Determining the molecular-packing arrangements on protein crystal faces by atomic force microscopy

Huayu Li,[a] **Mary A. Perozzo,**[b] **John H. Konnert,**[b] **Arunan Nadarajah**[a]* **and Marc L. Pusey**[c]

[a]Department of Chemical and Environmental Engineering, University of Toledo, Toledo, OH 43606, USA, [b]Laboratory for the Structure of Matter 6030, Naval Research Laboratory, Washington, DC 20375, USA, and [c]Biophysics Branch ES76, NASA/Marshall Space Flight Center, Huntsville, AL 35812, USA

Correspondence e-mail:
arunan.nadarajah@utoledo.edu

Previous atomic force microscopy (AFM) studies and periodic bond-chain (PBC) analyses of tetragonal lysozyme crystals have suggested that the (110) face consists of chains of molecules related to one another by 4_3 axes parallel to the crystal face. In this study, high-resolution AFM images of the (110) face were obtained and analyzed in order to verify this prediction. A computer program was employed which constructs the theoretical AFM image corresponding to a specific crystallographic molecular-packing arrangement and AFM tip shape. The packing arrangement and tip shape were varied in order to obtain the maximum possible correlation between experimental and theoretical images. The prediction from PBC analysis of an arrangement involving 4_3 helices was confirmed in this manner, while the alternate arrangement, consisting of molecules related to one another by 2_1 axes, was not observed. However, the surface structure was found to differ significantly even from this crystallographic arrangement. The molecules were found to pack slightly closer about what will become the 4_3 axes within the interior of the crystal, suggesting the occurrence of surface reconstruction or rearrangement on the tetragonal lysozyme (110) face. This study represents a new approach for more precise determination of the molecular-packing arrangements on protein crystal faces employing AFM.

1. Introduction

The advent of atomic force microscopy (AFM) has had a significant impact on investigations of the growth of single crystals in solution. Not only do AFM techniques have a high resolution, but they are also particularly well suited for *in situ* observation of the crystal-growth process in solution when employed with a fluid cell. Durbin and co-workers pioneered the use of this technique for protein crystals and were able to observe the motion of growth steps and growth by two-dimensional nucleation and screw dislocations (Durbin & Carlson, 1992; Durbin *et al.*, 1993). Konnert *et al.* (1994) were the first to observe individual molecular features on the surface of a growing protein crystal and to tentatively identify the structure of the exposed surface. The recent investigations of McPherson and co-workers (Land *et al.*, 1995; Malkin, Land *et al.*, 1995; Malkin, Kuznetsov *et al.*, 1995; Malkin, Kuznetsov, Glantz *et al.*, 1996; Malkin, Kuznetsov & McPherson, 1996a,b; Kuznetsov *et al.*, 1997; Malkin *et al.*, 1997; Ng *et al.*, 1997) and others (Yip & Ward, 1996; Baker *et al.*, 1997; Yip, Brader *et al.*, 1998; Yip, DeFelippis *et al.*, 1998) have added a wealth of information on the growth mechanisms of several proteins by observing two-dimensional and three-dimensional nucleation, defect formation, contaminant inclusion and molecular details.

Additionally, these investigations showed that the growth of these crystals does not always proceed by monolayers as expected. Instead, the heights of the growth steps on the crystal faces usually correspond to the repeating unit in that direction, which is often larger than a monolayer. For example, bimolecular growth steps were observed on the (110) face of tetragonal lysozyme crystals (Durbin & Carlson, 1992; Konnert et al., 1994), which is the minimum repeating unit in that direction (Nadarajah & Pusey, 1996). However, growth proceeds by monomer growth steps on the (101) face (Durbin & Carlson, 1992), corresponding to its repeating unit (Nadarajah & Pusey, 1996). Similarly, the (101) face of thaumatin crystals also displays bimolecular growth steps (Malkin, Kuznetsov, Glantz et al., 1996). This suggested that either the molecules which make up a repeating unit or the unit cell are nucleated on the surface during each growth event, or that aggregates corresponding to these units are first formed in solution prior to attachment on the crystal face. These and other observations have significantly advanced our understanding of protein crystal growth.

Despite these advances, important questions still remain. Some of the observations made on the growth process mentioned above can be thought of as being mesoscopic: that is, they are almost on the molecular scale but can still be described by continuous means. In particular, the measured growth rates were modeled employing continuum theories, and macroscopic quantities such as step free energies were determined (Land et al., 1995; Malkin, Kuznetsov et al., 1995; Malkin, Kuznetsov, Glantz et al., 1996). However, continuum models or methods of observation will not explain the growth process at the true molecular level. For example, they do not explain why growth steps and growth processes appear to proceed by multilayers rather than monolayers. Investigations at the molecular level are needed to understand such processes.

Periodic bond-chain (PBC) theory provides a formal means of investigating the molecular-growth mechanisms of crystal faces and their relationship to the internal structure of crystals. This theory assumes that crystal growth proceeds by the formation of consecutive bonds between the crystallizing units (Hartman & Perdok, 1955a,b,c; Hartman, 1987). These bonds form repeating (periodic) chains throughout the crystal. From the assumption that growth proceeds by the preferential formation of the stronger PBCs over the weaker ones, the growth mechanism of the crystals can be deduced from the relative magnitudes of these periodic bond chains.

The structure of tetragonal crystals of chicken egg-white lysozyme was recently analyzed employing PBC theory (Nadarajah & Pusey, 1996; Strom & Bennema, 1997a,b). The analysis revealed the existence of two sets of intermolecular bonds: one set connecting the molecules in a regular helix centered along the 4_3 crystallographic axes and the other connecting the molecules in an irregular helix centered along the 2_1 axes. The PBCs corresponding to each of these two sets of bonds were complete in themselves and mutually exclusive. In other words, tetragonal lysozyme crystals could be constructed by either one of these two sets, but not by both.

Because the interactions forming the 4_3 helices were stronger, the analysis predicted that tetragonal lysozyme crystals were formed by the PBCs corresponding to these helices only.

The above molecular mechanism was shown to explain numerous observations made on the growth of tetragonal lysozyme crystals, such as bimolecular growth step heights and anisotropic growth kinetics on the (110) faces (Nadarajah & Pusey, 1996; Nadarajah et al., 1997). This provides indirect confirmation of the validity of the mechanism and the PBC analysis used to obtain it. However, the mechanism has not yet been confirmed by direct experimental observations. Such a confirmation is necessary, as earlier simplified analyses of the crystal structure had suggested that growth proceeded by the alternate mechanism: the formation of 2_1 helices (Durbin & Feher, 1990; see also the discussion in Nadarajah & Pusey, 1996).

The molecular-growth mechanism of tetragonal lysozyme crystals discussed above also specifies that the molecular-packing arrangement on the (110) face is a unique one, consisting of rows of 4_3 helices. This offers the possibility of directly verifying this mechanism by AFM in the high-resolution mode, where the scanned area is in the order of nanometers or tens of nanometers on one side. This is too fine a resolution to observe large surface features, such as the dislocations and two-dimensional islands seen in earlier studies (Durbin & Carlson, 1992; Malkin, Kuznetsov, Glantz et al., 1996; Malkin, Kuznetsov & McPherson, 1996b), but is adequate to resolve individual protein molecules on the crystal faces.

Attempts have been made in some recent studies to discern the molecular-packing arrangements from such high-resolution AFM scans (Yip & Ward, 1996; Baker et al., 1997). However, it is possible to obtain much more information from these scans than was attempted in these studies. If the structure of the crystal and the molecular structure of the protein are known, the predicted AFM image for a given molecular-packing arrangement can be constructed. Such predicted AFM images can then be compared quantitatively with an actual high-resolution AFM image to obtain the correct packing arrangement. This approach of determining the surface-packing arrangements on protein crystals was pioneered by Konnert and co-workers for the (110) face of tetragonal lysozyme crystals (Konnert et al., 1994). Their AFM images were of sufficient quality to indicate that the face is relatively smooth, with periodic repeats which are consistent with the internal structure of the crystal. Of the two possible surface-packing arrangements, the AFM data was found to be more consistent with that containing the 4_3 helices, but the study was not definitive.

In this study, we further develop the method of determining the molecular-packing arrangements on protein crystal faces employing high-resolution AFM. The techniques of Konnert and co-workers for the quantitative modeling of AFM images, which have been further refined in recent years, will be applied to new high-resolution AFM images. We investigate the (110) face of tetragonal lysozyme crystals to verify the packing arrangement suggested

by the earlier AFM data and predicted by the PBC analysis.

2. Construction of predicted AFM images

Figs. 1(a) and 1(b) illustrate, in a simplified manner, the molecular packing of tetragonal lysozyme crystals and the two possible packing arrangements on the (110) face. The dotted line in Fig. 1(a) is the (110) face produced if 2_1 helices are the dominant building blocks of the crystal. The solid line in this figure shows the (110) face formed by 4_3 helices, which were found by the PBC analysis to be strongly bonded and most likely to control the growth process (Nadarajah & Pusey, 1996; Strom & Bennema, 1997a,b). The latter arrangement is also illustrated in Fig. 1(b).

