity is proportional to the degree of supercooling, \( \Delta T \), the entropy of fusion, \( \alpha \), and the fraction, \( f_B \), of acceptable sites on the interface (which reflects the steric constraints involved in packing solvent molecules into a different configuration), and inversely proportional to the solvent viscosity, \( \mu \):

\[
u \propto \frac{f_B \Delta T}{\mu T_m}
\]

(2)

\( T_m \) is the equilibrium melting point.

Steric effects, are generally more pronounced than are viscous effects in Eqn. 2. Branched hydrocarbons such as isohexane ((CH\(_3\))\(_2\)CH(CH\(_2\))\(_2\)CH\(_3\)) freeze at a much lower temperature (-153.7°C) than does straight chain n-hexane (CH\(_3\)(CH\(_2\))\(_5\)CH\(_3\))(-94°C), indicating that the steric restrictions imposed by the branching make crystallization much more difficult. Cyclohexane, on the other hand, adopts a fairly rigid conformation in the liquid phase that is easy to pack into a crystalline lattice and freezes at about 6°C (Roberts and Caserio, 1977).

Clearly, as many of the physical properties of these solvents are similar, to optimize the system for rapid freezing, isohexane, which is sterically hindered, is a much better choice than cyclohexane, which crystallizes readily. In aqueous solutions, salts, solutes and macromolecules that tend to disrupt water structure by hydrogen bonding can hinder crystallization. For instance, ice formation in polyacrylamide gels is suppressed to below -17°C (Tanaka et al., 1977).

Fracture and replication

Once frozen successfully, the sample must be replaced by a metal "replica" that is an accurate map of the biomaterial microstructure and is compatible with the requirements for STM imaging. The fracture, etching, and replication steps of the freeze-fracture technique are carried out at low temperature and high to ultrahigh vacuum. Typically, the "copper sandwich" samples are loaded under liquid nitrogen into a hinged brass block fracture stage. The fracture stage has sufficient thermal mass that the specimens do not heat up significantly during the brief time they are exposed to air during transfer into a vacuum chamber. The fracture stage is clamped to a temperature controlled coldfinger within the vacuum chamber. The specimen stage fractures the sandwiches on opening; a sharp, quick break is preferable to a long, steady pull for reasons discussed below. The stress on the specimen is primarily tensile.

The general behavior of any material under load can be classified as ductile or brittle depending on whether or not the material exhibits plastic deformation. A completely brittle material will fracture at the elastic limit. (The recovery of the original dimensions of a deformed body when the load is removed is known as elastic behavior. The limiting load beyond which a material no longer can recover its shape is known as the elastic limit). If a material is deformed beyond its elastic limit without fracturing, it is said to have undergone plastic deformation. For freeze-fracture replication, the goal is an ideal brittle fracture in which the sample fractures without deformation. An ideal brittle fracture limits deformation to only the molecules being pulled apart because a brittle material cannot redistribute local stresses to regions outside the fracture zone. The stress builds up in a very localized region, a crack forms at one or more points where the stress concentration overcomes the cohesive strength of the materials, then rapidly spreads, fracture the entire sample. It is important to point out that brittleness is not an absolute property; decreased temperature, increased rates of stress and the presence of cracks or notches increase the brittleness of a material (cf., Dieter, 1976). This is why it is necessary to fracture at the lowest practical temperature and to induce fracture by a quickly imposed, large
Griffith (cf., Dieter, 1976) proposed that, even prior to fracture, a brittle material contains a population of small cracks at imperfections, boundaries, and other imperfections. When placed under sufficient stress, one or more of these cracks spreads into a brittle fracture, thereby decreasing the elastic energy at the expense of increasing the surface area, and hence the surface energy of the material. A crack will spread when the decrease in elastic energy is at least equal to the energy required to create a new crack surface. The surface energy can be thought of as the product of the new surface area created and the specific energy per unit area of the fracture surface. The Griffith theory states that the fracture will follow the path of least resistance (smallest specific surface energy or smallest molecular cohesion) provided that the fracture area created is not too large. In most two or more phase colloidal dispersions, the fracture surface propagates along the interface between the two phases, usually at particle-solvent boundaries. Apparently, solvent-solvent cohesion and particle-particle cohesion are larger than solvent-particle cohesion. Alternately, small differences in the volume contraction on cooling between the solvent phase and the particles could lead to debonding prior to fracture, or to the formation of cracks at the particle-solvent interface. In either case, the weak zone appears to be at the interface and interpretation of freeze-fracture images is greatly simplified. Branton (1966) has shown that the weak zone in bilayer membranes is along the hydrocarbon interior of the membrane; this also appears to be true for bilayer phases in general.

**Etching**

The controlled sublimation of the solvent, known as etching, can be used to enhance the topographic variations in a fractured, microstructured fluid. However, removing too much of the solvent can alter the apparent location and distribution of dispersed particles, hide evidence of crystallization induced reorganization, and make the replicas difficult to pick up and clean. If the dispersed phase is entangled and self-supporting, as are polymer solutions and gels, a limited amount of etching can bring out the network structure (Zasadzinski et al. 1987a). The important parameter in etching a sample is temperature. The sublimation pressure, hence the sublimation rate, is set once the sublimation temperature is fixed. The sublimation rate, \( S \), in nanometers per second from a surface under vacuum can be obtained from gas kinetic theory:

\[
S = \frac{P_s - P_v}{\rho_e} \left( \frac{M_e}{2\pi kT} \right)^{1/2} \times 10^7
\]

in which \( P_s \) is the saturation vapor pressure and \( P_v \), the background pressure of the sublimating phase in dynes/cm\(^2\), \( M_e \) is the molecular weight (g/mole) and \( \rho_e \) (g/cm\(^3\)) is the density of the sublimating phase. \( R \) is the gas constant, and \( T \) is the sublimation temperature (K). The saturation vapor pressure for most solvents can be found in general engineering handbooks. If the background pressure of the solvent is greater than the vapor pressure at the temperature chosen, material will condense from the vacuum onto the sample, obscuring surface details. Hence, care is necessary to understand the relative composition of the residual gases in the vacuum chamber.

**Replication**

The goal in the replication process is to reproduce the fracture surface as accurately as possible with an electron opaque shadowing layer backed by a continuous, electron transparent, backing layer. The resolution in freeze-fracture electron microscopy is limited by imperfections in replication (Akahori et al., 1986; Gross et al., 1985). Ideally, the evaporated metal atoms, which are usually a mixture of platinum and carbon, stick exactly where they land and form a structureless layer. However, the surface energy of the metal layer is much higher than that of the original fracture surface of water or hydrocarbons; hence, the metal film does not spread or "wet" the surface, but aggregates into small droplets (Adamson, 1990. Woodward, 1994). The aggregates grow in size, eventually merging with neighboring aggregates to form a continuous film. For most electron microscopy applications, the evaporation is stopped prior to the formation of a continuous metal layer. In our laboratory, about 1.5 nm of platinum carbon followed by about 15 nm of carbon backing gives optimum results. For proper interpretation of the replica, the sample material must be completely removed from the replica before viewing. In our laboratory, the cleaning method of Fetter and Costello (1985) has always given the best results.

