Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*

Maria A. Schumacher, Barry K. Hurlburt & Richard G. Brennan

Nature **409,** 215–219 (2001).

....

In light of the recent X-ray structure determination of the SarR-maltose binding fusion protein by Zhang *et al.*[1], we re-examined our crystal structure determinations of the *Staphylococcus aureus* transcription regulator, SarA, which were reported in the 11 January 2001 issue of *Nature*. Although suggested homologues (27% identity), the SarA and SarR structures are significantly dissimilar. We now believe that our structures of the apo and DNA-bound forms of SarA may harbour some anomalies.

We are performing a comprehensive examination of the structure and function relationships of a global virulence gene regulator in *S. aureus*, SarA. The full-length, recombinant protein was expressed in *E. coli* and purified to apparent homogeneity. This protein, whether generated in the laboratory of B.K.H. or that of R.G.B., behaved as expected. It bound with high affinity to *cis* regulatory sequences upstream of virulence genes previously reported to be controlled by SarA. These data were reported in refs 2 and 3.

To investigate why the SarA and SarR structures are different, we first reassessed our structure-determination processes. The SIR and MAD (as well as averaged)-derived phases that were used to calculate our SarA–DNA complex structure resulted in electron density maps that showed consistent secondary structure features different from those of SarR. In addition, MAD data for one of the 'apo' SarA crystals revealed similar features to the DNA-bound form of SarA. These results suggested to us that we were working with a highly flexible molecule that could undergo remarkable structural changes. One point is the non-standard, although we feel correct, manner by which we calculated the value of R_{free}, that is, both structures were not refined with CNS and therefore we may have erred in the use of this quality-control standard. It remains a possibility that our intensity data were simply not of high enough quality to discern the structure of Zhang *et al.*[1]. Although attempted multiple times, however, molecular replacement experiments to solve the apo SarA and the SarA–DNA protein structures, using our intensity data and SarR as the search model, have failed. Moreover, our heavy-atom sites are inconsistent with their SarR location. We are left with the possibility that in the absence of a carrier protein, such as MBP, or cognate DNA of sufficient length, we have crystallized a protein with some anomalously folded regions, which in our structures display a great deal of conformational flexibility. In support of this are the facts that there is a remarkable change in space group (from $P2_12_12_1$ to $P2_12_12$) and c cell edge (from 141 Å to 27 Å) upon freezing; all cryocooled crystals were non-isomorphous; our protein becomes inactive over time and degrades; our SarA–DNA complex reveals only nonspecific contacts; and there is an unprecedented change in protein conformation upon ligand binding.

To test the validity of our structures and to assess the importance of residues shown in the structures as important for function, we generated mutants of SarA that should result in aberrant activity. Using both an *in vivo* assay for virulence gene regulation and an *in vitro* DNA-binding assay for SarA function, most of the mutants have significantly altered activity (K. Sterba, M. S. Smeltzer and B.K.H., unpublished results). At this point, we do not know the significance and origin of the differences between the reported structures of SarA and SarR. □

1. Liu, Y. *et al.* Crystal structure of the SarR protein from *Staphylococcus aureus. Proc. Natl Acad. Sci. USA* **98,** 6877–6882 (2001).
2. Rechtin, T. M. *et al.* Characterization of the SarA virulence gene regulator of *Staphylococcus aureus. Mol. Microbiol.* **33,** 307–316 (2001).
3. Blevins, J. S., Gillaspy, A. F., Rechtin, T. M., Hurlburt, B. K. & Smeltzer, M. S. The staphylococcal accessory regulator (sar) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (*cna*) in an *agr*-independent manner. *Mol. Microbiol.* **33,** 317–326 (2001).

Biodegradation of oil in uplifted basins prevented by deep-burial sterilization

A. Wilhelms, S. R. Larter, I. Head, P. Farrimond, R. di-Primio & C. Zwach

Nature **411,** 1034–1037 (2001).

....

Figure 2 inadvertently contained plotting errors for oils from the Tampen Spur and Viking Graben data sets and the plotting of pristane/(*n*-heptadecane + pristane) for the Barents Sea data set rather than pristane/*n*-heptadecane as described. The corrected figure is shown below with all the offshore data now correctly plotted as depth from sediment surface. The shallowest Barents Sea sample plots with pristane/*n*-heptadecane slightly greater than 1, but alkane, isotope and light hydrocarbon distributions suggest that this oil is not biodegraded. Reservoirs in the Barents Sea cooler than 80 °C at present are found at depths shallower than 2,000 m. Our conclusions remain unchanged. □
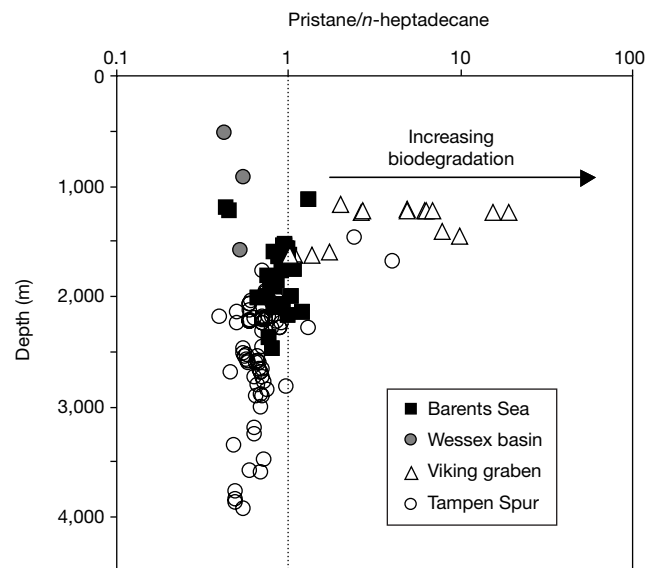

Figure 2 Degree of oil biodegradation, assessed by the pristane/*n*-heptadecane ratio, as a function of current reservoir depth below sediment surface for oils from four different settings.

computed as a function of the fragment length and position, and of the length and position distributions of fragments in the prey library (these distributions are calibrated using data from random sequencing).

