homogeneity and uniqueness of the 'molecular portraits' provided by the quantitative analysis of gene expression patterns. We infer that these portraits faithfully represent the 'tumour' itself, and not merely the particular tumour 'sample', because we could recognize the distinctive expression pattern of a tumour in independent samples. The finding that a metastasis and primary tumour were as similar in their overall pattern of gene expression as were repeated samplings of the same primary tumour, suggests that the molecular program of a primary tumour may generally be retained in its metastases. Finally, we have explicitly discussed only a tiny fraction of the genes whose expression patterns varied among these tumours. Attention to the thousands of individual genes that define the molecular portraits of each tumour, and learning to interpret their patterns of variation, will undoubtedly lead to a deeper and more complete understanding of breast cancers. □

Methods

Most of the techniques used in this work have been described elsewhere[2,3,22,23], and detailed protocols are available at ⟨http://cmgm.Stanford.EDU/pbrown/⟩. The methods and protocols are also included in the Supplementary Information, and the primary data tables can be obtained at ⟨http://genome-www.stanford.edu/molecularportraits/⟩.

Received 7 February; accepted 25 May 2000.

1. Tavassoli, F. A. & Schnitt, S. J. *Pathology of the Breast* (Elsevier, New York, 1992).
2. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303,** 179–205 (1999).
3. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24,** 227–235 (2000).
4. Aas, T. *et al.* Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nature Med.* **2,** 811–814 (1996).
5. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
6. Perou, C. M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96,** 9212–9217 (1999).
7. Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O. & Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.* **27,** 1517–1523 (1999).
8. Hoch, R. V., Thompson, D. A., Baker, R. J. & Weigel, R. J. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer* **84,** 122–128 (1999).
9. Pauletti, G., Godolphin, W., Press, M. F. & Slamon, D. J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* **13,** 63–72 (1996).
10. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23,** 41–46 (1999).
11. Ronnov-Jessen, L., Petersen, O. W. & Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol. Rev.* **76,** 69–125 (1996).
12. Taylor-Papadimitriou, J. *et al.* Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. *J. Cell Sci.* **94,** 403–413 (1989).
13. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537 (1999).
14. Dairkee, S. H., Mayall, B. H., Smith, H. S. & Hackett, A. J. Monoclonal marker that predicts early recurrence of breast cancer. *Lancet* **1,** 514 (1987).
15. Dairkee, S. H., Puett, L. & Hackett, A. J. Expression of basal and luminal epithelium-specific keratins in normal, benign, and malignant breast tissue. *J. Natl Cancer Inst.* **80,** 691–695 (1988).
16. Malzahn, K., Mitze, M., Thoenes, M. & Moll, R. Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. *Virchows Arch.* **433,** 119–129 (1998).
17. Guelstein, V. I. *et al.* Monoclonal antibody mapping of keratins 8 and 17 and of vimentin in normal human mammary gland, benign tumors, dysplasias and breast cancer. *Int. J. Cancer* **42,** 147–153 (1988).
18. Gusterson, B. A. *et al.* Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. *Cancer Res.* **42,** 4763–4770 (1982).
19. Nagle, R. B. *et al.* Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J. Histochem. Cytochem.* **34,** 869–881 (1986).
20. Berns, E. M. *et al.* Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumors: correlation with steroid receptors. *Eur. J. Cancer* **28,** 697–700 (1992).
21. Heintz, N. H., Leslie, K. O., Rogers, L. A. & Howard, P. L. Amplification of the c-erb B-2 oncogene and prognosis of breast adenocarcinoma. *Arch. Pathol. Lab. Med.* **114,** 160–163 (1990).
22. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686 (1997).
23. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511 (2000).

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

Acknowledgements

We thank W. Gerald and L. Norton for the three New York tumour specimens; M. Stampfer and P. Yaswen for the 184 sample mRNAs; and members of the P. O. Brown, D. Botstein and A.-L. Børresen-Dale labs for discussions. We are grateful to the NCI and the Howard Hughes Medical Institute who provided support for this research. C.M.P. is a SmithKline Beecham Pharmaceuticals Fellow of the Life Sciences Research Foundation. T.S. is a research fellow of the Norwegian Cancer Society. M.B.E. is an Alfred P. Sloan Foundation Postdoctoral Fellow in Computational Molecular Biology. D.T.R. is a Walter and Idun Berry Fellow. P.O.B. is an Associate Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.B. (e-mail: botstein@genome.stanford.edu) or P.O.B. (e-mail: pbrown@cmgm.stanford.edu).

Potential for biomolecular imaging with femtosecond X-ray pulses

Richard Neutze*, Remco Wouts*, David van der Spoel*, Edgar Weckert†‡ & Janos Hajdu*

** Department of Biochemistry, Biomedical Centre, Box 576, Uppsala University, S-75123 Uppsala, Sweden*
† Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-76128 , Germany

Sample damage by X-rays and other radiation limits the resolution of structural studies on non-repetitive and non-reproducible structures such as individual biomolecules or cells[1]. Cooling can slow sample deterioration, but cannot eliminate damage-induced sample movement during the time needed for conventional measurements[1,2]. Analyses of the dynamics of damage formation[3–5] suggest that the conventional damage barrier (about 200 X-ray photons per Å^2 with X-rays of 12 keV energy or 1 Å wavelength[2]) may be extended at very high dose rates and very short exposure times. Here we have used computer simulations to investigate the structural information that can be recovered from the scattering of intense femtosecond X-ray pulses by single protein molecules and small assemblies. Estimations of radiation damage as a function of photon energy, pulse length, integrated pulse intensity and sample size show that experiments using very high X-ray dose rates and ultrashort exposures may provide useful structural information before radiation damage destroys the sample. We predict that such ultrashort, high-intensity X-ray pulses from free-electron lasers[6,7] that are currently under development, in combination with container-free sample handling methods based on spraying techniques, will provide a new approach to structural determinations with X-rays.