**Capillary condensation and feature height amplification in STM imaging in air**

Up to this point, the preparation for TEM and STM examination are interchangeable; in fact, it was our original intention to be able to image the same sample with both TEM and STM to compare features and eliminate artifacts. However, we have found that there is a fundamental difference in the sample requirements
for the two imaging techniques. In our initial studies of freeze-fracture replicas with the STM, we found that the height information was very erratic. Images of similar objects on the same replica could differ by more than a factor of ten in apparent height (Woodward et al., 1991). Other times, the height of an object would grow or shrink with continued scanning (Hansma et al., 1988). The literature also includes a number of accounts of unusual height measurements on a variety of samples (Coleman et al., 1985; Hallmark et al., 1987; Hansma and Tersoff, 1987). A number of theoretical models have proposed that specific quantum and atomic scale interaction scan lead to amplifications in specific tip-sample systems (Tersoff and Lang, 1990; Ciraci et al., 1990; Yuan and Shao, 1990; Chen, 1992). Other theories propose that bulk compression by the scanning tip at regions of low conductivity or solid surface contamination could lead to amplifications of surface features (Soler et al., 1986; Mamin et al., 1986).

In most descriptions of scanning tunneling microscopy, there is no physical coupling between the STM tip and sample, even though there are at least three significant interactions to consider when imaging is done in air: (1) electrostatic interaction due to the potential difference between the tip and sample, (2) van der Waals attraction, and (3) capillary attraction due to the Laplace pressure generated by the formation of a highly curved fluid meniscus connecting the tip and sample (Woodward et al., 1991; Woodward and Zasadzinski, 1994). These forces can cause the STM tip, while traversing a surface feature with an actual height, Z, to distort the surface being measured, resulting in an amplified height, Z*.

The role of capillary forces in scanning probe microscopy has been recognized in the AFM literature (Erlandsson et al., 1988; Weisenhorn et al., 1989; Blackman et al., 1990) while an appreciation of their importance during STM imaging has lagged (Yuan et al., 1991; Anselmetti et al., 1993; Guckenberger et al., 1994).

The real forces that couple the tip to the sample can be modeled by a spring of spring constant, k₁. A second spring of spring constant, k₂, connects the sample to the STM base. Regardless of the functional form of k₁ and k₂, this combination of springs will amplify a surface feature of actual height Z by:

\[ Z^* = Z(k_1 + k_2)/k_2 \]  

where \( Z^* \) is the height measured by the STM. In most applications where the sample is a uniformly rigid solid firmly mounted to the base, \( k_2 \) is very large compared to \( k_1 \) and no amplification is evident (\( k_2 \) would be related to the bulk compressibility or elastic modulus of the sample being imaged (Mamin et al., 1986)). However, if the sample is soft, non-uniform, or weakly connected to the base, as is often the case for organic or biological films, or even layered solids, \( k_2 \) can be much less than \( k_1 \), leading to large amplifications.

To estimate these two important spring constants, we first assume that the force and spring constants between the tip and sample can be approximated by the linear combination of electrostatic, van der Waals and capillary forces. For a spherical tip of radius R and a tip-sample distance D, (Fig. 1) the van der Waals force is:

\[ F_{vdW} = \frac{AR}{6D^2} \]  

where A is the Hamaker constant (Israelachvili, 1991). The Hamaker constant for two platinum wires in air or an electrolyte is 20x10^-20 J (Derjaguin et al., 1978). The electrostatic force between a sphere held at a potential, \( V_0 \), and a grounded plane is given by

\[ F_e = \frac{\pi e V_0^2 R}{D} \]  

in the limit of \( R >> D \).

The capillary force due to the meniscus between the

---

**Figure 1:** Schematic showing a fluid meniscus connecting the STM tip to the replica. The figure is not to scale.

---

128
Table 1: Sample-Tip Spring Constants as a Function of Relative Humidity

<table>
<thead>
<tr>
<th>Interaction</th>
<th>(P/P_s = 0.0)</th>
<th>(P/P_s = 0.10)</th>
<th>(P/P_s = 0.50)</th>
<th>(P/P_s = 0.75)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(k_i(N/m))</td>
<td>(F(nN))</td>
<td>(k_i(N/m))</td>
<td>(F(nN))</td>
</tr>
<tr>
<td>Electrostatic</td>
<td>0.01</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>van der Waals</td>
<td>3.3</td>
<td>1.7</td>
<td>3.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Laplace</td>
<td>0.0</td>
<td>10.0</td>
<td>11.0</td>
<td>7.6</td>
</tr>
<tr>
<td>Total</td>
<td>3.31</td>
<td>14.4</td>
<td>15.4</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Calculated values for the spring constants and forces of the three interactions we have considered both in a dry environment \((P/P_s = 0.0)\) and for a water meniscus at various relative humidities. Note that \(k_i\) is nearly constant over a wide range of relative humidities, and the dominant interaction is the Laplace pressure due to the meniscus.

\[
F_L = \frac{4\pi R \gamma \cos \theta}{(1 + D/d)} \quad (7)
\]

where \(\gamma\) is the surface tension of the condensate and \(\theta\) is the liquid-solid contact angle, \(d\) is the distance the tip extends into the meniscus and is given by \((V/RT) \cos \theta/\ln(P/P_s)\) where \(V\) is the condensate molar volume, \(R\) the gas constant, \(T\) the temperature, and \(P/P_s\) is the relative humidity (Israelachvili, 1991). For water, \(\gamma = 0.073 J/m^2\) and \(d = -(1.08 \text{ nm})/\ln(P/P_s)\). Using these approximations, the spring constant \(k_i = -\partial F/\partial d\) is:

\[
k_i = \frac{\pi \gamma V^2 R}{D^2} + \frac{AR}{3D^3} + \frac{4\pi \gamma R \cos \theta}{d(1 + D/d)^2} \quad (8)
\]

and is governed by the Laplace pressure for relative humidities over 10% as shown in Table 1. From the exact meniscus profiles and interaction forces (Orr et al., 1975; Zasadzinski et al., 1987b), we calculate that the pressure within the meniscus can exceed 100 atm, which is sufficient to damage soft materials, although it would only lead to small deformations of crystalline solids such as graphite (Mamin et al., 1986). The wide range of local pressures experienced by samples during scanning may explain the irreproducibility typically found in imaging biological materials.