Since the high-resolution molecular structure of tetragonal lysozyme is available, the surface morphology of the idealized crystal face corresponding to each of the above two molecular-packing arrangements can be calculated (Konnert et al., 1994). The coordinates for each atom of the reference molecule in the asymmetric unit is given by Protein Data Bank file 193L (Young et al., 1994). For the $P4_32_12$ space group of tetragonal lysozyme, the reference molecule can be suitably translated and rotated to produce the packing arrangements illustrated in Fig. 1(a). The coordinates of all atoms in the two arrangements can then be determined. A sphere of 2–3 Å radius representing the van der Waals radius will be placed at each atom location in order to obtain the surface morphology of the crystal face.

Although it may seem that images obtained by AFM and associated techniques show the morphology of the scanned surface, this is not the case. Since these images are obtained by scanning the surface with a tip, the resulting image depends on both the tip shape and the surface morphology. We shall refer to these images as images of the tip-accessible surfaces. When the AFM is used in the low-resolution mode to study surface features, as in most previous studies of protein crystal growth (Durbin & Carlson, 1992; Malkin, Kuznetsov & McPherson, 1996b), the scan resolution is in micrometres, which is much larger than the tip radius. In such instances the tip-accessible surface can be assumed to correspond to the actual surface morphology. However, when attempts are made to resolve individual molecules with AFM in the high-resolution mode, this assumption is no longer valid.

There are two possible ways to overcome this drawback. If the shape of the AFM tip is known (from electron-microscopy scans or by scanning a standard substrate), then deconvolution routines may be used to obtain the true surface morphology (Markiewicz & Goh, 1994; Wilson et al., 1995; Williams et al., 1996; Villarrubia, 1997). This is the usual approach to obtain the true surface morphology from a high-resolution AFM scan. For example, in biological applications it has been used to obtain the three-dimensional shape of blood proteins adsorbed on surfaces (Siedlecki et al., 1996) and even to observe the activity of a DNA strand on a surface (Kasas et al., 1997). This approach does not produce a unique solution and the surface shape obtained by this procedure is of relatively low resolution. For some applications, such as in determining the approximate shape of an adsorbed protein molecule on a surface, this approach is adequate. For the application considered here, where the AFM scans are being employed to distinguish between two molecular-packing arrangements of the same protein, an approach which more uniquely defines the shapes to be compared is needed.

The second approach to overcoming this problem satisfies this requirement. Konnert and co-workers first suggested that the theoretical tip-accessible surface be constructed with an assumed molecular model and a tip defined by one or two parameters (Konnert et al., 1994). The parameter(s) for the tip and the relative positions of the experimental and theoretical images are adjusted to maximize the correlation between the two. This approach requires that the precise morphology of the surface be assumed, which is generally not possible except for crystalline surfaces. For example, such an approach cannot be employed to image the shape of adsorbed protein or nucleic acid molecules on a surface referred to earlier (Siedlecki et al., 1996; Kasas et al., 1997).

The tip-accessible surface is illustrated in two dimensions in Figs. 2(a) and 2(b). On the left in Fig. 2(a), the morphology of a section of a hypothetical protein crystal face, which can be constructed from crystallographic data, is shown. This illustrates two molecules, or two sections
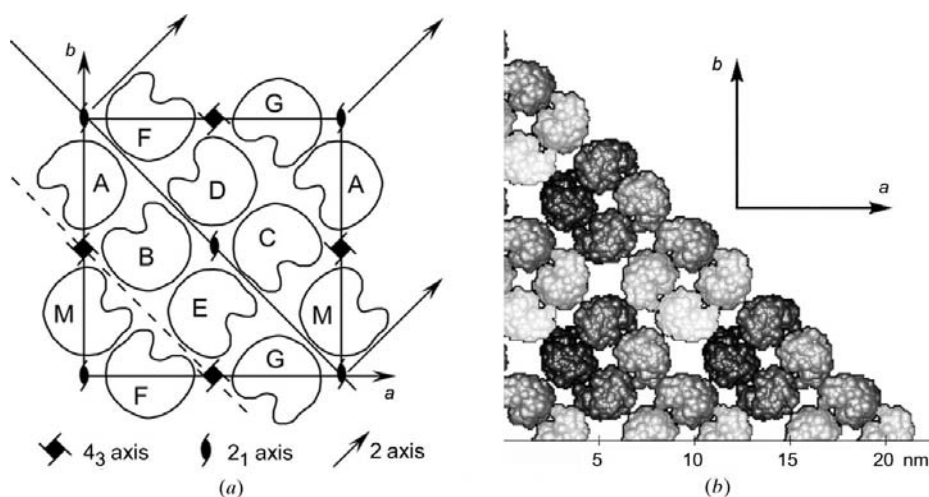

Figure 1
(a) The unit cell of tetragonal lysozyme crystals showing a simplified representation of the eight molecules labeled A–G, with the reference molecule labeled M. The molecules are related to one another by the 4_3 and 2_1 screw axes as well as by the twofold symmetry axes. The solid and dotted lines show the two possible packing arrangements on the (110) face. (b) The molecular-packing arrangement of the (110) face, corresponding to the solid line in (a) showing the 4_3 helices. Each molecule is drawn at 70% of its original size centered at its true position. The packing arrangement for the dotted line in (a) is obtained by removing the layer of surface molecules in (b).

of a single molecule, with individual atoms visible on them. This figure also shows a hypothetical 'infinitely sharp' AFM tip (Fig. 2a). It is easy to see that this procedure results in the loss of some atomic information. Fig. 2(b) shows the result of employing a more realistic AFM tip on the same surface. This results in the loss of almost all atomic information, but the individual molecules or the two sections of a single molecule can still be distinguished in the resulting image.

We will briefly describe the features of the computer program used to compute the theoretical images (Konnert *et al.*, 1994). The sequence performed is as follows.

(i) Read the atomic coordinates of the molecules in the asymmetric unit from a Protein Data Bank file.

(ii) Perform the translations and rotations of the reference molecule required to obtain a given molecular-packing arrangement on the crystal face and determine the atomic coordinates of the entire set of molecules.

(iii) Determine the surface morphology (*i.e.* calculate the height at any given position on the surface) from all the atomic coordinates and the atomic radii.

(iv) Determine the tip-accessible surface for a given AFM tip shape from the surface morphology in (iii).

(v) Convert the convoluted surface-morphology data into a file which can be read by an AFM analysis program, in this case the program driving a Digital Instruments Nanoscope IIIa scanning-probe microscope, and displayed like an image from an actual AFM scan.

This computer program was then used to create the expected high-resolution AFM images for the two possible molecular-

Figure 2
Illustration of the convolution process between the AFM tip and the crystal surface to produce the predicted AFM image. A hypothetical infinitely sharp tip produces (*a*) an image with greater surface detail than (*b*) an image produced with a more realistic AFM tip.

packing arrangements on the (110) face shown in Fig. 1(*a*). These two images were compared with the images obtained from actual AFM scans on the (110) face of tetragonal lysozyme.

3. Construction of predicted AFM images

3.1. Protein preparation and crystallization

Chicken egg-white lysozyme was purchased from Sigma and repurified by cation-exchange and size-exclusion chromatography as previously described (Ewing *et al.*, 1996). The final protein solutions were maintained at pH 4.0 with 0.1 M acetate buffer. Protein concentrations in these solutions were determined by UV absorbance (Aune & Tanford, 1969). Tetragonal crystals were grown in specially designed cells at room temperature ($\sim$293 K) in 20–40 mg ml^{-1} protein solutions with 5% NaCl. Following crystallization, the remaining protein solution was drained and the cell plate with the crystals was transferred to the AFM fluid cell. The fluid cell was then filled with fresh protein solution of a known concentration close to the solubility limit ($\sim$3.5 mg ml^{-1}) under these conditions (Cacioppo & Pusey, 1991). The system was allowed to reach equilibrium before the AFM scans were begun.

3.2. Performing high-resolution AFM scans on crystal faces

The high-resolution AFM scans were carried out with a Digital Instruments Nanoscope IIIa scanning-probe microscope. The instrument was calibrated by scanning freshly cleaved mica at atomic resolution and measuring the interatomic distance. The distance from three separate scans yielded an average value of 5.5 Å, while the standard value is 5.2 Å. The ratio of these two values was used to scale all the images from AFM scans of protein crystal faces.

The details of performing AFM scans on protein crystals are described in many publications (*e.g.* Konnert *et al.*, 1994). All scans were carried out *in situ* at room temperature in the contact mode. Silicon nitride tips were used throughout and the scan frequency was around 10 Hz. The square scan area ranged from edges of 25 to 50 nm on the (110) face of tetragonal lysozyme crystals.

3.3. Averaging the experimental data

The conclusions which can be drawn from interpreting an AFM image are limited by the quality of the images. An example of the experimental data from an AFM scan is shown in Fig. 3(*a*) and consists of a 512 × 512 array of surface height information. This image was produced from the raw experimental data by the application of a 5 × 5 median filter followed by convolution with a Gaussian. Here, a 5 × 5 median filter replaces each point in the image with the median value of a cell surrounding that point, while the Gaussian employed was $\exp(-i^2/2\sigma^2)/\sigma(2\pi)^{1/2}$, where i is in units of pixels and the standard deviation σ is 3 Å.