Radiation damage is caused by X-ray photons depositing energy directly into the sample. At 1 Å wavelength, the photoelectric cross-section of carbon is about 10 times higher than its elastic-scattering cross-section, making the photoelectric effect the primary source of damage. The photoelectric effect is a resonance phenomenon in which a photon is absorbed and an electron ejected[8], usually from a low-lying orbital of the atom (about 95% of the photoelectric events remove K-shell electrons from carbon, nitrogen, oxygen and sulphur), producing a hollow ion with an unstable electronic configuration. Relaxation is achieved through an electron from a higher shell falling into the vacant orbital. In heavy elements this usually gives rise to X-ray fluorescence, whereas in light elements the falling electron is more likely to give up its energy to another electron, which is then ejected in the Auger effect. Auger emission is predominant in carbon, nitrogen, oxygen and sulphur ($>95\%$)[9]; thus, most photoelectric events ultimately remove two electrons from these elements. These two electrons have different energies ($\sim$12 keV for photoelectrons and $\sim$0.25 keV for Auger electrons),

‡ Present address: HASYLAB at DESY, Notkestr. 85, D-22607 Hamburg, Germany.

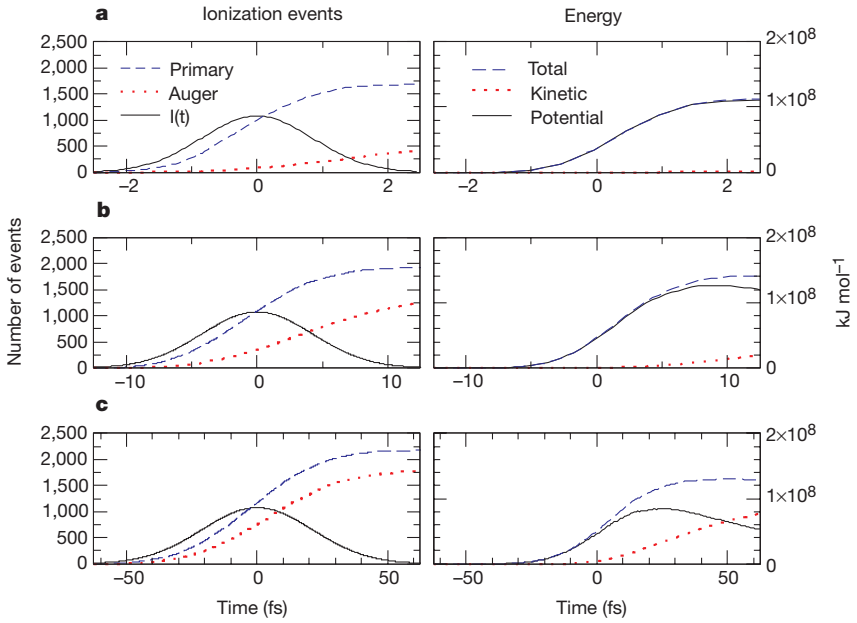

Figure 1 Ionization of a lysozyme molecule in intense X-ray pulses. The full width at half maximum (FWHM) of the pulse was 2 fs (**a**), 10 fs (**b**) or 50 fs (**c**). The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^{6} photons per Å^2) in all cases. The creation of a large number of positive charges owing to primary ionization by X-rays (photoelectric effect and Compton scattering), and the subsequent Auger emissions (left) result in a rise in the potential (mainly electrostatic) energy of the sample (right). The degree of conversion of potential energy into kinetic energy during the X-ray exposure is inertia limited, and so depends strongly on the duration of the pulse. The gaussian shape of the X-ray pulse is indicated.