We evaluate \(k_2\), the spring constant between the sample and the STM base, for thin, metal surface replicas for TEM imaging as described above. However, the general principles used should be valid to estimate \(k_2\) for many of the weakly supported samples common to STM imaging. Replicas for TEM are thin metal films mounted on porous metal mesh grids; where the replica is in direct contact with the mesh, \(k_2\) is large. However, for the part of the replica that is loosely suspended over a pore of area, \(a\):

\[
k_2 = \frac{ET^3}{0.14a(1-\nu^2)} \quad (9)
\]

where \(E\) is the Young’s modulus of the replica, \(\nu\) is Poisson’s ratio, and \(T\) is the thickness of the replica (Roark, 1965). Metal surface replicas prepared for transmission electron microscopy are about 25 nm thick with \(E = 2 \times 10^{10} \text{ N/m}^2\) and \(\nu = 0.3\), and are mounted on a mesh with 30 \(\mu\text{m}\) spacings. Hence, \(k_2 = 0.003 \text{ N/m}\) in the center of the mesh. Taken together with the estimate of \(k_1\), Eq. 4 suggests a possible 1000 fold amplification, which is clearly outside the limits of our approximations, but also indicative of potential artifacts due to tip-sample interactions. Most importantly to reliable imaging, we cannot determine whether the particular area being imaged is well supported or not during a particular STM scan. Clearly, \(k_2\) will vary significantly depending on the relative location of the sample with respect to the support mesh. As a result, we expect, and have found, that feature heights often vary with location on the surface. This effect will not be appreciable for chemically and physically homogeneous "bulk" samples (>1 mm thick). However, even if the sample is "bulk", any weak point in the mounting or in the local adhesion of the sample to the substrate or the substrate to the STM base can lead to amplification of surface features. This also suggests that certain areas of a sample might be highly amplified, leading to permanent local distortion of the surface, where other areas might not be distorted at all. As controlling the properties of the sample surface are difficult, and we do not know the details of the interaction a priori, it is important to identify the impact of capillary forces on test images, then minimize these forces for reliable imaging of unknown surfaces.
Modifications to freeze-fracture replicas for STM imaging

From Eqn. 9 above, it is clear that making the sample thicker and reducing the relative humidity are the best routes to increase \( k_f \) and minimize sample height amplification. To properly test these ideas, we made two types of specimens, one "bulk" and one "thin film", with identical chemical composition and surface features and with well-known feature heights. We chose metal shadowed Langmuir-Blodgett multilayer islands of cadmium arachidate deposited on 1 mm thick mica substrates. The original specimen surface consisted of bilayer islands of cadmium arachidate of 1-2 \( \mu \)m in extent that vary in height by a bilayer from adjacent areas (Schwartz et al., 1992a,b). The thickness of the bilayer is known from X-ray diffraction (Tippmann-Krayer, 1992) and AFM images to be 5.5 nm (Schwartz et al., 1992a,b). To make the surfaces conductive and chemically uniform, 1.5-2 nm of platinum was applied by electron beam evaporation at a 45° angle while the sample was rotated in a Balzers 400K freeze-etch device (Balzers, Hudson, NH). This sample could then be imaged directly with the STM as the "bulk" sample. For the thin film samples, we deposited an additional 30 nm of carbon as a backing film to a second set of platinum coated LB island surfaces on mica, followed by a 200 nm thick layer of silver. These multilayer metal films were then stripped from the mica substrate using hydrofluoric acid, washed repeatedly in Millipore filtered water, and mounted, platinum side up, on silver wire mesh filters (SPI Supplies, West Chester, PA.) with a nominal 0.2 \( \mu \)m pore size. STM (Nanoscope II, Digital Instruments; Santa Barbara, CA) images were obtained with cut platinum/iridium tips at 100 mV bias voltage, a 1.0 nA tunneling current and a scan rate of 5.8 Hz. Step heights were measured by taking bearing plots of roughly equal areas on both sides of a step and taking the difference between the peak values to be the step height.

To control the environment and eliminate condensation between the tip and sample, we enclosed the STM in a bell jar with a base plate that allowed us to evacuate the chamber with a rotary pump to <0.1 torr and backfill with dry nitrogen (Woodward, 1994). For the "bulk" samples on mica, 46 bilayer steps were measured on 18 different images taken with different tips in both ambient conditions and a dry nitrogen atmosphere. Seventy percent of the measurements fell within eight percent of the mean, and there was no difference between samples imaged in air or in dry nitrogen. This variance in the step height was similar to that achieved measuring the bilayer steps directly with the AFM (Garnaes et al., 1993). We attribute most of the scatter to nonlinear effects in the piezoelectric crystal in the scanning head as well as some coupling between the z calibration and the x and y position. As the sample is a bulk solid and the adhesion of the surface film to the substrate was good, \( k_f \) is large in Eqn. 9 above and we did not expect nor did we find any amplification of feature heights. These bilayer steps of 5.5 nm also make excellent vertical calibration samples for the STM in a range difficult to find elsewhere.

We then imaged the thin film replicas of the LB bilayer steps in a dry atmosphere. In the dry atmosphere, the images were stable for more than one hour and lateral and vertical feature dimensions were reproducible for hundreds of scans. Ten images taken with three different tips gave 28 bilayer height measurements. The mean height was 91% of the mean height from the "bulk" sample. Seventy percent of the measurements fell within eight percent of the mean as with the coated sample. Again, the variance was similar to that achieved measuring the bilayer steps directly with the AFM (Garnaes et al., 1993). There was no significant difference in heights measured with different tips.