At this point no periodicity has been imposed on the image, although it is clear that the figure displays near-periodicity in two nearly orthogonal directions. It was assumed that the
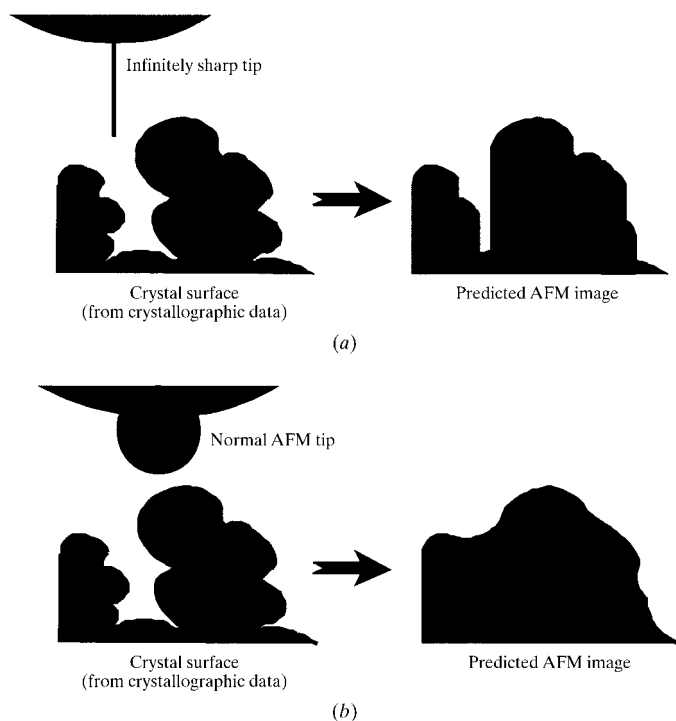
image would display the same or multiples of the periodic dimensions of the bulk crystal (see below). With this assumption, the nearly orthogonal axis in Fig. 3(a) should be orthogonal for the (110) faces of tetragonal space groups ($\alpha = \beta = \gamma = 90°$). The approximately 2° offset from perpendicular between these axes was attributed to sample drift and corrected for with a single parameter. The experimental image was first rotated so that one crystallographic or periodic direction coincided with the horizontal axis of the figure. Each horizontal line of this image was then shifted so that the second crystallographic or periodic direction was vertical. The resulting image is shown in Fig. 3(b).

The next step was to assume that the surface is periodic and obtain the periodic structure. This is necessitated by the fact that the theoretical image is perfectly periodic, while the experimental image cannot be expected to be perfectly periodic owing to instrumental distortions, vibrations and sample drift. To enable a comparison between the two, a periodic image must be constructed from the experimental data by a suitable averaging procedure. The objective was to average the information contained in the experimental image in such a way as to obtain the most accurate image possible of a single unit cell. FFT filtering (which is often used for this purpose) Fourier transforms the image, applies a periodic mask and back-transforms to obtain the filtered image. However, since the image is not periodic owing to the above-mentioned errors, the approach described below can yield a more accurate averaged unit cell.

The method employed here was to use the two-dimensional unit cells, which can readily be discerned as illustrated in Fig 3(b). An average unit cell was obtained from the many contained in this figure by determining the cell which maximized the correlation coefficient between it and the individual cells. The quantitative measure of agreement between the experimental image and the image obtained from the averaged unit cell was taken to be the correlation coefficient, CC, defined by

$$CC = \frac{\left(\sum_{x=\text{surface}} A_x B_x \right)}{\left(\sum_x A_x^2 \sum_x B_x^2 \right)^{1/2}}. \quad (1)$$

Here, A and B represent the height values at the x coordinate on the two-dimensional projection of the scanned surface, scaled to have mean values of zero, for the two images designated A and B. The value of CC may range from 1.0 for perfect correlation through 0 for no correlation and -1.0 for perfect anti-correlation. The sequence of steps used to obtain the averaged image with the maximum correlation coefficient was as follows.

(i) Choose a unit cell, such as that shown in Fig. 3(b).

(ii) Optimize the correlation between this unit cell and a unit cell just above it by shifting the first unit cell to approximately coincide with a second unit cell; the correlation between the two is then maximized by translating the images both horizontally and vertically relative to one another in increments of 0.6 Å over a range of ± 5 Å (256 positions) to find the translations corresponding to
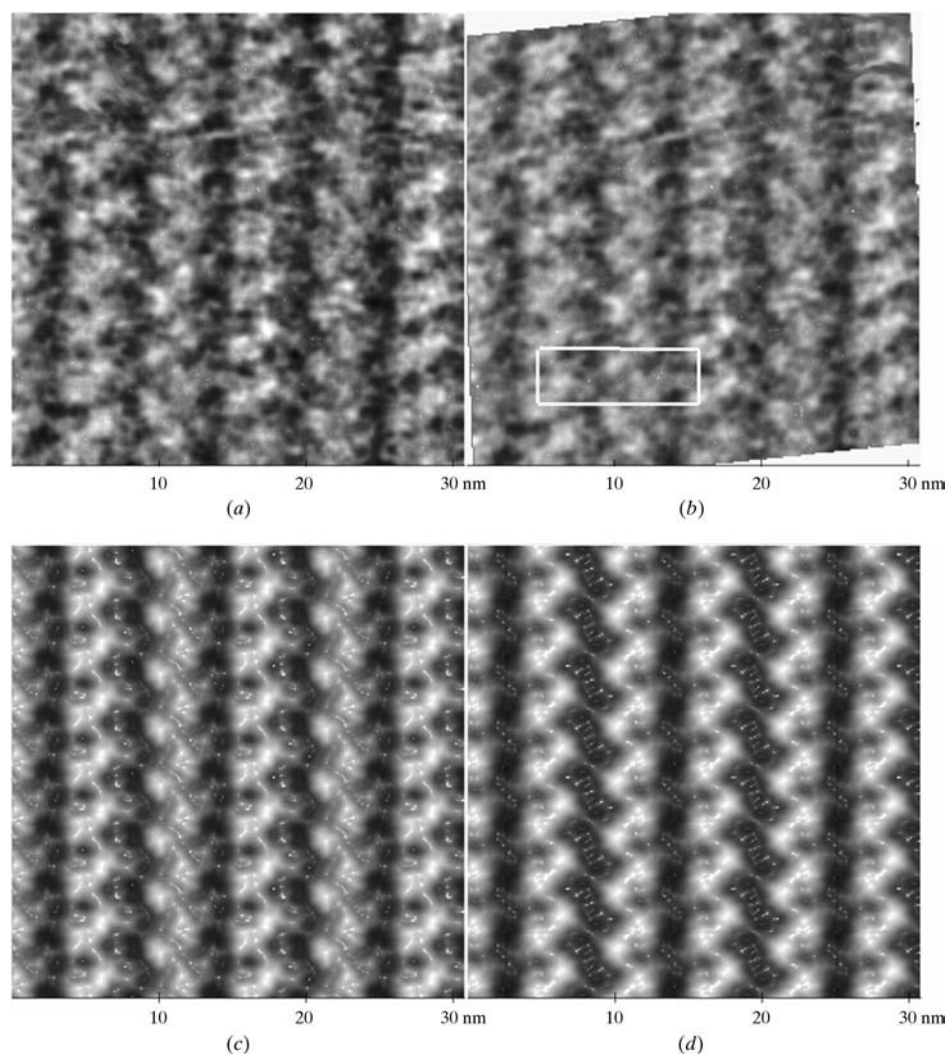

Figure 3
Sequence of images showing the averaging of the experimental AFM image. (a) The resulting image after the application of a median filter and a Gaussian convolution to the raw data. (b) The image in (a) after rotation and correction of the the 2° skewness to make the 4_3 axes and the unit cell as vertical as possible, where the dimension of the two-dimensional unit cell shown is 112×38 Å. (c) The image obtained by translationally averaging the image in (b). (d) The twofold averaged image obtained from the image in (c). Note that in these and subsequent AFM images the dark areas are of lower elevation and the lighter ones of a higher elevation.

maximum correlation. The two images are then averaged.

(iii) Repeat step (ii), correlating the remaining cells with the accumulating average image.

(iv) Repeat (i)–(iii) starting with the average image and cycle until convergence is reached.

The result of this averaging operation on Fig. 3(b) is shown in Fig. 3(c), obtained by averaging over 11 unit cells. For this

(a)

(b)

Figure 4
Theoretical AFM images of the (110) face for a hypothetical infinitely sharp tip. (a) Image for the packing arrangement containing 4_3 helices on the surface (solid line in Fig. 1a) and (b) image for the packing arrangement containing 2_1 helices on the surface (dotted line in Fig. 1a).

case, the average value of all correlation coefficients between the individual cells in Fig. 3(b) and the average unit cell of Fig. 3(c) was 0.88.

The crystal structure has twofold rotation axes perpendicular to the (110) face which relate adjacent 4_3 helices. The highest quality experimental images possessed very nearly this symmetry on the surface. Therefore, the assumption was made that twofold symmetry was present and, in addition to the translational averaging described above, a further twofold averaging was carried out. For the case of Fig. 3(c), this would involve rotating the image by 180° and averaging the rotated image with the unrotated Fig. 3(c). The resulting twofold-averaged image is shown in Fig. 3(d). The correlation coefficient between Fig. 3(c) and its twofold image was 0.72. The average correlation coefficient between Fig. 3(d) and all the individual cells in Fig. 3(b) was 0.82.