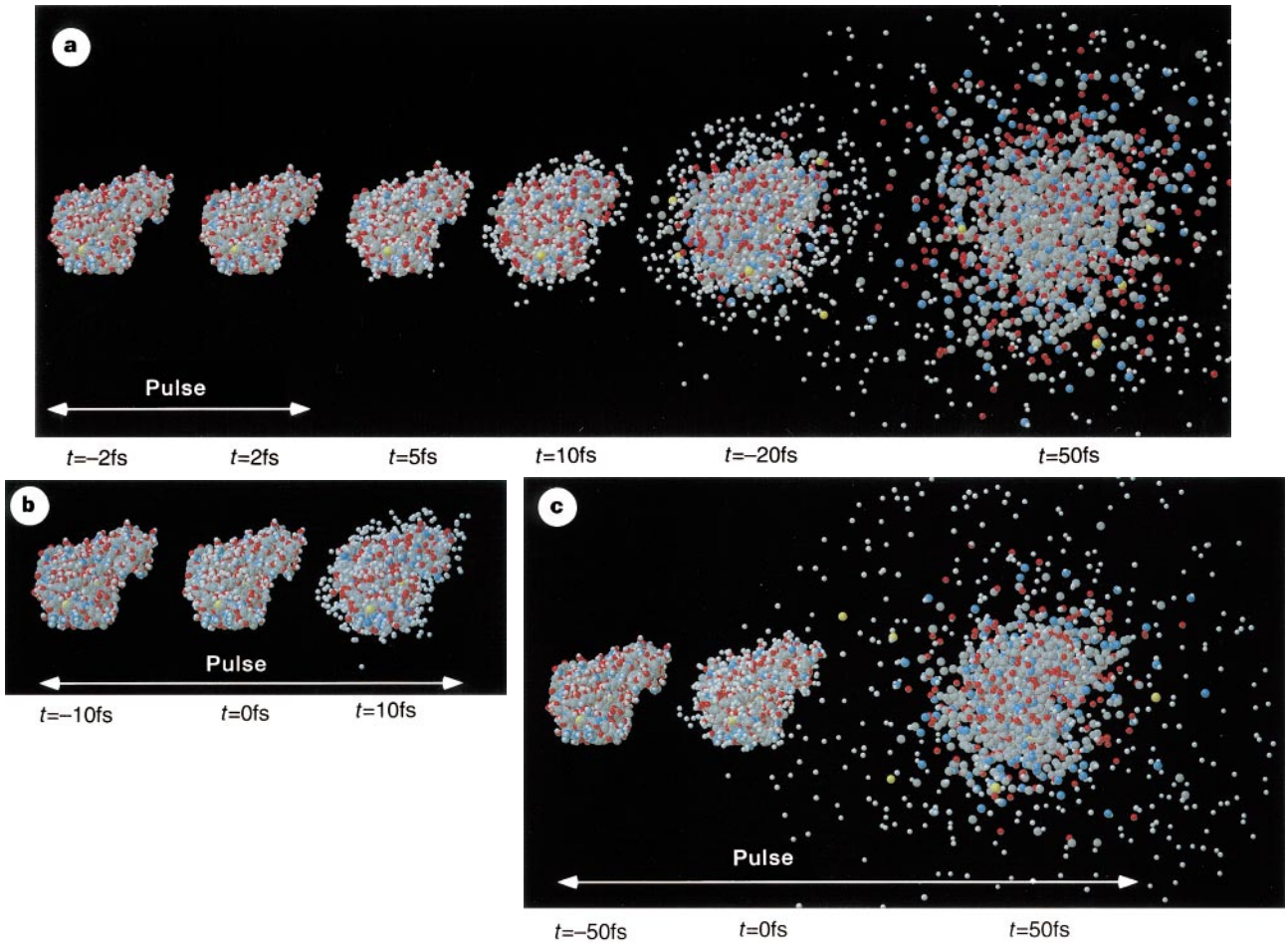

Figure 2 Explosion of T4 lysozyme (white, H; grey, C; blue, N; red, O; yellow, S) induced by radiation damage. The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^{6} photons per Å^2) in all cases. **a**, A protein exposed to an X-ray pulse with an FWHM of 2 fs, and disintegration followed in time. Atomic positions in the first two structures (before and after the pulse) are practically identical at this pulse length because of an inertial delay in the explosion. R_{nucl} = 3%, R_{elec} = 11% **b**, Lysozyme exposed to the same number of photons as in **a**, but the FWHM of the pulse was 10 fs. Images show the structure at the beginning, in the middle and near the end of the X-ray pulse. R_{nucl} = 7%, R_{elec} = 12% **c**, Behaviour of the protein during an X-ray pulse with an FWHM of 50 fs. R_{nucl} = 26%, R_{elec} = 30%.

and are released at different times. Relevant K-hole lifetimes determined from Auger line-widths[9] are 11.1 fs (C), 9.3 fs (N), 6.6 fs (O) and 1.3 fs (S). Shake-up excitations and interference between decay channels will modulate this picture.

Another effect is Compton (or inelastic) scattering, which represents a direct momentum transfer from an X-ray photon to a bound electron, so that the X-ray photon is scattered with a reduced energy. If the recoil energy taken up by the electron is greater than its shell binding energy, the atom will be ionized. The inelastic cross-section of carbon, nitrogen and oxygen is about 3% of the corresponding photoelectric cross-sections[10,11], whereas the inelastic cross-section of hydrogen is much higher than its photoelectric cross-section.

The average velocities of photoelectrons (43 nm fs^{-1}) and Auger electrons (7 nm fs^{-1}) enable these electrons to escape the protein environment in less than a femtosecond during early phases of an exposure. At these velocities the inelastic electron scattering cross-sections for carbon[1] lie between 0.1 and 2.0 Å^2, such that roughly one electron in five will deposit additional energy into the molecule from which it escapes, and may also remove outer-shell electrons[12]. In late phases of an exposure, a significant fraction of the emitted