However, when we changed the environment of the sample during imaging, we could record significant changes in feature heights due to contamination from the lab air. Fig. 2 shows one such series of images (Woodward and Zasadzinski, 1994). Imaging the replica gives an inverted image of the surface so the large feature resembling Australia is a bilayer higher than most of the replica surface and corresponds to a bilayer deep hole in the original LB film. When imaged in the dry nitrogen atmosphere (Figs. 2a-b), the image is stable. However, when lab air of 65% relative humidity is allowed into the bell jar, the height difference between the background and the bilayer 'hole' increased by about 5.5 nm to more than 7 nm, (see Fig. 3) and several of the small features on the background showed a significant increase in height. The difference between the highest and lowest points (z-range) in the image increased from 18.6 nm in 1b to 40.0 nm in 2d. New "features" or high spots appear to grow on other areas of the replica as the replica is exposed to humid air for longer periods of time (Fig. 2d). The feature heights would vary from scan to scan when imaging in humid air, and would often appear and disappear. Upon flushing with dry nitrogen the effects of the condensation were quickly reversed. Figs. 2e, 2f show that the high spots vanish within minutes and the bilayer step returns to nearly the original height. However, although the "pseudo-features" that appeared in Figs. 2c and 2d have disappeared from the image and the z-range of the image was restored, some of the smaller bilayer islands seen in Fig. 1a have also shrunk or disappeared. Hence, the amplification of the sample during imaging in humid air.
Figure 2: Series of six images of a platinum replica of a cadmium arachidate multilayer Langmuir-Blodgett film. Figs. a-b were taken under a dry nitrogen environment at time 15 and 31 minutes respectively and are nearly identical. The sharp differences in gray scale correspond to bilayer steps of 5.5 nm in height (See Fig. 3). Figs. c-d were taken after exposure to humid air at times 38 and 41 minutes and show significant variations in surface topology, most likely due to transient sample deformation due to capillary coupling of the STM tip to the sample (full arrows). Figs. e-f were taken with the sample under nitrogen again at times 50 and 57 minutes. Compare e-f with a-b; most of the surface features have returned to their original height and shape. All images are 200 nm x 200 nm and the height is linearly gray scaled over 44 nm from black (low) to white (high). Although most of the "pseudo-features" that appeared in c-d disappeared from the image and the z-range of the image was restored, some of the smaller bilayer islands seen in Fig. 2a have also shrunk or disappeared between 2a and 2f. Hence, the amplification of the sample during imaging in humid air can lead to permanent surface deformation, and hence should be avoided.
Figure 3: Height histograms of Figs. 2b (top), 2d (middle) and 2f (bottom) taken from areas on both sides of the large bilayer feature. The left peak in each plot corresponds to the background height and has been set to zero. The right peak corresponds to the height of the bilayer island. The plots are scaled so that the height difference between the peaks in the top figure is calibrated to the known bilayer height of 5.5 nm. The bilayer height increases to 7.0 nm upon exposure to air (middle) and returns to 5.4 nm (bottom) when returned to a dry nitrogen environment.

can lead to both transient and permanent surface deformation, and hence should be avoided. Fig. 3 shows histograms of the height difference between areas on either side of the large feature resembling Australia from the images in Figs. 2b, 2d and 2f. Fig. 3a shows that the step height (measured from the peak of the distributions) is 5.5 nm. (The width of the distribution is related to the finite size of the metal grains that make up the replica which are about 2 nm in diameter.) Fig. 3b shows that in humid air, the step height grows to more than 7 nm, and Fig. 3c shows that the step height returns to 5.4 nm on return to the dry atmosphere. Clearly, as these plots were from identical areas, capillary condensation from humid air is responsible for this amplification.

Quantitative Height Measurements by Correlation Averaging

From Figs. 2 and 3, it is clear that the resolution in any individual STM image of a freeze-fracture replica is limited by the granularity of the replicating film. In general, the quality of the replica is determined by how well the metal film replicates the surface on which it was deposited. The ideal replicating film would conform at the molecular level to the sample surface, be highly conductive and continuous, and would not reorganize over times long compared to typical imaging times. However, the metal replicas are made up of small metal grains several nanometers in diameter (Zasadzinski and Bailey, 1989; Ruben, 1989; Wepf et al., 1991). Ideally, the evaporated metal atoms, which are usually a mixture of platinum and carbon, stick exactly where they land and form a structureless layer. However, the surface energy of the metal layer is much higher than that of the original fracture surface of water or hydrocarbons; hence, the metal film does not spread or "wet" the surface, but aggregates into small droplets (Adamson, 1990; Woodward, 1994). The aggregates grow in size, eventually merging with neighboring aggregates to form a continuous film. In STM images, the grains contribute a "noisy" background that can be removed by correlation averaging techniques or other image analysis techniques. However, as the grey scale in an STM image contains quantitative three-dimensional information, the type of correlation averaging technique becomes important to preserve this information unaltered (Woodward, et al., 1995).

Correlation averaging is used to enhance images that contain a repeating pattern. Computer algorithms based on cross-correlation principles have been developed by several researchers (Frank et al., 1978; Crepeau and Fram, 1981; Saxton and Baumeister, 1982; Henderson et al., 1986) to locate and average patterns found in digitized images of electron micrographs. Correlation averaging is widely used to analyze images of biological structure with poor signal to noise ratios or those that do not form large crystal arrays. Correlation averaging schemes are a standard part of many commercially available software packages for electron microscopy image analysis (Hegerl, 1992). The dramatic improvement in image quality and resolution often reveals
structural information that is obscured in any individual image. More recently, correlation averaging has been applied to images taken with scanning tunneling microscopes (Soehnout et al., 1988; Amrein et al., 1989; Stemmer et al., 1989; Wang et al., 1990) and atomic force microscopes (Weigrabe et al., 1991).

While the details of correlation averaging routines in the literature vary, they all follow the same basic outline:

1. an image of the structure of interest is digitized,
2. a small area containing at least one "unit" is chosen as the test image,
3. this image is cross correlated with the main image,
4. peaks in the resulting correlation correspond to the location of images similar to the test image,
5. a criteria for selecting the largest peaks is then applied to determine which areas of the image are used to make the composite image, and
6. these areas are averaged to form a composite image.

Many of the current image analysis routines contain a variety of additional techniques for enhancing images including corrections for the microscope transfer function, 3D reconstructions from a series of different tilts, and multivariate statistical analysis (Henderson et al., 1990; Hegerl, 1992). The review by Hegerl describes many of the techniques included in commercially available image processing software. The following discussion will focus solely on correlation averaging.

**Vertical distortions in correlation averaging**

Although the composite images created by correlation averaging provide resolution superior to that in the original image, they also introduce a bias as described below. For electron micrographs, this bias is not important as the quantitative values of the pixel are rarely significant. The variety of contrast mechanisms in electron microscopy make it difficult to relate the absolute value of a pixel element to the properties of the original sample. In most images, the contrast in the composite is used primarily to determine the position of features in the image plane. In other cases, the ratio of two pixels is used; the absolute value of the pixel value is irrelevant (Zasadzinski and Bailey, 1989). In scanned probe microscopies, however, the pixel values from calibrated images represent absolute height measurements. Thus, any bias introduced by the correlation function leads to quantitative errors in "height" measurements made on the composite and could lead to possible misinterpretations of the image.

For calculation by computer, the discrete form of the correlation function

\[ G(x',y') = \sum_{x=0}^{m-1} \sum_{y=0}^{m-1} T(x,y)M(x+x',y+y') \]

is usually used in conjunction with the fast Fourier transform (FFT) algorithm. Here \( T(x,y) \) is a \((m\times m)\) pixel test image, \( M(x,y) \) is an \((n\times n)\) pixel main image, and \( x, x', y, y' \) are all integers corresponding to the pixel location. Direct calculation using Eqn. (10) requires \((n-m+1)^2(m)^2\) multiplications while the use of the FFT algorithm requires \(5(2\log_2(n))^2 + (2\log_2(m))mn + 4m^2\) multiplications (Niblack, 1986). The value of \( T(x,y) \) or \( M(x,y) \) is the digitized pixel intensity of the original image for electron micrographs or relative height values for scanning probe images. In the correlation averaging techniques, the coordinates \((x',y')\) of the peaks of the correlation function \( G(x',y') \) are taken to correspond to the locations in the original image that most closely resemble the test image. By averaging over many areas, random noise is minimized, giving a composite image that resembles the actual structure more closely than do any of the individual images. However, a close look at the correlation function shows that it does not treat positive and negative deviations from the test image in an equitable way.