The dimensions of the averaged two-dimensional unit cell in Fig. 3(d) are 112×38 Å. The range of reported dimensions of the three-dimensional unit cells of tetragonal lysozyme crystals is from $78.4 \times 78.4 \times 37.3$ to $79.6 \times 79.6 \times 38.3$ Å, with accepted values of $79.3 \times 79.3 \times 38$ Å (Salunke et al., 1985). This translates to a range of 110.9×37.3 to 112.6×38.3 Å for the two-dimensional unit cell for the (110) face. The dimensions of the averaged cell obtained here fall well within this range and are extremely close to the accepted values of 112.1×38.0 Å. Therefore, it was assumed that there was no difference in the repeating-unit dimensions of the bulk crystal obtained from X-ray crystallography and of the crystal faces as measured by AFM. The same dimensions were employed for the repeating units in the construction of the theoretical images for comparison with the above experimental ones. It should be noted that no adjustable parameters were employed in this averaging procedure and that all experimental images analyzed in this study were averaged in this manner prior to comparison with the theoretical images.

4. Results and discussion

The predicted AFM images for two possible packing arrangements on the (110) face were created as described in §2. Fig. 4(a) shows the predicted AFM image employing a hypothetical infinitely sharp tip on the surface described by cleavage along the solid diagonal line in Fig. 1(a). Individual atoms may be distinguished, but the molecular boundaries are less clear as the molecules are in contact with one another. There are five complete 4_3 helices of bimolecular width in this figure. The axes of these helices run vertically and coincide with the crystallographic c axis.

The predicted AFM image for an infinitely sharp tip for the alternate surface-packing arrangement, described by cleavage along the dotted diagonal line in Fig. 1(a), is shown in Fig. 4(b). Neither Fig. 4(a) nor Fig. 4(b) resembles the averaged experimental AFM images of the (110) face, such as those shown in Figs. 3(c) and 3(d). For example, the correlation coefficient between these images and that of Fig. 3(c) are only 0.27 and 0.12, respectively. The correlation coefficient here is the maximum value of coefficients defined by (1) for all
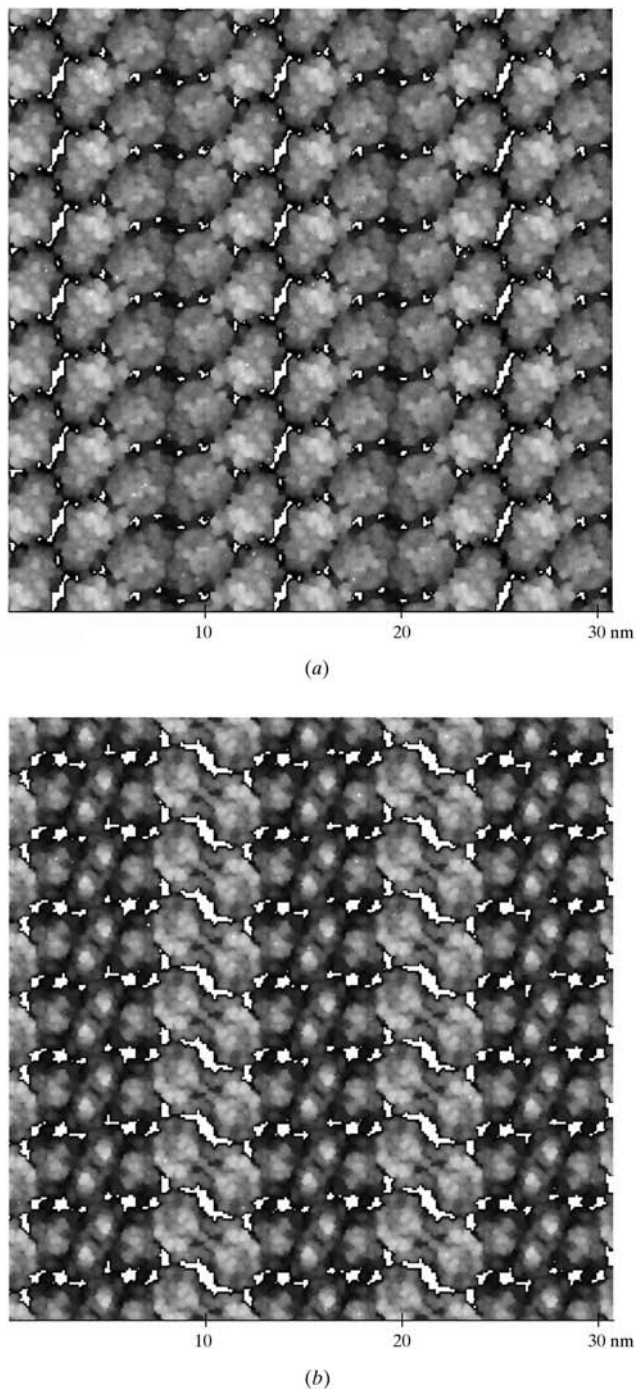
possible translations of the two surfaces relative to one another.

Clearly, an experimental AFM image obtained with real tips cannot be expected to correlate well with a theoretical image constructed assuming the tip to be infinitely sharp. Theoretical images incorporating actual AFM tip shapes need to be constructed. The tips are specified to have a nominal tip radius of 200–400 Å. A spherical tip of this large a radius would not yield the molecular details observed. It is believed that the tip has protrusions of a smaller size which are mainly responsible for the molecular details observed on high-resolution scans (for further discussions on this topic see, for example, Konnert *et al.*, 1994; Baker *et al.*, 1997). For this reason, we constructed a tip model for which a tip of smaller radius protruded from a tip of radius 300 Å. The radius of the smaller tip and the extent to which it protrudes were the two adjustable parameters for the model of the tip. Values for these two parameters were adjusted until a maximum value was obtained in the correlation coefficients between the two images. For the experimental image of Fig. 3(d), a maximum correlation with the theoretical image for the surface shown by the solid line in Fig. 1(a) was obtained for a tip of radius 28 Å protruding 5 Å from the larger sphere. This tip is illustrated in Fig. 5.

The above approach to obtain the tip and surface shapes is referred to as the 'blind tip reconstruction' approach (Williams *et al.*, 1996; Villarrubia, 1997). It should be noted here that the tip radius and protrusion are the only two adjustable parameters employed in this study. The relative position of the experimental and theoretical images is also varied to select the one which produces the maximum correlation. However, a position has to be chosen for this comparison and selecting the one which gives the best correlation is as good as any other. Thus, the relative position of the compared images should not be considered to be an adjustable parameter.

For the experimental image shown in Fig. 3(d), the corresponding theoretical image with a 28 Å radius tip and a 5 Å protrusion is shown in Fig. 6(a). As expected, individual atoms can no longer be resolved in this image, unlike in Fig. 4(a), but the individual lysozyme molecules can clearly be distin-

guished. More significantly, the vertical 4_3 helices of bimolecular width are still clearly visible, as shown by the rectangular blocks representing the molecules which constitute a single

(a)

(b)