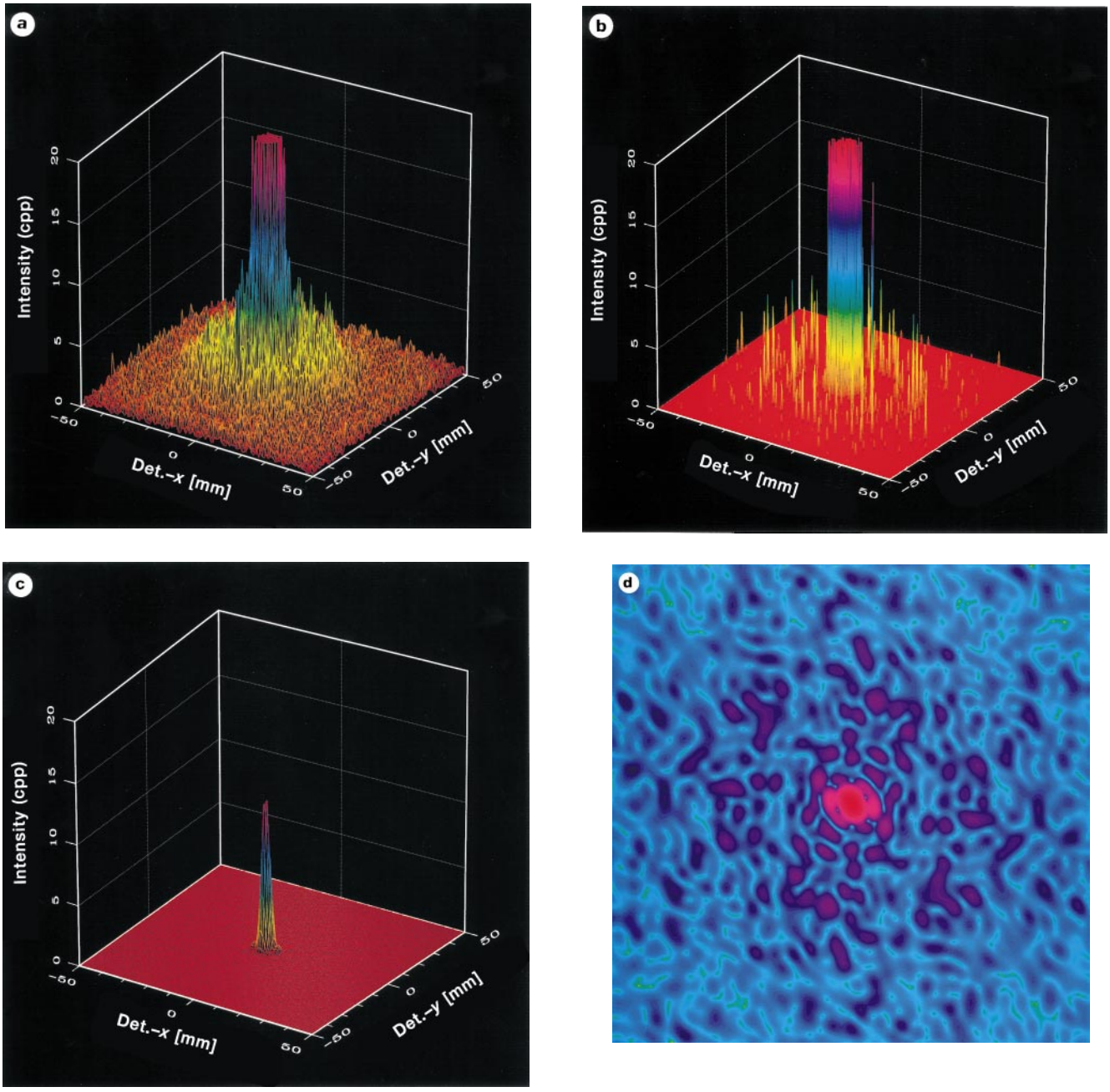

Figure 3 Elastic scattering from a variety of samples. **a**–**c**, Simulated diffraction images on a 128 × 128 pixel planar detector (100 mm × 100 mm) normal to and centred at the beam, and placed 100 mm from the sample. The background was not modelled, and 100% detective quantum efficiency was assumed. The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^6 photons per Å^2); the pulse length was 10 fs. The resolution is 2.2 Å at the rim in **a**–**c**. cpp, counts per pixel. **a**, Scattering from a single tomato bushy stunt virus capsid. **b**, Scattering from a 5 × 5 × 5 cluster of lysozyme molecules with an average r.m.s. conformational deviation of 0.2 Å to model an imperfect lattice. **c**, Scattering from a single molecule of lysozyme. **d**, A planar section through the molecular transform (that is, a simulated continuous scattering image) of a single T4 lysozyme molecule under ideal conditions without sample movement or damage. Resolution at the rim of **d** corresponds to 2.0 Å. Structure factor amplitudes are coloured logarithmically (magenta, high; green, low). The section is perpendicular to the *z* axis, and crosses through the origin at the centre of the image, revealing centric symmetry.

electrons will not be able to escape the increased positive potential of the sample. Trapped electrons will increase the kinetic energy of the sample through thermal equilibration, but they will also slow down the Coulomb explosion of the sample by partially neutralizing the positively charged protein core. These opposing effects have not been considered here, but are factors that are likely to influence the dynamics of larger systems.

In addition to the effects listed above, it has been suggested that extreme X-ray intensities could strip all outer-shell electrons from their parent atoms on a timescale of attoseconds[5]; however, a quantum mechanical analysis of the electric-field-induced tunnelling shows that atoms actually become stabilized against ionization by this mechanism at high photon frequencies[13] like those of X-rays.

We constructed a model in which X-ray-induced damage is described stochastically on the basis of the probability of a photoelectric or an inelastic event. The instantaneous probability of ionization of atom j at time t was calculated as the product of its photoelectric and inelastic cross-section[10,11], and the X-ray intensity $I(t)$. We modelled Auger emission as a stochastic exponential decay to reproduce appropriate K-hole lifetimes. The direction of photoemission was distributed according to a random deviate that followed a gaussian distribution[8]. A recoil velocity for the ionized atom owing to inelastic scattering or the emission of a photo- or Auger electron was determined from energy and momentum conservation. For inelastically scattered photons, the angle of deflection was determined by a random deviate following a Rayleigh distribution[8]. For each inelastic scattering the electron's recoil energy was calculated, and where this was greater than the binding energy of the electron an ionization event was modelled. An inventory was kept of what electrons remained on what atoms, and changes in the photoelectric, elastic and Compton scattering cross-sections of all atoms were computed and updated during exposures.

Energies and ionization events for a set of representative simulations on T4 lysozyme[14] and its 118 crystallographically determined water molecules are shown in Fig. 1. The protein with its bound solvent molecules was considered to be in the gas phase, under conditions similar to a non-destructive electrospray mass spectrometry experiment (see below). For each simulation, a total flux of 3×10^{12} (12 keV) X-ray photons passed through the 100-nm diameter focal spot, corresponding to 3.8×10^{6} photons per Å^2, and causing roughly 2,000 primary ionization events, or more than one ionization event for each non-hydrogen atom in the sample. For this set, the full-width at half-maximum (FWHM) for each pulse was 2 fs (Fig. 1a), 10 fs (Fig. 1b) and 50 fs (Fig. 1c). For an FWHM of 2 fs, only a quarter of the K-holes created by inner-shell ionization events had time to decay by Auger electron emission; therefore, the total number of positive charges on the sample at the completion of the 2-fs pulse was only 60% of the total number of charges at the completion of the 50-fs pulse. Furthermore, during short intense pulses numerous K-holes may be present at any one time, reducing the photoelectric cross-sections of atoms in which they were produced and thus lowering the total number of primary ionization events in the sample (see trend in Fig. 1). This effect makes the system radiation hardened to photo-ionization during very short exposures, and is more pronounced at higher radiation intensities (not shown).