There are two potential problems with the correlation function: (1) positive deviations from the actual structure lead to an increase in the value of the correlation function while negative ones lead to a decrease in the value, and (2) the relative effect of deviations from the test image strongly depends on the pixel value of the test image. The first problem tends to skew the entire image toward higher values while the second problem tends to increase higher valued pixels in the test image more than lower valued ones. To illustrate these problems we can examine the simple test case of a true image with pixel values \((2, 5, 10, 5)\). For simplicity we will assume that these are also the values in our test image. Let us then imagine that there are two different areas in the main image, the first with each pixel value one lower than the test image, \((1, 4, 9, 4)\), and the second with each pixel value one higher, \((3, 6, 11, 6)\). The correlation value of the first area is 132 while that of the second area is 176. If the cut off for including a peak in the average falls between 132 and 176, the high valued area will be included in the average, while the low valued area will not. The higher pixel values are clearly favored. We should note that for our simple example, biasing the composite toward higher pixel values is not a problem for either TEM or SPM images as neither has an absolute reference scale. In all three of the simple images, the peak value is eight units above
the lowest value.

To illustrate the second problem, take four more areas: (4, 5, 10, 5), (0, 5, 10, 5), (2, 5, 12, 5) and (2, 5, 8, 5). The respective correlation values with the test image are 158, 150, 174 and 134. The noise located at the pixel in the test image with the highest value has a much larger effect on the correlation value. When a threshold is selected, areas with positive valued noise will be selected over those with negative valued noise, leading toward a bias of the whole composite toward higher values. Furthermore, negative valued noise from high spots in the test image is more likely to be rejected than the same noise at a lower valued pixel. This causes the correlation to enhance the peaks in the test image. In the resulting composite, the image appears stretched with high areas extended disproportionately. In our example, if the threshold is \( \geq 150 \) only the first three areas will be averaged to form the composite \((2, 5, 10, 7, 5)\). Note that the peak value has grown while the others have remained the same. The corresponding slight change in the visual appearance of the image is probably unimportant to the TEM user, but the SPM user will now measure the difference between the high and low point to be 8.7. Hence, the correlation averaging algorithm has the potential to cause a quantitative errors in our results.

Following Schulz-DuBois and Rehberg (1981), we propose that the structure function is superior to the correlation function when absolute pixel values are important. In its discrete form

\[
S(x', y') = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} [T(x, y) - M(x', y')^2]
\]

the structure function is easy to recognize as the sum of the squared deviations. The main benefits of the structure function are: (1) it treats positive and negative noise equally, and (2) it does not discriminate based on the pixel value of the test image. The minima of the structure function give the best areas to be included in the composite. Direct computation of the structure function requires the same number of multiplications as the correlation function, but requires an extra addition for every multiplication. Calculation using the FFT algorithm requires \(9(2\log_2(n))^2 + (2\log_2)m + 12n^2\) multiplications. For a 512x512 image with a 32x32 test area this is only twice the number required for the same calculation using the correlation function.

To test the relative merits of the correlation and structure functions we wrote a simple averaging program. The program calculates the correlation or struc-

---

Figure 4: Computer generated images used to test the correlation averaging and structure averaging routines. (a) The image with no noise added. (b) The image with a noise level of 2.

---

ature function between an image file and a test area taken either from a portion of the image itself or from another file. The calculation is done in real space using Eqn. (10) or (11). The calculation is only done for points where the test image fits entirely within the main image, so there is no padding or wrapping around at the edge of the image. This results in a \((m-n)\times(m-n)\) field for the
J.T. Woodward and J.A. Zasadzinski

Table 2: Errors in correlation and structure averaging of test sample

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>Correlation</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>324</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>0.0±0.0</td>
<td>0.0±0.0</td>
</tr>
<tr>
<td></td>
<td>0.0±0.0</td>
<td>0.0±0.0</td>
</tr>
<tr>
<td>1</td>
<td>0.0±0.0</td>
<td>0.00±0.01</td>
</tr>
<tr>
<td></td>
<td>0.0±0.0</td>
<td>0.02±0.06</td>
</tr>
<tr>
<td>2</td>
<td>0.0±0.0</td>
<td>0.06±0.03</td>
</tr>
<tr>
<td></td>
<td>0.0±0.0</td>
<td>0.24±0.05</td>
</tr>
<tr>
<td>4</td>
<td>0.0±0.0</td>
<td>0.24±0.05</td>
</tr>
<tr>
<td></td>
<td>0.0±0.0</td>
<td>0.49±0.07</td>
</tr>
<tr>
<td>8</td>
<td>0.47±0.07</td>
<td>0.87±0.11</td>
</tr>
<tr>
<td></td>
<td>0.69±0.06</td>
<td>1.12±0.09</td>
</tr>
</tbody>
</table>

The results of the correlation and structure averaging routines. The programs were run on the same images and either 324, 162, 81 or 32 areas were used to form the composite images. The top value in each entry is the mean deviation in pixel value from the true image. The mean deviation is always near zero for the structure averaging routine, indicating that there is no bias induced by the averaging. The mean deviation is always positive for the correlation averaging routine, indicating that the correlation averaging tends to generally increase pixel values. The bottom value is the root mean square deviation per pixel which is also lower for the structure averaging routine. The test area used was the true image with no noise added.

correlation or structure function. The program then finds and sorts all the local maxima (correlation) or minima (structure) in the function. Extrema at the edge are ignored. The user pre-selects the number of areas to be averaged to form the composite image. The areas corresponding to highest local maxima in the correlation function, or lowest local minima in the structure function, are averaged and a file containing the output is created. Our program does not interpolate between data points. The program was written in C and all the results presented were obtained on a Silicon Graphics Indigo computer.

**Test on a known image**

To demonstrate the advantages of the structure function over the correlation function we made a test area and test image composed of a 20x20 array of test areas as can be seen in Fig. 4a. The test area is 5 pixels x 5 pixels with the following values:
A pseudo-random number generator was then used to add varying degrees of noise to the image. A noise level of 1 corresponds to -1, 0 or 1 being added to each pixel, a noise level of 2 corresponds to -2, -1, 0, 1 or 2 being added, and so on. An image with a noise level of 2 is shown in Fig. 4b.

An averaging routine was used to make composite images using both the structure and correlation functions on images containing noise levels of 1, 2, 4 and 8. These correspond to images with a signal to noise ratio of 13.0, 4.3, 1.30 and 0.36 respectively. As our routine ignores the areas near the edge of the main image, both the correlation and the structure function find 324 peaks in the zero noise main image corresponding to the 18x18 matching areas on the interior. We ran the program using the best 324, 162, 81 and 32 maxima or minima to find the areas to be averaged for the composite. The average deviation per pixel between the composites and the true image and the root mean square (RMS) deviation per pixel were calculated. The average results of ten runs are shown in Table 2.