Figure 6
(a) Theoretical AFM image for the packing arrangement containing 4_3 helices (solid line in Fig. 1a), with a tip radius maximizing correlation with the image in Fig. 3(d). The molecules M–C–B–A which constitute the tetramer unit corresponding to a single turn of this helix are also shown by rectangles. The rectangles represent the space occupied by each molecule in the tetragonal crystal packing arrangement (for details, see Nadarajah & Pusey, 1996). (b) Image in (a) embedded in the image in Fig. 3(d), where the CC between the images is 0.62. The whole image in (b) has been slightly tilted to show the surface relief in three dimensions.
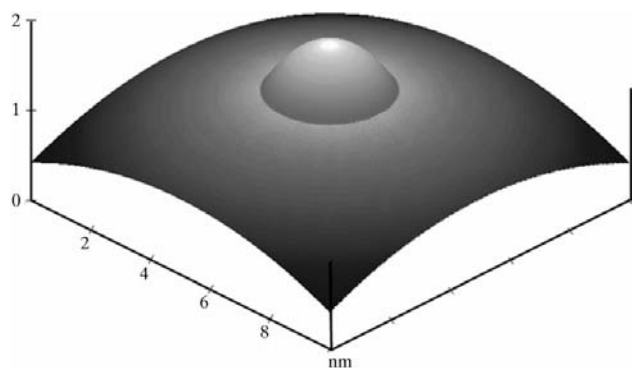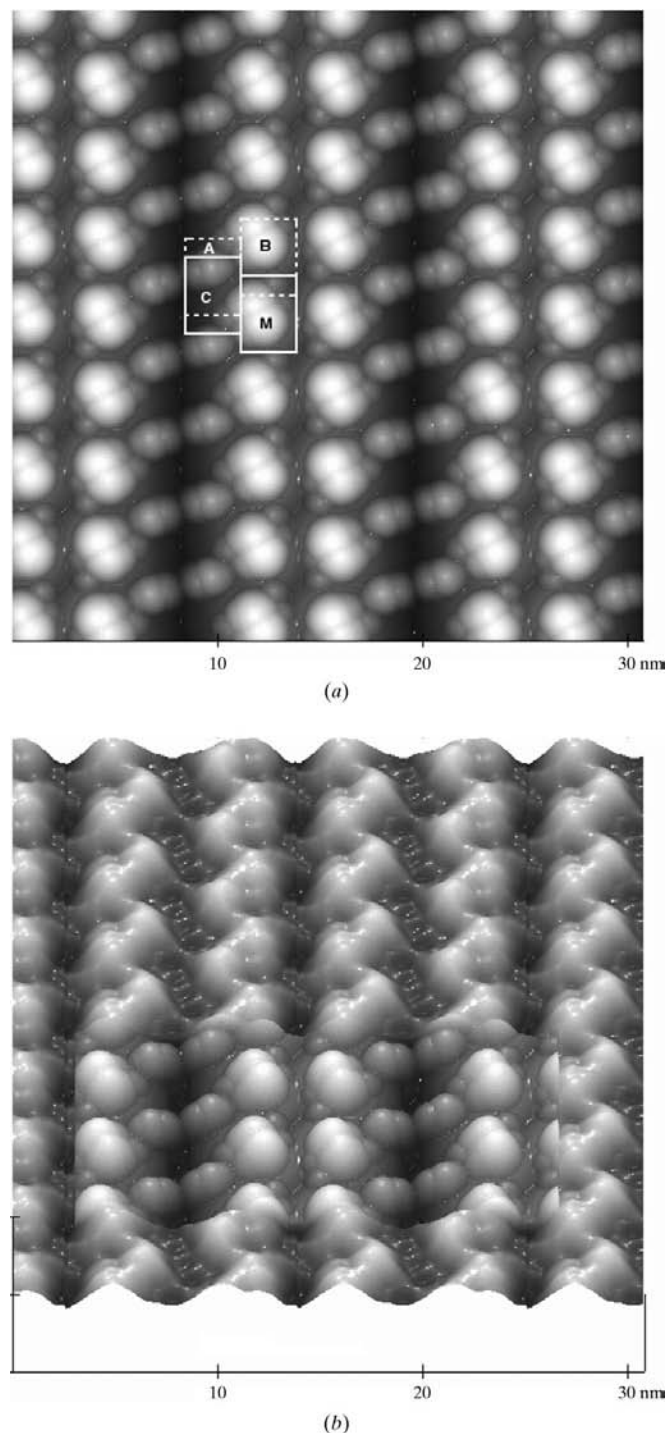

Figure 5
The shape of the AFM tip employed to obtain the theoretical AFM images. The base has a radius of 30 nm, with a protrusion of radius 2.8 nm extending out 0.5 nm from the base.

turn of this helix. However, there is somewhat less separation by dark vertical regions between the individual helices than in Fig. 3(d). Fig. 6(b) shows the theoretical image (Fig. 6a) embedded in the experimental image (Fig. 3d) for comparison. The pronounced resemblance between the theoretical and the
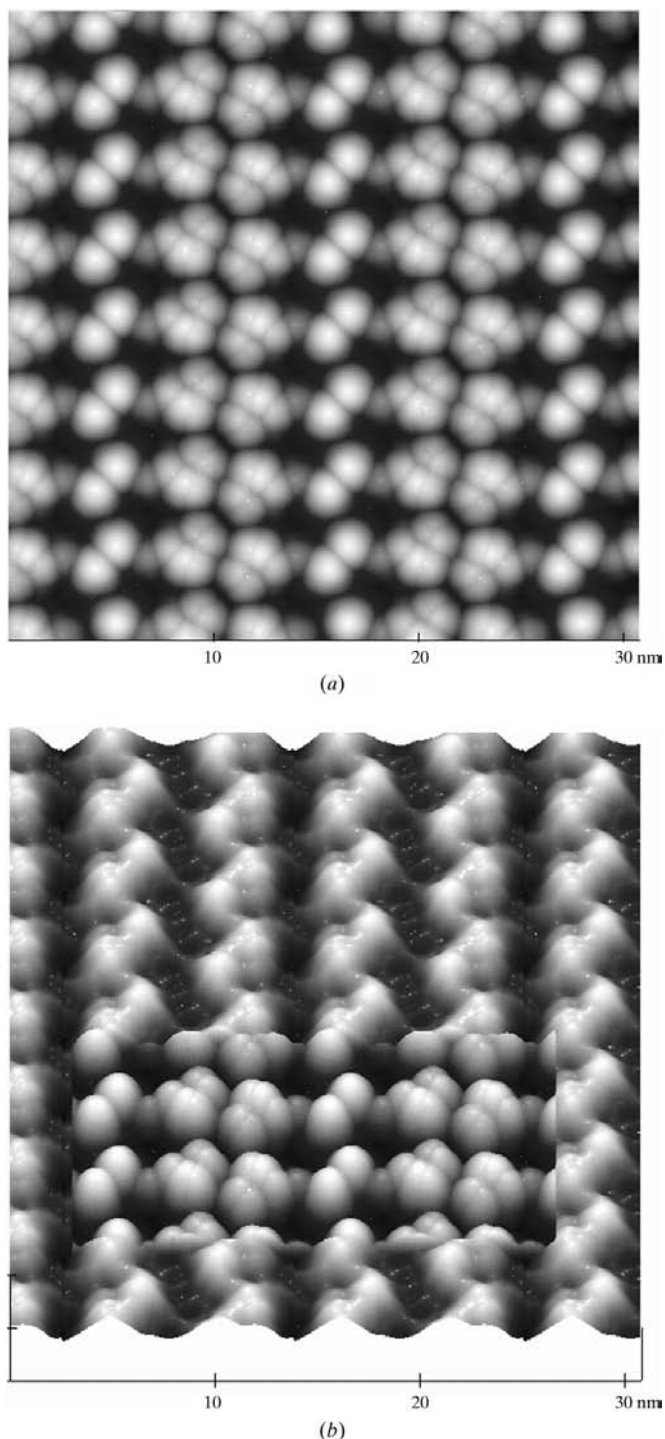

experimental images is visible here. The correlation coefficient for these two images is 0.62.

The predicted AFM image, generated using the alternative packing arrangement (shown by the dotted line in Fig. 1a) and the AFM tip parameters which maximize correlations with the experimental image (Fig. 3d), is shown in Fig. 7(a). The packing is reflective of the side-by-side molecular arrangement of the irregular 2_1 helices (Nadarajah & Pusey, 1996). There are five complete 2_1 helices shown in Fig. 7(a) with their axes aligned with the c axis, but unlike the 4_3 helices shown in Fig. 6(a), they are not easily discernible. These characteristics clearly differentiate Figs. 6(a) and 7(a). Fig. 7(b) shows the theoretical image (Fig. 7a) embedded in the experimental image (Fig. 3d). There is little visual resemblance between these two images and this is reflected by the correlation coefficient between Figs. 3(d) and 7(a), which is 0.25.

The above comparisons strongly suggest that the molecular-packing arrangement on the (110) face corresponds to the formation of 4_3 helices on it, as given by the solid line in Fig. 1. More significantly, all our high-resolution AFM scans on the (110) face only produced images similar to Fig. 3(d). That is, the scanned images resembled Fig. 6(a) and not Fig. 7(a). The resulting correlation coefficients with constructed theoretical images for three different experimental AFM images are given in Table 1 (image A is the one shown in Figs. 3a–3d). Different theoretical images were constructed for each of the three comparisons, as each image was produced with a different tip. Table 1 shows the different optimal tip radii used for each construction. Similar to image A, images B and C did not correlate with the image for the alternate packing arrangement for the (110) face. Other experimental images besides these three were collected. They also showed a close visual resemblance to Fig. 6(a), although the detailed analyses described above for obtaining the correlation coefficients were not carried out.

The visual resemblance, coupled with the correlation coefficient of 0.62 between the averaged experimental image in Fig. 3(d) and the predicted image in Fig. 6(a), is adequate to confirm that the molecular-packing arrangement on the (110) face is comprised of 4_3 helices. However, as described above, the image in Fig. 3(d) was produced by averaging over all unit cells in Fig. 3(b), with an average correlation coefficient of 0.82 between all the individual cells and the averaged one. Thus, the individual unit cells have a strong correlation with each other. Consequently, the 0.62 correlation coefficient between the averaged experimental image in Fig. 3(d) and the theoretical image in Fig. 6(a) suggests a systematic deviation from predictions. Table 1 shows that other experimental images have similar trends.

The images were further compared in detail in order to discern the cause of this deviation. A visual comparison of Figs. 3(d) and 6(a) suggests that in the experimental AFM image the protein molecules are more closely packed along the 4_3 axes, resulting in greater separation between the 4_3 helices. Accordingly, a new theoretical image was constructed by translating the molecules from their crystallographic positions in order to improve the correlation coefficient. The 4_3

Figure 7
(a) Theoretical AFM image for the packing arrangement containing 2_1 helices (dotted line in Fig. 1a), with a tip radius maximizing correlation with the image in Fig. 3(d). (b) Image in (a) embedded in the image in Fig. 3(d), where the CC between the images is 0.25. The whole image in (b) has been slightly tilted to show the surface relief in three dimensions.

Table 1
Results of the detailed analyses of the three experimental AFM images.