The creation of a large number of positive charges in close proximity results in a rise in the electrostatic energy of the sample (Fig. 1), which drives its eventual explosion (Fig. 2). Each sample will survive only a single shot. The degree of conversion of potential energy into kinetic energy during the X-ray exposure is inertia limited, and so depends strongly on the duration of the pulse. During the 2-fs pulse there was insufficient time for the kinetic energy to grow appreciably (Fig. 1a). In contrast, by the completion of the 50-fs pulse the kinetic energy of the sample had surpassed its potential energy (Fig. 1c), indicating that the explosion of the sample was well under way. The destruction of the sample by the X-ray pulse is illustrated by snapshots from the trajectories for an FWHM of 2, 10 or 50 fs (Fig. 2a–c). For the two shorter X-ray pulse widths, only very small changes in the atomic positions have had time to develop.

Hydrogen ions and highly ionized sulphurs are the first to escape the immediate vicinity of the protein (at 12 keV the photoelectric cross-section for sulphur is about 50 times larger than that for carbon). In the 50-fs FWHM pulse, the molecule is destroyed before the pulse is over. The 2-fs FWHM simulation was continued to 50 fs beyond the pulse, and in this case the molecule explodes as well, but only after the structural information has been gathered. These results are in agreement with other observations. Molecular dynamics simulations of the response of small van der Waals clusters of atoms ionized by intense visible femtosecond laser pulses[12] also showed a delay of a few tens of femtoseconds before the creation of significant structural disorder, and when a femtosecond X-ray pulse generated by a laser plasma was used to probe the dynamics of a rapidly heated organic sample[15] a delay of about 100 fs was observed before significant growth in disorder.

We calculated scattering intensities for a range of samples

Table 1 Calculated limits of resolution

Pulse duration*	1 fs	5 fs	10 fs	50 fs	100 fs	
Photons per pulse in a 100-nm diameter spot*	5×10^{13}	1×10^{13}	5×10^{12}	8×10^{11}	3×10^{11}	
Relative scattering power†	0.32	0.53	0.71	0.96	0.97	
Single lysozyme molecule	15 Å	24 Å	26 Å	30 Å	> 30 Å	1 photon per pixel‡§
	> 30 Å	> 30 Å	> 30 Å	>30 Å	>30 Å	9 photons per pixel‡‖
	682	223	150	33	12	Σphotons (2–30 Å)¶
2×2×2 cluster of lysozymes	2.5 Å	3.1 Å	4.8 Å	12 Å	17 Å	1 photon per pixel
	6.5 Å	12 Å	16 Å	30 Å	> 30 Å	9 photons per pixel
	5,795	1,901	1,277	277	105	Σphotons (2–30 Å)
3×3×3 cluster of lysozymes	< 2.0 Å	< 2.0 Å	< 2.0 Å	3.0 Å	6.5 Å	1 photon per pixel
	2.2 Å	3.0 Å	3.0 Å	12 Å	17 Å	9 photons per pixel
	17,254	5,661	3,803	824	313	Σphotons (2–30 Å)
5×5×5 cluster of lysozymes	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	1 photon per pixel
	< 2.0 Å	< 2.0 Å	< 2.0 Å	2.9 Å	3.9 Å	9 photons per pixel
	76,346	25,050	16,828	2,746	1,387	Σphotons (2–30 Å)
Single viral capsid (TBSV)	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	1 photon per pixel
	< 2.0 Å	2.5 Å	2.5 Å	4.7 Å	22 Å	9 photons per pixel
	160,640	52,708	35,969	7,795	2,964	Σphotons (2–30 Å)

* Pairs of pulse lengths and integrated photon intensities were selected from the line of 'maximum tolerable damage' in Fig. 4b (R_{elec} = 15%).
† Integrated scattering power of the sample over the exposure relative to the integrated scattering power of an undamaged, idealized sample.
‡ Two resolution limits are listed, corresponding to peak heights with 1 or 9 elastically scattered photons per pixel. Under ideal conditions, these values would correspond to intensity measurements with $I = 1\sigma(I)$ and $I = 3\sigma(I)$, respectively, where $\sigma(I)$ is the standard deviation of the intensity measurement. The background was not modelled, and 100% detective quantum efficiency was assumed. Wavelength = 1.0 Å (12 keV energy), planar detector (100 mm × 100 mm) normal to and centred at the beam, sample to detector distance = 100 mm, pixel size = 0.8 mm × 0.8 mm.
§ Resolution may be extended beyond this limit by averaging, if orientation can be determined accurately.
‖ Information extending to this limit could be used for determining sample orientation.
¶ Number of elastically scattered photons on the detector between 2.0 and 30 Å resolution. An even larger number of photons will be scattered between 30 Å resolution and infinity (Fig. 3), but these are not listed.

(Fig. 3a–c), using pulse parameters from Fig. 1b. A single capsid of the tomato bushy stunt virus (TBSV)[16] would scatter X-rays to atomic resolution with a 10-fs pulse containing 3×10^{12} (12 keV) photons in a 100-nm diameter spot (Fig. 3a). The capsid particle of TBSV is a $T=3$ icosahedral nanocluster of 60×3 subunits. It is not a crystalline structure, as it lacks translational symmetry. For a tiny crystalline structure (Fig. 3b), high-resolution scattering may be obtained from a nanocrystal of $5\times5\times5$ lysozyme molecules arranged on a rectangular primitive lattice. Molecules in the nanocrystal and subunits in TBSV were given an average r.m.s. conformational deviation of 0.2 Å to simulate imperfect conditions. Although these structures diffract well with these pulse parameters, Fig. 3c shows that the same intensity would be scattered to only about 40 Å resolution (Fig. 3c) from a single lysozyme molecule. For comparison, Fig. 3d shows a planar section through the continuous molecular transform of lysozyme, shown here under ideal conditions without sample movement or damage. An analysis of maximum attainable resolutions in single exposures is given below (Table 1, Fig. 4).