The results clearly show that for any appreciable noise level the correlation function skews the composite toward higher pixel values, while there is no significant skewing at any noise level for the correlation function. In addition the (RMS) deviation from the true image is significantly lower for the structure function generated composites than for the correlation function generated composites. The increase in the RMS deviation for both functions when fewer areas are included in the composite is to be expected, as it is less likely that the noise will average to zero when fewer data points are included in the average. Similarly, as the level of the noise increases, the typical deviation of the average from the true value grows.

In Fig. 5 we look at the extreme case of the results from taking the best 32 areas of the noise level 8 images. Fig. 5a shows the actual 5x5 unit cell. Figs. 5b and 5c are the composites from the correlation averaging and structure averaging programs respectively. The actual unit cell had pixels valued 1, 3, 5 and 10 as seen in (12). The average value of these pixels in the correlation composite is 1.8, 2.7, 7.8 and 14.5 respectively and in the structure composite they are 1.2, 3.0, 4.7 and 8.5. Although in practice it would be rare to restrict oneself to using so few of the available areas to form a composite, the example serves to illustrate the effect that the averaging routines can have on the composites.

Figure 5. The results of the averaging routines on images with noise level 8 when only the top 32 areas are used in the average. (a) The true unit cell. (b) The composite unit cell formed by the structure averaging routine. Other than the central peak all pixels retained their true value. (c) The composite unit cell formed by the correlation averaging routine. The values of the pixels are generally greater than those in the true unit cell with the peak values especially enlarged.
error in the difference between the peak and low values is more than twice as large using the correlation function compared to the structure function.

The case we have presented assumes that the true image can be used as a test function, which is usually not the case in most applications. When we ran our averaging programs using a 10x10 pixel test area chosen from within the image we found that the biasing effect of the correlation function remains, although it is decreased by nearly half, while the RMS deviations are little changed. More surprisingly, the RMS deviations for the structure function increase to become comparable with, but still less than, those of the correlation function. There is still no bias with the structure function.

Measurements of ripple wavelength and amplitude in DMPC bilayers

One of the most difficult structures to examine by any technique are lyotropic liquid crystals - the general materials class into which bilayers, liposomes, and related lipid phases fall. Conventional x-ray diffraction methods usually cannot provide sufficient resolution as these materials are difficult to align, domain sizes and correlation lengths are small, and thermal fluctuations tend to obscure structural details (Sirota et al., 1988). The freeze-fracture technique has been used primarily to investigate membrane structure and has led to a general acceptance of the fluid mosaic model of cell membranes, in which integral proteins are embedded in a lipid bilayer (Branton, 1966; Singer and Nicolson, 1972). Recent finding have shown that the lipid membrane is not only a passive matrix; protein function can be modified by the composition, phase and local structure of the lipids (Keller et al., 1993).

Hence, it is important to study the phase behavior and structures of phospholipids, and saturated phosphatidylcholines (PC) in particular, as PC's are present in many cell membranes and are the major component of human lung surfactant (Longo et al., 1993). Saturated PC's undergo three distinct structural transitions when dispersed in water: a subtransition, pretransition and main transition separating the phases L_n, L_{g'}, P_r, and L_s (Tardieu et al., 1972; Janiak et al., 1979; Wack and Webb, 1989). In the high temperature L_n phase, the order within each bilayer is short-range and the trans-gauche intramolecular order is low (Tardieu et al., 1972). The main transition is associated with lipid chain melting (Tardieu et al., 1972). The bilayers are smooth and the molecules, on average, are normal to the bilayer (Sirota et al., 1988). The L_{g'} phase is characterized by flat bilayers with the lipid chains fully extended (all-trans configuration) and tilted with respect to the bilayer normal. The magnitude and direction of tilt depends on the water fraction (Sirota et al., 1988). The low temperature L_{g'} -> L_n transition involves a modification of the chain packing and dehydration of the head groups (Janiak et al., 1979).

A satisfactory explanation of the P_r phase remains a theoretical and experimental challenge. In the P_r phase, the lipid chains retain much of their all-trans configuration, and the molecules are packed into a two-dimensional hexagonal lattice with long range correlations (Janiak et al., 1979; Ruppel and Sackman, 1983; Zasadzinski and Schneider, 1987). The bilayers are characterized by regular three-dimensional corrugations, hence the common name of ripple phase. X-ray diffraction and freeze-fracture electron microscopy find the ripple wavelength in excess water to be about 10 - 15 nm (Tardieu et al., 1972; Luna and McConnell, 1977; Janiak et al., 1979; Ruppel and Sackman, 1983; Zasadzinski and Schneider, 1987; Zasadzinski et al., 1988; Wack and Webb, 1989). The ripples in the P_r phase are capable of aligning large molecules such as proteins along their length; diffusion within the plane of the bilayer appears to be anisotropic (Schneider et al., 1983). Replacing one or more methyl groups from the choline headgroup by hydrogen, thereby decreasing the size of the headgroup relative to the chains, eliminates the P_r phase (Zasadzinski and Schneider, 1987). Models of the P_r phase fall into two broad classes: phenomenological models based on modulations of membrane thickness (Goldstein and Leibler, 1988; Cevc, 1991) or curvature (Doniach, 1979; Lubensky and MacKintosh, 1993) and molecular models based on packing frustration between the lipid headgroup and chains (Carlson and Sethna, 1987; McCullough and Scott, 1990; Scott and McCullough, 1991; Schwartz et al., 1994). The structural details of the ripples necessary to test these theories cannot be determined by X-ray diffraction or freeze-fracture electron microscopy. The particular features of this phase makes it an ideal system to examine by the freeze-fracture-STM technique (Woodward and Zasadzinski, 1995).