Experimental AFM image	A	B	C
Number of individual unit cells in the translationally averaged image	11	13	12
Average CC (and range) of individual cells with translationally averaged cell	0.88 (0.82–0.91)	0.75 (0.67–0.85)	0.85 (0.78–0.89)
CC of translationally averaged cell with its twofold image	0.72	0.66	0.86
Average CC (and range) of individual cells with twofold and translationally averaged cell	0.82 (0.76–0.85)	0.70 (0.60–0.77)	0.83 (0.74–0.86)
CC with theoretical image (molecules in crystallographic positions)	0.62	0.45	0.65
Optimum AFM tip radius that protrudes 5 Å from sphere of 300 Å radius (Å)	28	28	50

Table 2
Construction of a new theoretical image for maximizing correlation with the experimental images.

Experimental AFM image	A	B	C
Distance moved by reference molecule in x direction for maximum in CC (Å)	−1.7	−2.0	−2.4
Distance moved by reference molecule in y direction for maximum in CC (Å)	6.3	5.5	5.9
Distance moved by reference molecule in z direction for maximum in CC (Å)	−4.5	−3.4	−3.0
Total distance moved by reference molecule for maximum in CC (Å)	7.9	6.8	7.0
CC with the new theoretical images	0.93	0.84	0.86
CC with the single new theoretical image (molecules shifted by 7.2 Å)	0.93	0.84	0.83

axes within the face as well as the twofold axes perpendicular to the face were retained for the surface molecules during this refinement. In order to maximize correlation with the experimental image in Fig. 3(d) with these restrictions, it was found necessary to move each molecule ~7 Å from its original crystallographic position for the new theoretical image. As suggested earlier, this resulted in an image with the molecules being more closely packed around the 4_3 axes. This new theoretical image had a correlation coefficient of 0.93 with Fig. 3(d).

This new image is shown in Figs. 8(a) constructed with the optimized AFM tip radius of 28 Å with a protrusion of 5 Å. When Figs. 6(a) and 8(a) are compared, Fig. 8(a) has wider deeper valleys between the 4_3 helices than Fig. 6(a). In Fig. 8(b), this new image is embedded in the averaged experimental image shown in Fig. 3(d). Fig. 8(b) clearly shows how well the molecules of the 4_3 helices in the predicted and experimental AFM images coincide. Table 2 shows that this dramatic improvement in the correlation between the experimental and theoretical images is not an artifact unique to one experimental image. For each of the three images, a theoretical image producing a maximum value of the corre-

lation coefficient was constructed separately by moving the individual molecules from their crystallographic positions in the x, y and z directions. Table 2 shows the distance moved by the reference molecule, with the remaining molecules moving the same distance as dictated by the twofold and fourfold screw symmetries. It can be seen that despite the separate optimizations, in each case the molecule was moved approximately the same distance in each of the three directions. This means that the three new theoretical images would coincide very closely. This suggests that these images tend to correlate well with a single new theoretical image where the molecules have been moved a total average distance of 7.2 Å, resulting in closer packing around the 4_3 axes (they were moved 5.7 Å toward the 4_3 axes). This single image was employed for comparisons with all experimental images.

All three experimental images analyzed show pronounced increases in correlation with this new theoretical image (Table 2). This increased correlation is also seen visually in Figs. 8(c) and 8(d), which correspond to Fig. 8(b) for the experimental images B and C in Tables 1 and 2 (Fig. 8b corresponds to image A). Visual inspection of the other experimental images which were collected but not systematically analyzed also show a greater separation between the 4_3 helices than in Fig. 6(a). This suggests that they too would produce a better correlation with the image in Fig. 8(a) than that in Fig. 6(a).

It should be noted that the high-resolution AFM images analyzed here were collected near the resolution limit of atomic force microscopy. Many of the images collected had to be discarded because they lacked the periodicity expected of molecular arrangements on crystal faces. There could be many reasons for this, but a principal one is instrument limitation. When a new scanning experiment is begun, it takes a significant amount of time for the instrument to stabilize sufficiently to enable the collection of high-resolution images. Images collected prior to stabilization are invariably of poor quality. This problem can be minimized readily by allowing adequate time for scanning experiments. However, the second important cause of incoherent images, namely the accumulation of debris on the AFM tip, is harder to avoid.

The distortion of scanned images owing to tip debris is best illustrated by an image labeled C*. This image was collected on the same crystal face with the same tip shortly after image C was collected. Thus, almost all variables were kept the same

for both images. The filtered experimental image of *C** is shown in Fig. 9(*a*). Similar to image *C*, this figure displays a pronounced periodicity (*cf*. Fig. 3*a*), although visually it seems somewhat more distorted. When the translational averaging was performed, it produced the image shown in Fig. 9(*b*). This image had an average correlation coefficient of 0.89 with the individual unit cells, indicating good translational periodicity comparable with that of image *C* (see Table 1).

However, despite the periodicity, it is obvious that Fig. 9(*b*) is a distorted image when it is compared with the corresponding Fig. 3(*b*). The nature of this distortion is seen when twofold averaging is attempted. The correlation coefficient between Fig. 9(*b*) and its twofold image is only 0.38, indicating that this image lacks twofold symmetry. We next attempted to construct a theoretical image with an optimum tip shape corresponding to Fig. 9(*b*), similar to Figs. 6(*a*), 7(*a*) or 8(*a*). Here, too, little correlation was found between any such theoretical image and Fig. 9(*b*).

In order to produce a correlation between a theoretical image and Fig. 9(*b*), it was necessary to modify the tip shape considerably. A tip shape which preserves twofold symmetry, such as a sphere, will always produce theoretical images with twofold symmetry. Therefore, in order to obtain a theoretical image which will resemble Fig. 9(*b*), it is necessary to employ a nonsymmetrical tip shape. This suggests that the distortion in image *C** arose because of the attachment of debris to the AFM tip disrupting its twofold symmetry. This would have occurred after the completion of the image *C* scan but before the start of the image *C** scan. This was confirmed when a nonsymmetrical tip was employed to construct the theoretical image, along with closer packing of molecules around the 4_3 helices. This image is shown in Fig. 9(*c*) embedded in the experimental image in Fig. 9(*b*). The correlation coefficient between the two images in Fig. 9(*c*) now improves sharply to 0.83, which is the same value as obtained for image *C* (Table 2).

The above result also suggests a reason for the somewhat low agreement between the translationally averaged images of *A*, *B* and *C* and their twofold images, which have an average correlation coefficient of 0.75 between them (see Table 1). Even if the tips have not accumulated any debris, the tips themselves may not be perfectly spherical. The tip shapes obtained by scanning standards also suggest this. This implies that removing the restriction that the tip shape be spherical in the construction of the theoretical image should produce even greater agreements between the experimental and theoretical images. This was performed for image *A* and a theoretical image was produced with a tip which deviated slightly from a perfect spherical shape. This image is shown in Fig. 10, embedded in
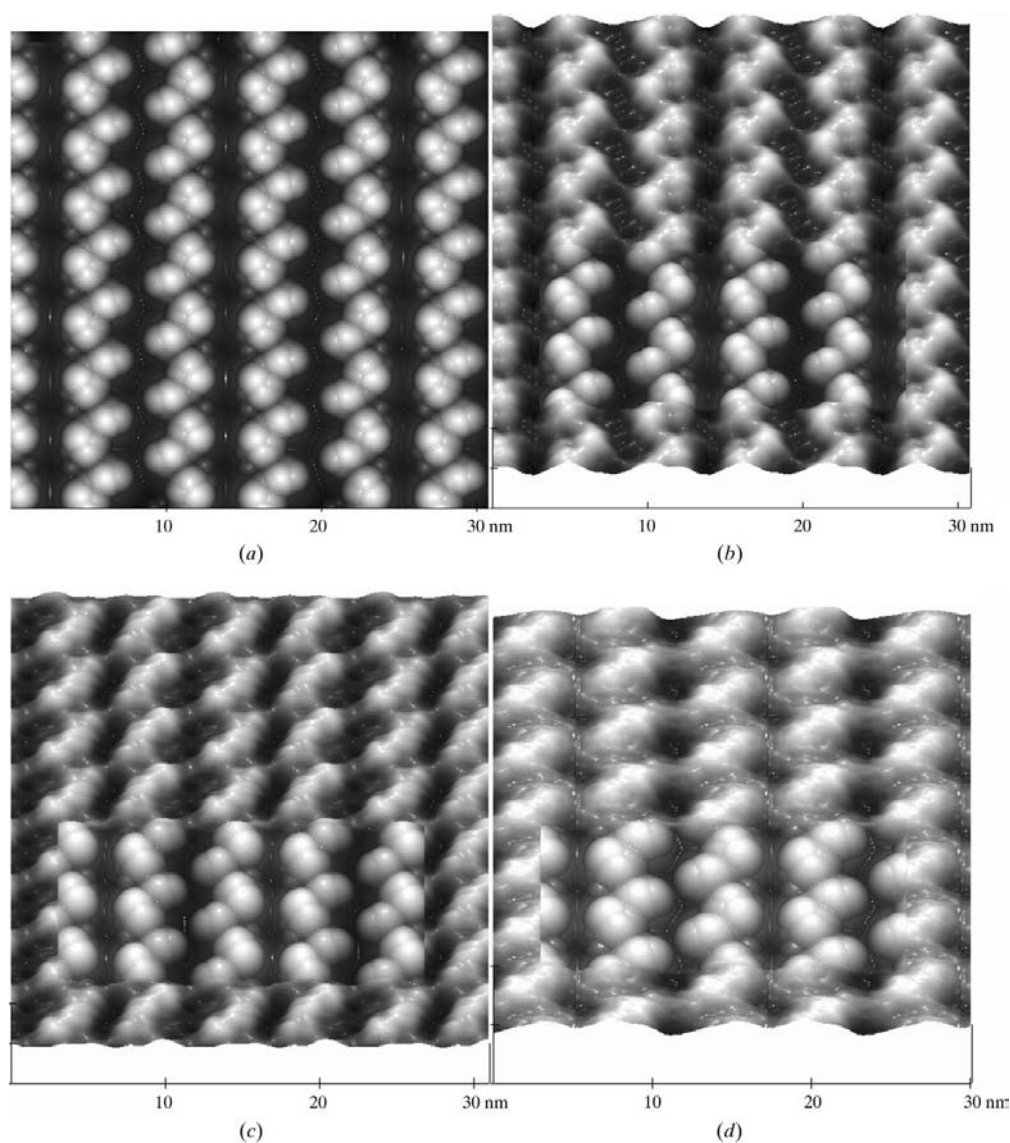