Larger samples scatter to higher resolutions even without internal symmetry, and resolution for reproducible samples can be extended further by numerical alignment and averaging procedures[17–19]. Such procedures are well established in electron cryomicroscopy. Averaging techniques exploit the fact that photon counts from the structural signal grow more rapidly than counts from an incoherent

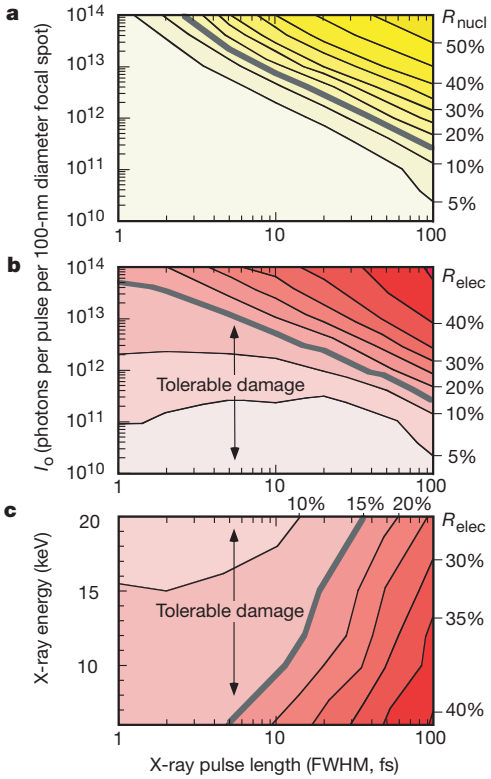

Figure 4 The landscape of damage tolerance. Contour plots of the nuclear (R_{nucl}) and electronic (R_{elec}) weighted-average R factors (equation (2)) as functions of the X-ray flux, pulse duration and photon energy made by interpolating the R value from the individual simulations. **a**, R_{nucl} for an X-ray beam of 12 keV, illustrating the extent to which the average positions of the sample's nuclei are perturbed by radiation damage. **b**, R_{elec} for a X-ray beam of 12 keV, illustrating the extent to which the information content of the elastically scattered X-rays is degraded by radiation damage. We regard damage as acceptable if $R_{elec}\leq15\%$ which is indicated by the grey line. **c**, The variation of R_{elec} as the X-ray photon energy is changed, but with the total number of elastic-scattering events per carbon atom held constant. These intensities (I_{tot}) were 1.33×10^{12} (6 keV), 1.85×10^{12} (8 keV), 2.36×10^{12} (10 keV), 3.0×10^{12} (12 keV), 3.96×10^{12} (15 keV) and 6.0×10^{12} (20 keV) photons per 100-nm diameter spot.

background (sources of noise here include inelastic scattering, Bremsstrahlung from photoelectrons colliding with instrument walls, and an imperfect sample environment). Multi-image alignment and averaging methods developed in electron microscopy produced substantially increased resolutions, for example, for the ribosome[19]. With particles displaying high symmetry, resolution can be extended further by exploiting the symmetry of the structure[17,20]; however, averaging procedures can only be applied when a reproducible sample scatters a sufficiently large number of photons for its orientation to be determined.

Phases for scattering images can be recovered in a number of ways, including the oversampling of continuous molecular transforms[21], holographic imaging methods[22,23], holographic data evaluation methods[22], classical methods of crystallography and techniques for phase extension from lower-resolution electron cryomicroscopy images.

The structural information, which is theoretically recoverable from the sample during an exposure, can be quantified by the introduction of a weighted-average agreement factor (R factor, see equation (2) in Methods), which provides a direct assessment of the data quality. The analysis is based on differences between scattering from a sample that suffers radiation damage and scattering from a hypothetical sample that suffers no radiation damage during X-ray exposure. Radiation damage interferes with the atomic scattering factors and positions. If the atomic scattering factors for the damaged sample are taken as unchanged from those of the parent atoms, then R provides information on the extent to which the average positions of the sample's nuclei are perturbed by X-ray-induced damage ($R\equiv R_{nucl}$). If the degree of ionization of each atom is included when modelling the atomic (or ionic) scattering factors, then R provides information on the extent to which the elastically scattered radiation is perturbed by X-ray-induced damage ($R\equiv R_{elec}$). R_{nucl} depends only on the movements of atoms during an exposure, whereas R_{elec} depends on both the movements and the changes in the electronic structure of atoms. Whereas R_{nucl} delivers useful insight into the explosion process, R_{elec} corresponds directly to the quality of the data that would be obtained in an actual experiment. Macromolecular crystal structures in the Protein Data Bank have crystallographic R factors of about 20%. Many of the structures, especially those collected earlier on photographic film, represent data sets with merging R factors in the 5% to 15% range. Taking the latter value as an arbitrary upper limit, we regard damage as acceptable if $R_{elec}\leq15\%$. A survey of the landscape for damage tolerance (Fig. 4) shows many combinations of wavelength, integrated intensity and pulse length where $R_{elec}\leq15\%$. Figure 4a and b shows contour plots of weighted-average R_{nucl} and R_{elec} values for 12 keV photons as functions of pulse duration and the total photon flux; R_{nucl} and R_{elec} display different response dynamics, as atomic positions do not change as fast as the electronic configuration of the atoms. Table 1 shows the calculated limits of resolution at five different points along the line of the 'maximum tolerable damage' in Fig. 4b. The total dose delivered to the sample at either end of this line represents an increase of several orders of magnitude above the previously postulated limit of about 200 photons per Å^2 for conventional experiments at low X-ray intensities and cryogenic temperatures[2]. Increasing I_{tot} above 5×10^{13} photons per spot will probably be of little value even at very short pulse lengths, as above this dose most of the sample's electrons are stripped from the atoms during the X-ray exposure (see values for relative scattering powers in Table 1). Thus, the total number of elastically scattered photons no longer increases linearly with increasing dose, and X-rays scattered from unbound electrons will contribute additional noise. With increasing X-ray energy, the ratio of elastic-scattering events to damaging events becomes more favourable[10,11], and data quality improves as the X-ray probe moves to higher photon energies (Fig. 4c).