Metal replicas of DMPC bilayers in excess water, equilibrated at 16, 18, 20 and 23°C were prepared (as described above) for both TEM and STM examination to determine the temperature dependence of the ripple amplitude, waveform, and wavelength (the P_r phase of DMPC exists from 14-24°C). For each sample, 200 mg of DMPC (Avanti Polar Lipids, Atlanta, GA) was added to 0.2 ml Milli-Q water (Millipore, Bedford, MA); the resulting 50% DMPC/50% water mixture assured that the bilayers are fully hydrated (Wack and Webb, 1989). The samples were alternately centrifuged at low speed, vortexed, then heated at > 30°C for at least 24 hours to allow complete mixing. Thin films of the DMPC-water mixture were sandwiched between copper freeze-fracture planchettes, then equilibrated at 100% relative humidity
Figure 6: (a) TEM image of DMPC in excess water at 23°C. The ripples change orientation over μm length scales, typically at 120° angles, indicating an underlying hexagonal packing. The ripple asymmetry can be inferred from the shadow patterns - narrow, dark gray - wide, light gray lines in the upper left of the image (open arrow) adjacent to narrow, white - wide, dark gray lines in the lower right of the image (filled arrow). (b) TEM image of DMPC in excess water at 16°C (The sample at 18°C was similar). The ripples are much less well defined and appear like a string of beads. This beading is induced by a self-shadowing process that enhances small features. The ripple amplitude is much smaller and the asymmetry in the shadowing has disappeared.

at 23, 20, 18 or 16°C prior to rapid quenching in liquid propane cooled by liquid nitrogen. Freeze-fracture replication was done in a Balzers 400K freeze-etch machine as described above. The TEM samples were shadowed with 1.5 nm of Pt/C at a 45° angle relative to the surface, while the STM samples were coated normal to the surface while the sample table was rotated to insure a continuous coating. A 15 nm thick film of carbon was added to stabilize the shadowing film. Replicas for STM had an additional 0.5 μm thick layer of silver deposited by sputtering to increase film rigidity. Replicas of each sample were examined first with TEM (JEM 100CXII) to determine replica quality and identify the ripple phase. STM (Digital Instruments, Santa Barbara, CA) imaging was done with a 12 μm scanning head in the constant current mode under a dry nitrogen atmosphere. Each sample was examined with several

Table 3: Amplitude of first two harmonics and wavelength of ripples from different images as determined by correlation averaging

<table>
<thead>
<tr>
<th>Temperature</th>
<th>A₁ (nm)</th>
<th>A₂ (nm)</th>
<th>λ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20°C</td>
<td>0.62</td>
<td>0.12</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.16</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.07</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.06</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.06</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.07</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.10</td>
<td>12.9</td>
</tr>
<tr>
<td>23°C</td>
<td>1.20</td>
<td>0.25</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>0.20</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>1.03</td>
<td>0.20</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.20</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>0.20</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>1.72</td>
<td>0.36</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1.82</td>
<td>0.38</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1.29</td>
<td>0.36</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.26</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>0.68</td>
<td>0.21</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>0.16</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>0.86</td>
<td>0.20</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>1.16</td>
<td>0.28</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.27</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>1.03</td>
<td>0.30</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.34</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1.08</td>
<td>0.40</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>1.18</td>
<td>0.40</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>0.74</td>
<td>0.14</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>1.34</td>
<td>0.30</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.25</td>
<td>10.1</td>
</tr>
</tbody>
</table>

138
Figure 7: (a) Predicted ripple shadow pattern for symmetric ripples if $\theta_2 > \theta$, most of the ripple is uncoated by platinum and will appear gray-white in micrographs. If $\theta_2 \leq \theta$, all the ripple will be coated and will appear dark gray-light gray in micrographs. $\varphi$ is the macroscopic shadow angle, which is typically 45° in these experiments. (b) Asymmetric ripples will appear different depending on their relative orientation with respect to the shadow direction. Gray-white if the long side is oriented towards the shadow directions, dark gray-light gray if away. Compare to the patterns in Fig. 6a.
samples are shown in Fig. 6. In the 23 and 20°C samples, the ripples are continuous and the ripple asymmetry is apparent from the variation in shadowing across the image (Fig. 6a) (Zasadzinski and Schneider, 1987). Near the open arrow, the ripples appear as wider light gray lines with narrow, dark gray lines. At the filled arrow, the pattern is reversed - narrow white lines with wide, dark gray lines (Fig. 7). These patterns are only consistent with an asymmetric waveform for the ripples. In TEM images of the 18 and 16°C samples, the ripples are discontinuous and appear to be made up of a line of beads. This pattern of beads is consistent with a self-shadowing artifact; the oblique metal deposition enhances small features in the fracture surface (Ruben, 1989). It is clear that the ripple is of smaller amplitude, the shadowing pattern has lost its asymmetry and is less well defined, suggesting a temperature dependence of the ripple amplitude. When the replica is made by depositing the metal film normal to the surface (which eliminates self-shadowing), TEM images of the 18 and 16°C samples are featureless. The wavelength of the ripples is 11.0±1.0 nm for all four temperatures. The measured values for the wavelength are in good agreement with the value of 11-12 nm found by x-ray diffraction (Janiak et al., 1979; Wack and Webb, 1989) or freeze-fracture TEM of DMPC (Luna and McConnell, 1977; Ruppel and Sackmann, 1983; Zasadzinski and Schneider, 1987).

Survey STM scans of the 23°C and the 20°C samples show that the STM and TEM images have the same defect patterns and general features (Fig. 8a, b). To quantify the ripple amplitude and wavelength, roughly twenty 250 by 250 nm STM images with the ripples oriented in a single direction were analyzed for each sample to quantify the ripple features observed by TEM. At this image size, each ripple has about 20 pixels per wavelength (images are 512 by 512 pixels). Figs. 8a and b are representative images from the 23°C and the 20°C samples. Samples prepared at 18 and 16°C showed no modulated textures. Fourier transforms of both the smaller and larger scale images were used to evaluate the ripple wavelength, which was consistent with the TEM results (see Table 3). STM images suitable for analysis were averaged using a structure function program described earlier on a Silicon Graphics Indigo computer. A test area size of 32 by 32 pixels was used to insure that the test area contained at least one complete ripple. Areas that showed a minimum square deviation from the test image were then averaged to make a composite image (Woodward et al., 1994). The composite image was then used as the test image to determine a new composite and eliminate any bias in the original choice of test image. Typically 1000-4000 areas were used to make a composite. For any given image,
we determined four different composites using 4 different test areas chosen arbitrarily from the original image. There was no significant difference in the composites that came from the different initial test areas.
Imaging soft materials with STM

![Image of a graph showing the autocorrelation of the sample quenched from 23°C and cross sections along and perpendicular to the ripple direction.](image)

**Figure 11:** (a) Autocorrelation of the sample quenched from 23°C shown in Fig. 9a. (b) Cross sections of (a) along (top) the ripple direction and (bottom) perpendicular to the ripple direction. The lack of modulation perpendicular to the ripple direction rules out any significant secondary modulation seen previously (Zasadzinski et al., 1988).

et al., 1988). X-ray results have suggested that this secondary ripple may be the result of molecular tilt perpendicular to the ripple wavevector (Hentschel and Rustichelli, 1991). However, Fourier transforms of our STM and TEM images do not have any peaks other than those of the primary ripple. This secondary ripple, and the larger amplitudes are likely artifacts caused by the height amplification discussed earlier as those images were taken in air. As an additional check for the secondary ripple we looked at autocorrelations of the STM images and found no periodicity perpendicular to the primary ripple. Fig. 11 shows the autocorrelation and cross-sections from the image shown in Fig. 9a.