Figure 8
(*a*) Theoretical AFM image corresponding to Fig. 6(*a*), but with the molecules shifted by 7.2 Å. (*b*) Image in (*a*) embedded in the image in Fig. 3(*d*) for the image *A* in Table 1. (*c*) The theoretical and experimental images corresponding to (*b*) for the image *B* in Table 1. (*d*) The theoretical and experimental images corresponding to (*b*) for the image *C* in Table 1. The whole images in (*b*), (*c*) and (*d*) have been slightly tilted to show the surface relief in three dimensions.

the experimental image in Fig. 3(*d*). The correlation coefficient between the two images is now 0.98.

Although such precise determinations of tip shapes and scanned images may have other important applications, they are not necessary for this study, where the goal was to determine the correct molecular-packing arrangement from two possible ones. It is also possible that the deviation from perfect correlation between the images may arise from other factors and not exclusively from the lack of spherical symmetry in the tip. As shown above, assuming that the tip is perfectly spherical is quite adequate for distinguishing between the two packing arrangements. It produces a high correlation for the experimental image with the theoretical one corresponding to the correct packing (CC $\simeq$ 0.9) and little correlation with the image corresponding to the incorrect packing (CC $\simeq$ 0.25). However, the above exercise is useful in elucidating the origin of imperfections in the experimental images and pointing out the care which must be taken to avoid artifacts caused by the accumulation of debris on AFM tips.

5. Conclusions

The original predicted AFM image shown in Fig. 6(*a*) was calculated from crystallographic data, which reflects the molecular packing in the bulk of the crystal. In order to obtain better agreement with the experimentally measured packing on the crystal surface, the molecules had to be translated 7 Å from their crystallographic positions to pack more tightly about the 4_3 axes. This change may be further illustrated by comparing Fig. 11, in which this new packing arrangement is viewed along the axis, with Fig. 1(*b*), which displays the structure from the crystallographic data. This suggests that the molecules on the crystal surface are packed slightly differently from those in the bulk. As the crystal grows and the surface becomes covered by newer layers of molecules, the molecules on the original surface relax into the correct crystallographic arrangement.

If this rearrangement of molecules on the tetragonal lysozyme (110) face does occur, it would be the first time that a surface rearrangement has been observed on protein crystals, corresponding in some ways to the surface reconstruction widely observed in inorganic crystals. Reconstruction is the process by which molecules on the surface rearrange from the bulk crystalline order in order to minimize the surface energy of the system. However, the term is usually applied to structures which rearrange to larger repeating units than their bulk units. Such transformations are often quite complex, such as the (7 × 7) surface reconstruction observed on the (111) faces of silicon crystals (see, for example, the review by Kubby & Boland, 1996). The surface rearrangement suggested by this study is one which retains the same repeating unit, which would correspond to merely a (1 × 1) reconstruction. As such, it may be an indicator of
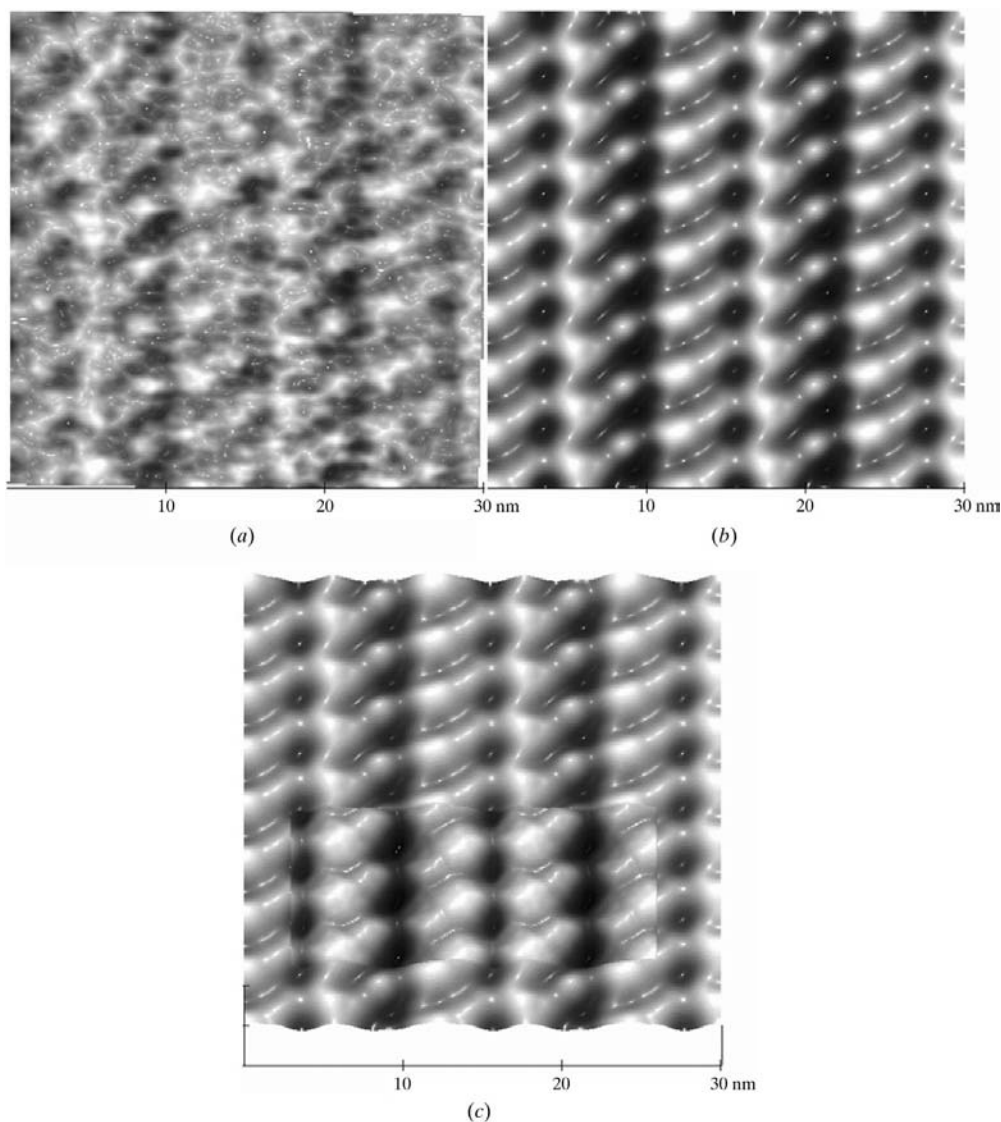

Figure 9
Series of images showing the effects of tip contamination. (*a*) The image *C** after the application of a median filter and Gaussian convolution to the raw data. (*b*) The image in (*a*) following translational averaging. (*c*) The theoretical image corresponding to Fig. 6(*a*), but for a non-symmetric tip, embedded in the experimental image in (*b*). The whole image in (*c*) has been slightly tilted to show the surface relief in three dimensions.

relevant solution-phase interactions between the molecules.