With ultrasmall samples, standard procedures for sample selec-

tion and handling will no longer be applicable. New container-free methods are being developed[24] for characterizing, selecting and injecting single molecules, particles or nanoclusters into intense X-ray pulses. These methods are based on techniques used in electrospray ionization mass spectrometers and cell sorters, and allow the injection of individual hydrated protein molecules, virus particles[25,26] and intact ribosomes[26] into the vacuum chamber of the mass spectrometer while maintaining the structural integrity of the samples at cryogenic temperatures. The transit time of the sample is a few microseconds in the instrument. The sample entering the vacuum chamber is in random orientation, encapsulated in a micro-droplet. Such spraying methods could be used to characterize and inject well defined sample particles into intense X-ray pulses. Assembling protein molecules into nanoclusters (M. Svenda and J.H., manuscript in preparation) increases the intensity of scattered radiation from otherwise small proteins (Fig. 3). Such nanoclusters could be created by the specific attachment of target proteins on the surface of reproducible scaffolds, such as icosahedral viruses[27].

Should this new femtosecond window in imaging provide a path to high-resolution structural information without the need for macroscopic crystals, its impact on structural biology would be tremendous. □

Methods

Modelling

We created a program (XMD) that extends the GROMACS molecular dynamics package[28] to simulate electronic and structural changes triggered by X-rays in the sample. The GROMOS[29] force field was modified to incorporate Morse potentials for the description of all chemical bonds, thereby enabling bonds with sufficiently high energy to break. For water, the simple point charge model[30] was used and adapted in the same manner. Elastic, inelastic and photoelectric cross-sections of atoms were incorporated, and changes in the cross-sections during the X-ray exposure were modelled, using theoretical values[10,11]. We did not model femtosecond collisional electron transfer and slower radical reactions. Modification of the dissociation energy of two atoms owing to electron emission was not taken into account.

We used T4 lysozyme[14] as a model, including its 118 crystallographically determined water molecules. Hydrogens were added to polar and aromatic groups and to water oxygen atoms; the initial charge on the protein was +8. After several rounds of energy minimization the r.m.s. deviation from the crystal structure of the protein was 0.21 Å for all non-hydrogen protein atoms. This structure was used as the starting point for subsequent molecular dynamics simulations. Simulations were performed with a time step of 50 attoseconds, taking all non-bonded interactions explicitly into account.

Calculations

The X-ray pulse with intensity $I(t)$ was taken to have a gaussian temporal profile with pulse duration (FWHM) of 1, 2, 5, 10, 20, 50 or 100 fs. The total integrated X-ray flux, I_{tot}, was expressed as the number of 12-keV X-ray photons passing through a 100-nm diameter circular focal spot, and simulations were run with intensities I_{tot} of 10^{10}, 3×10^{10}, 10^{11}, 3×10^{11}, 10^{12}, 3×10^{12}, 10^{13}, 3×10^{13} and 10^{14} photons for each of the pulse widths. A simulation with $I_{tot} = 0$ was the reference simulation for calculating elastic-scattering properties. The maximum intensity of the X-ray pulse was set at time $t = 0$, and the simulations were run from $t = -1.2$ FWHM to $t = +1.2$ FWHM (covering 99% of the integrated pulse). All simulations were repeated with different random number seeds for the stochastic events to test reproducibility. The results deviated by less than 5% from each other. We carried out further sets of simulations to estimate wavelength-dependent effects (see Fig. 4c, legend).

For unpolarized X-rays, the mean number of elastically scattered photons $I(\mathbf{u},\Omega)$ to be detected by an idealized detector pixel of projected solid angle Ω centred at a positional vector $\mathbf{u}$ is

$$I(\mathbf{u},\Omega) = 1/2(1+\cos^2 2\theta)\Omega r_e^{\,2}\int_{-\infty}^{\infty} \mathrm{I}(t)\left|\sum_j f_j(t)\exp\left\{i\Delta\mathbf{k}(\mathbf{u})\cdot\mathbf{x}_j(t)\right\}\right|^2 \mathrm{d}t \qquad (1)$$

where r_e is the classical electron radius (2.81785×10^{-5} Å); $\mathrm{I}(t)$ is the intensity of the X-ray pulse; $f_j(t)$ is the atomic scattering factor for the jth atom as a function of time; $\mathbf{x}_j(t)$ is the position of this atom as a function of time; and $\Delta\mathbf{k}$ is the change in the wave vector of the X-ray photon when scattered through 2θ radians towards the pixel centred at $\mathbf{u}$. Radiation damage interferes with $f_j(t)$ and $\mathbf{x}_j(t)$. A direct assessment of imaging quality is given by the R-factor, defined as:

$$R \equiv \sum_{\mathbf{u}}\left|\frac{K^{-1}\sqrt{I_{real}(\mathbf{u},\Omega)}-\sqrt{I_{ideal}(\mathbf{u},\Omega)}}{\sum_{\mathbf{u}}\sqrt{I_{ideal}(\mathbf{u},\Omega)}}\right| \qquad (2)$$

where

$$K = \frac{\sum_{\mathbf{u}}\sqrt{I_{real}(\mathbf{u},\Omega)}}{\sum_{\mathbf{u}}\sqrt{I_{ideal}(\mathbf{u},\Omega)}} \qquad (3)$$

Scaling factor K describes the relative scattering power of the sample (Table 1). Values for I_{real} and I_{ideal} were computed from molecular dynamics simulations by evaluating equation (1) from 60 snapshots of each MD trajectory. I_{real} was derived from the time-dependent atomic coordinates $\mathbf{x}_j(t)$ and scattering factors $f_j(t)$ of a sample exploding in the X-ray pulse, while I_{ideal} was determined from the reference molecular dynamics simulation of an unexposed sample.