Our combined STM and TEM results show that in excess water, bilayers of DMPC have uniaxial, asymmetric ripples with a temperature dependent amplitude, varying from 2.4 nm peak to peak near the chain melting temperature to near zero near the chain crystallization temperature. However, the ripple wavelength of 11 nm does not change with temperature. The asymmetric ripple shape is consistent with a recent theory that suggests the ripple is the result of a coupling between molecular tilt and bilayer bending (Lubensky and MacKintosh, 1993). This theory depends, to some extent, on the chiral nature of DMPC although experimental results show that the ripple is unchanged when a racemic mixture of DMPC is used (Zasadzinski, 1988b; Katsaras and Raghunathan, 1995) However, the temperature dependence of the ripple amplitude has not previously been observed experimentally nor been predicted theoretically. However, it is not unexpected as the bend elasticity of the membrane is expected to decrease dramatically over the transition from crystalline to fluid bilayers as the temperature is raised through the \( P_r \) phase. Both STM and TEM show that the wavelength of the ripples is unchanged at all four temperatures. This indicates that models of the ripple phase based on a fixed offset of neighboring molecules due to packing frustration must be modified (Schwartz et al., 1994). Such theories predict that the wavelength of the ripple would scale with the amplitude, in contradiction to our results. However, a change in molecular tilt with temperature might change the vertical component of a fixed molecular offset as the temperature increases from the tilted \( L_{p} \) phase to the untitled \( L_{c} \) phase. We have also carefully investigated the possibility of a secondary ripple normal to the first set of ripples by examining autocorrelations of the ripple images. No secondary ripple structure was found, indicating that previous indications of such ripples were likely artifacts (Zasadzinski et al., 1988).

Summary

The freeze-fracture replica technique is uniquely valuable to extend the utility of STM to imaging bulk suspension, liquid crystals, and other biological materials in excess water. The technique is substrate-free and can examine interior interfaces impossible to see with other scanned probe microscopies at high resolution. The quantitative, three dimensional information available by this technique is a necessary extension of TEM images of replicas, especially for situations where a repetitive, three-dimensional structure is expected and quantitative information about the third, or vertical dimension is
needed. By incorporating the necessary modifications described in this paper, replicas for STM can be made as easily as those for TEM with no significant additional investment. We have found that it is necessary to do our STM imaging of these replica surfaces in as dry an atmosphere as possible to inhibit amplification of surface features. However, images of the STM-replicas are reproducible and simple to interpret once these precautions are taken. On any particular replica, the granularity of the replica film limits resolution to about 2 nm laterally and about 1 nm vertically. However, the effects of the granularity can be removed by scanning over a large area, especially for periodic surfaces. The ultimate resolution of the technique for the ripple phase of phospholipids is 1 nm lateral and 0.3 nm vertical resolution. The replica/STM technique is, therefore, certainly as high resolution as direct STM or AFM imaging of supported biological materials, and can often be easier to interpret and reproduce. It can also be used as a check for tapping mode AFM images of soft surfaces.

References


Crepeau RH, Fram EK (1981) Reconstruction of imperfectly ordered zinc-induced tubulin sheets using


J.T. Woodward and J.A. Zasadzinski


Imaging soft materials with STM


Discussion with Reviewers

H.J.K. Hörber: It is stated that Chiruvolu et al. (1994) determined the resolution of freeze-fracture electron microscopy, due to the film grains, was limited to about 2 nm. In this paper, the authors claim 1 nm lateral and 0.3 nm vertical resolution. How does this fit together?

P.M. Frederik: In the abstract the resolution of the STM replica technique is stated to be better than 1 nm, whereas in the summary a more precise statement is given; 1 nm lateral resolution and better than 0.5 nm vertical resolution. In the Introduction this is further refined and better than 0.5 nm turns out to be 0.3 nm. Please clarify.

Authors: On any individual image, the apparent resolution is limited by the granularity of the replicating film to about 2 nm laterally, and somewhat less than this vertically. However, for periodic surfaces, such as the ripples examined here, image analysis techniques such as correlation averaging, can be used to remove much of the effect of the replica grain. By averaging over many images in a consistent way, we have found that the average shape of the ripple can be determined much more precisely - a statistical analysis of the data shows that after fitting the ripple shape to a series of harmonics, the standard deviation of the amplitude and wavelength of the ripple reduce to ± 1 nm laterally and ± 0.3 nm vertically. Similar enhancements in lateral resolution are possible in TEM imaging as well by correlation averaging.

H.J.K. Hörber: If one tries to clean a metal surface that once was in contact with organic material, one notices that the only way to clean it completely is to remove some layers of the metal. Is this done in the cleaning procedure for replicas? If yes, how does this affect the resolution? If not, is it really possible to be sure that contaminations will not affect height measurements, which only on a homogeneous material will give exact results for STM height measurements.

Authors: Our cleaning procedures are designed to be quite gentle, and we have no indication that we are removing any of the metal of the replica; there is only about 1-2 nm of platinum to start with, and even removing a few layers would destroy the replicas. However, we do not need the surfaces extremely clean; by extremely we mean that all organic materials are removed from the metal. The STM treats adsorbed organic layers as just part of the tunneling gap between the metal tip and the metal sample; insulating contaminants do not appear to have any effect on the topography of the surface, as long as they are sufficiently thin that tunneling can occur. Chemically and physically, the metal replica itself is homogeneous at the resolution we require. We further insure that our height measurements on replicas are accurate by careful calibration to other replicas with known features such as our Langmuir-Blodgett films (See Fig. 2). The precision and reproducibility of our measurements suggest that our height measurements are accurate.

H.J.K. Hörber: Is there a way to detect crystallization artifacts of the water, which really might occur, as the sample provides in most cases crystallization seeds?

Authors: Crystallization artifacts can occur in poorly prepared samples, and typically manifest themselves by segregation of liposomes to ice crystal grain boundaries. We checked for this with the TEM replicas that were prepared simultaneously with the STM replicas. This was one of the main benefits of the combined STM/TEM study - the TEM quickly showed which samples were frozen properly, fractured properly, and replicated properly. The question of whether or not ice crystals exist at the nanometer scale is much harder to address. It is clear from direct TEM images of water frozen in a similar way that the water is amorphous (Dubochet and McDowall, 1981). The lipid bilayers have a significantly higher viscosity and lower diffusivity that water, hence, we do not expect any significant rearrangement of the lipid structure.

P.M. Frederik: A roof-tile shape of ripples in DMPC is described and analyzed. How to the authors accommodate areas with line defects in their averaging routine?

Authors: The first step of the correlation averaging
routine is to identify areas that have a strong correlation to the "average" feature. Areas of imperfection, either due to defects in the ripples or to defects in replication, have low correlation to the average feature and are eliminated from further averaging by using a cutoff value for the correlation.