While there is little impossible contact between the molecules of the surface-rearrangement model, it is likely that some rotation of the molecules and rearrangement of side chains occurs as the molecules on the surface interact more
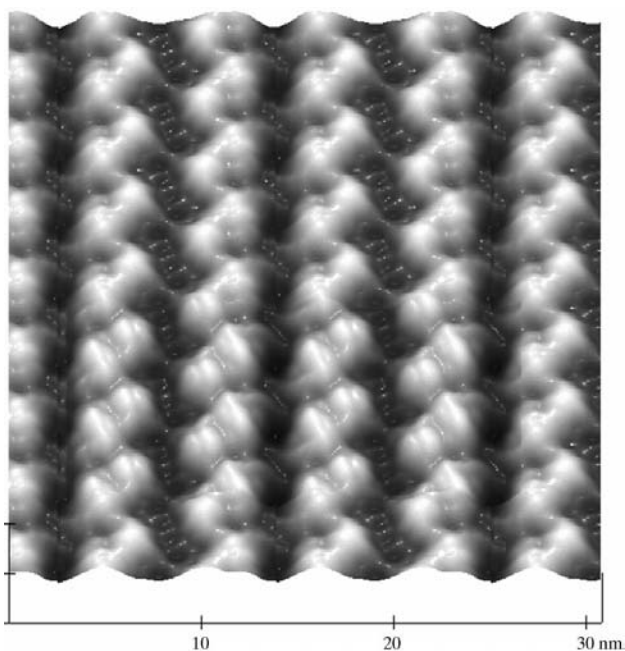

Figure 10
The theoretical and experimental images in Fig. 8(*a*), but with the theoretical image optimized with a nonsymmetric tip. The CC between the images is 0.98. The whole image has been slightly tilted to show the surface relief in three dimensions.
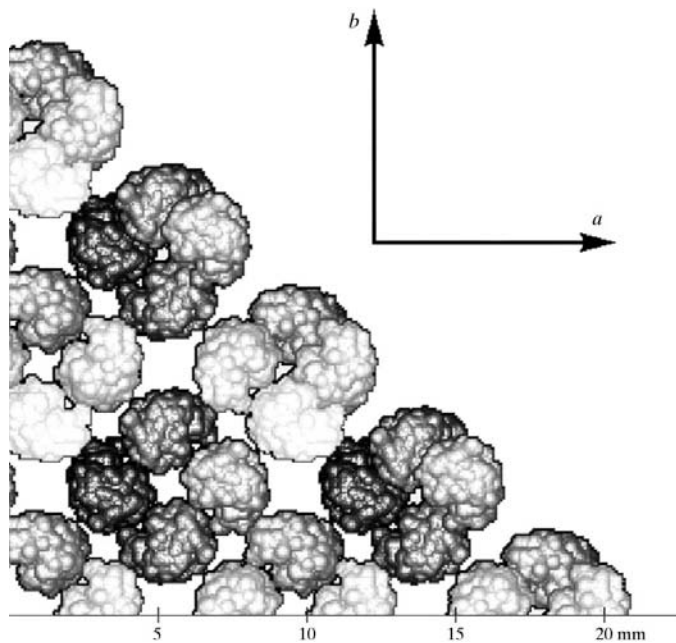

Figure 11
The molecular-packing arrangement of the (110) face similar to Fig. 1(*b*), but with the molecules in the surface bilayer shifted by 7.2 Å, resulting in closer packing about the 4_3 axes. Each molecule is drawn at 70% of its original size centered at its true position.

strongly about what will become the 4_3 helix within the crystal. More accurate determinations of the orientations and structure of the surface molecules require further analyses, which are ongoing. Similar differences between bulk and surface packing arrangements in other protein crystals have also been noticed in previous AFM studies (Baker *et al.*, 1997).

This study confirms the results of the earlier AFM experiment and the key prediction of the PBC analysis of the structure of tetragonal lysozyme crystals, namely that the molecular packing on the (110) faces of tetragonal lysozyme consists of what will become 4_3 helices when incorporated into the bulk of the crystal. The combination of PBC theory and the high-resolution AFM techniques developed here may provide a powerful method for probing the molecular-growth mechanisms of protein crystals. However, since the AFM data indicate that the surface structure differs significantly from the internal crystal structure with correspondingly different intermolecular contacts, caution should be applied when interpreting PBC analyses in the absence of AFM data. Additionally, it is important to note the limitations of the experimental part of this approach: the images produced are extremely sensitive to the shape of the tip and are easily distorted by tip anomalies, such as the accumulation of debris. This can result in images which lack the spatial periodicity and symmetries expected from the molecular packing. With care, these problems can be avoided, and it is not necessary to know the exact shape of the tip to determine the correct packing arrangement on a crystal face.

Besides its application to protein-crystal growth, the AFM technique developed here represents yet another application, in an ever-increasing list, for atomic force microscopy. Although it is limited to precisely ordered systems, such as crystalline materials, it provides a systematic method of determining the molecular-packing arrangements on the surfaces of these systems. It also affirms the importance of including the effect of tip shapes in high-resolution AFM studies.

References

Aune, K. C. & Tanford, C. (1969). *Biochemistry*, **8**, 4579–4590.
Baker, A. A., Helbert, W., Sugiyama, J. & Miles, M. J. (1997). *J. Struct. Biol.* **119**, 129–138.
Cacioppo, E. & Pusey, M. L. (1991). *J. Cryst. Growth*, **114**, 286–292.
Durbin, S. D. & Carlson, W. E. (1992). *J. Cryst. Growth*, **122**, 71–79.
Durbin, S. D., Carlson, W. E. & Saros, M. T. (1993). *J. Phys. D*, **26**, B128–B132.
Durbin, S. D. & Feher, G. (1990). *J. Mol. Biol.* **212**, 763–774.

Ewing, F. L., Forsythe, E. L., van der Woerd, M. & Pusey, M. L. (1996). *J. Cryst. Growth*, **160**, 389–397.

Hartman, P. (1987). *Morphology of Crystals*, edited by I. Sunagawa, pp. 269–319. Tokyo: Terra.

Hartman, P. & Perdok, W. G. (1955a). *Acta Cryst.* **8**, 49–52.

Hartman, P. & Perdok, W. G. (1955b). *Acta Cryst.* **8**, 521–524.

Hartman, P. & Perdok, W. G. (1955c). *Acta Cryst.* **8**, 525–529.

Kasas, S., Thomson, N. H., Smith, B. L., Hansma H. G., Zhu, X., Guthold, M., Bustamante, C., Kool, E. T., Kashlev, M. & Hansma, P. K. (1997). *Biochemistry*, **36**, 461–468.

Konnert, J. H., D'Antonio, P. & Ward, K. B. (1994). *Acta Cryst.* D**50**, 603–613.

Kubby, J. A. & Boland, J. J. (1996). *Surf. Sci. Rep.* **26**, 61–204.

Kuznetsov, Yu. G., Malkin, A. J., Land, T. A., DeYoreo, J. J., Barba, A. P., Konnert, J. H. & McPherson, A. (1997). *Biophys. J.* **72**, 2357–2364.

Land, T. A., Malkin, A. J., Kuznetsov, Yu. G., McPherson, A. & DeYoreo, J. J. (1995). *Phys. Rev. Lett.* **75**, 2774–2777.

Malkin, A. J., Kuznetsov, Yu. G., Glantz, W. & McPherson, A. (1996). *J. Phys. Chem.* **100**, 11736–11743.

Malkin, A. J., Kuznetsov, Yu. G., Land, T. A., DeYoreo, J. J. & McPherson, A. (1995). *Nature Struct. Biol.* **2**, 956–959.

Malkin, A. J., Kuznetsov, Yu. G. & McPherson, A. (1996a). *Proteins*, **24**, 247–252.

Malkin, A. J., Kuznetsov, Yu. G. & McPherson, A. (1996b). *J. Struct. Biol.* **117**, 124–137.

Malkin, A. J., Kuznetsov, Yu. G. & McPherson, A. (1997). *Surf. Sci.* **393**, 95–107.

Malkin, A. J., Land, T. A., Kuznetsov, Yu. G., McPherson, A. & DeYoreo, J. J. (1995). *Phys. Rev. Lett.* **75**, 2778–2781.

Markiewicz, P. & Goh, M. C. (1994). *Langmuir*, **10**, 5–7.

Nadarajah, A., Li, M. & Pusey, M. L. (1997). *Acta Cryst.* D**53**, 524–534.

Nadarajah, A. & Pusey, M. L. (1996). *Acta Cryst.* D**52**, 983–996.

Ng, J. D., Kuznetsov, Yu. G., Malkin, A. J., Keith, G., Giegé, R. & McPherson, A. (1997). *Nucleic Acids Res.* **25**, 2582–2588.

Salunke, D. M., Veerapandian, B., Kodandapani, R. & Vijayan, M. (1985). *Acta Cryst.* B**41**, 431–436.

Siedlecki, C. A., Lestini, B., Marchant-Kottke, K., Eppell, S. J., Wilson, D. L. & Marchant, R. E. (1996). *Blood*, **88**, 2939–2950.

Strom, C. S. & Bennema, P. (1997a). *J. Cryst. Growth*, **173**, 150–158.

Strom, C. S. & Bennema, P. (1997b). *J. Cryst. Growth*, **173**, 159–166.

Villarrubia, J. S. (1997). *J. Res. Natl Inst. Stand. Technol.* **102**, 425–454.

Williams, P. M., Shakesheff, K. M., Davies, M. C., Jackson, D. E., Roberts, C. J. & Tendler, S. J. B. (1996). *Langmuir*, **12**, 3468–3471.

Wilson, D. L., Krump, K. S., Eppell, S. J. & Marchant, R. E. (1995). *Langmuir*, **11**, 265–272.

Yip, C. M., Brader, M. L., DeFelippis, M. R. & Ward, M. D. (1998). *Biophys. J.* **75**, 1172–1179.

Yip, C. M., DeFelippis, M. R., Frank, B. H., Brader, M. L. & Ward, M. D. (1998). *Biophys. J.* **74**, 2199–2209.

Yip, C. M. & Ward, M. D. (1996). *Biophys. J.* **71**, 1071–1078.

Young, A. C., Tilton, R. F. & Dewan, J. C. (1994). *J. Mol. Biol.* **235**, 302–317.