Received 23 December 1999; accepted 14 June 2000.

1. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28,** 171–193 (1995).
2. Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and X-ray diffraction. *Proc. R. Soc. Lond. B* **241,** 6–8 (1990).
3. Solem, J. C. Imaging biological specimens with high-intensity soft X-rays. *J. Opt. Soc. Am. B* **3,** 1551–1565 (1986).
4. Hajdu, J. in *Frontiers in Drug Research* (eds Jensen, B., Jorgensen, F. S. & Kofod, H.) 375–395 (Alfred Benzon Symposium 28, Munksgaard, Copenhagen, 1990).
5. Doniach, S. Studies of structure of matter with photons from an X-ray free electron laser. *J. Synchr. Rad.* **3,** 260–267 (1996).
6. Winick, H. The linac coherent light source (LCLS): A fourth-generation light source using the SLAC linac. *J. Elec. Spec. Rel. Phenom.* **75,** 1–8 (1995).
7. Wiik, B. H. The TESLA project: an accelerator facility for basic science. *Nucl. Inst. Meth. Phys. Res. B* **398,** 1–8 (1997).
8. Dyson, N. A. *X-Rays in Atomic and Nuclear Physics* (Longman, London, 1973).
9. Krause, M. O. & Oliver, J. H. Natural width of atomic K and L level Kα X-ray lines and several KLL Auger lines. *J. Phys. Chem. Ref. Data* **8,** 329–338 (1979).
10. Veigele, W. J. Photon cross sections from 0.1 keV to 1 MeV for elements with Z = 1 to Z = 94. *Atomic Data* **5,** 51–111 (1973).
11. Hubbell, J. H. *et al.* Atomic form factors, incoherent scattering functions, and photon scattering cross sections. *J. Phys. Chem. Ref. Data* **4,** 471–494 (1975).
12. Ditmire, T. Simulation of exploding clusters ionized by high-intensity femtosecond laser pulses. *Phys. Rev. A* **57,** 4094–4097 (1998).
13. Reiss, H. R. Physical basis for strong-field stabilization of atoms against ionization. *Laser Phys* **7,** 543–550 (1997).
14. Weaver, L. H. & Matthews, B. W. Structure of bacteriophage T4 lysozyme refined at 1.7 ångstroms resolution. *J. Mol. Biol.* **193,** 189–199 (1987).
15. Rischel, C. *et al.* Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390,** 490–492 (1997).
16. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276,** 368–372 (1978).
17. Bottcher, B., Wynne, S. A. & Crowther, R. A. Determination of the fold of the core protein of hepatitis b virus by electron cryomicroscopy. *Nature* **386,** 88–91 (1997).
18. van Heel, M. *et al.* Angular reconstitution in three-dimensional electron microscopy. Historical and theoretical aspects. *Scanning Micros.* **11,** 195–210 (1997).
19. Mueller, F. *et al.* The 3D arrangement of the 23 S and 5 S rRNA in the *Escherichia coli* 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 angstrom resolution. *J. Mol. Biol.* **298,** 35–59 (2000).
20. Stowell, M. H. B., Miyazawa, A. & Unwin, N. Macromolecular structure determination by electron microscopy: new advances and recent results. *Curr. Opin. Struct. Biol.* **8,** 595–600 (1998).
21. Miao, J. W., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400,** 342–344 (1999).
22. Szöke, A. Holographic methods in X-ray crystallography. 2. Detailed theory and connection to other methods of crystallography. *Acta Crystallogr. A* **49,** 853–866 (1993).
23. Faigel, G. & Tegze, M. X-ray holography. *Rep. Progr. Phys.* **62,** 355-393 (1999).
24. Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. Electrospray time of flight mass spectrometry of the intact MS2 virus capsid. *J. Am. Chem. Soc.* **122,** 3550–3551 (2000)
25. Siuzdak, G. *et al.* Mass-spectrometry and viral analysis. *Chem. Biol.* **3,** 45–48 (1996).
26. Rostom, A. A. Detection and characterization of intact ribosomes in the gas phase of a mass spectrometer. *Proc. Natl Acad. Sci. USA* **97,** 5185–5190 (2000).
27. Lin, T. W., Porta, C., Lomonossoff, G. & Johnson, J. E. Structure-based design of peptide presentation on a viral surface: The crystal structure of a plant/animal virus chimera at 2.8 Ångström resolution. *Fold. Des.* **1,** 179–187 (1996).
28. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **91,** 43–56 (1995).
29. van Gunsteren, W. F. & Berendsen, H. J. C. *Gromos-87 Manual* (Biomos BV, Groningen, 1987).
30. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 331–342. (D. Reidel, Dordrecht, 1981).

Acknowledgements

We are grateful to G. Faigel, S. Hovmöller, A. Szöke, M. van Heel and A. Pratt for discussions. This work was supported by the Swedish Research Councils NFR and TFR as well as the EU-BIOTECH programme.

Correspondence and requests for materials should be addressed to J.H. (e-mail: janos@xray.bmc.uu.se).