The local score is the probability for a given SID to be obtained under the equal chance hypothesis, that is, as a result of random noise. It is deduced by combining probabilities p (using a binomial law) from each of the independent fragment defining it. A (global) PBS is computed for each protein interaction after pooling results from all screens. On the basis of an independence hypothesis, scores from different screens are combined together when the same protein domain pair is involved. The resulting PBS thus represents the probability that the protein–protein interaction is due to noise. Scores are real numbers ranging from 0 to 1, but are grouped in four categories (A, B, C and D) for practical purposes. Finally, the global connectivity of the interaction map is analysed to tag separately (category E) SIDs found as prey with frequency greater than a fixed threshold: the PBS of each protein–protein interaction involving highly connected SIDs is set to 1. Both the intercategory thresholds and the high-connectivity threshold were defined manually, taking into account the nature of the studied organism, the relevant library and the current coverage of the proteome ($A < 1^{e-10} < B < 1^{e-5} < C < 1^{e-2.5} < D$; the E category corresponds to prey SIDs selected with more than 4 baits and was arbitrarily attributed a PBS value of 1).

Bioinformatics

Several algorithms and software were implemented in the production database to facilitate experimental steps, such as a 'bait program' that designed automatically oligonucleotides for PCR amplification and sequencing of bait constructs, a 'prey program' that determined the position of each fragment in the genome and its coding capacity (such as intergenic, antisense, nucleotide position in an ORF, coding frame). The interactions were then analysed through a web-based software platform, the PIMRider developed at Hybrigenics and accessible through the web interface (http://pim.hybrigenics.com). Academic users will be granted a free licence. Other users will have to purchase a commercial licence. The *H. pylori* PIMRider platform is linked to the PyloriGene database.

1. Fields, S. The future is function. *Nature Genet.* **15,** 325–327 (1997).
2. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16,** 277–282 (1997).
3. Bartel, P. L., Roecklein, J. A., SenGupta, D. & Fields, S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet.* **12,** 72–77 (1996).
4. Flajolet, M. *et al.* A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* **242,** 369–379 (2000).
5. McCraith, S., Holtzman, T., Moss, B. & Fields, S. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc. Natl Acad. Sci. USA* **97,** 4879–4884 (2000).
6. Ito, T. *et al.* Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA* **97,** 1143–1147 (2000).
7. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
8. Walhout, A. J. M. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122 (2000).
9. Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
10. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28,** 45–48 (2000).
11. Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397,** 176–180 (1999).
12. Moszer, I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Lett.* **430,** 28–36 (1998).
13. Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J. P. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nature Struct. Biol.* **5,** 25–29 (1998).
14. Cussac, V., Ferrero, R. L. & Labigne, A. Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. *J. Bacteriol.* **174,** 2466–2473 (1992).
15. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59,** 451–480 (1995).
16. Skouloubris, S., Thiberge, J. M., Labigne, A. & De Reuse, H. The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival *in vivo*. *Infect. Immun.* **66,** 4517–4521 (1998).
17. Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. A H⁺-gated urea channel: the link between *Helicobacter pylori* urease and gastritic colonization. *Science* **287,** 482–485 (2000).
18. Dong, Z., Onrust, R., Skangalis, M. & O'Donnell, M. DNA polymerase III accessory proteins. I. holA and holB encoding delta and delta′. *J. Biol. Chem.* **268,** 11758–11765 (1993).
19. Liu, X. & Matsumura, P. An alternative sigma factor controls transcription of flagellar class-III operons in *Escherichia coli*: gene sequence, overproduction, purification and characterization. *Gene* **164,** 81–84 (1995).
20. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824 (1999).
21. Mooney, R. A. & Landick, R. RNA polymerase unveiled. *Cell* **98,** 687–690 (1999).
22. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27,** 919–929 (1999).
23. Ferrero, R. L., Cussac, V., Courcoux, P. & Labigne, A. Construction of isogenic urease-negative mutants of *Helicobacter pylori* by allelic exchange. *J. Bacteriol.* **174,** 4212–4217 (1992).

Acknowledgements

Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*

Maria A. Schumacher*†, **Barry K. Hurlburt**‡ & **Richard G. Brennan***

* *Department of Biochemistry and Molecular Biology and* † *Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA*
‡ *Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA*

Staphylococcus aureus **is a major human pathogen, the potency of which can be attributed to the regulated expression of an impressive array of virulence determinants. A key pleiotropic transcriptional regulator of these virulence factors is SarA, which is encoded by the** *sar* **(staphylococcal accessory regulator) locus**[1-3]**. SarA was characterized initially as an activator of a second virulence regulatory locus,** *agr,* **through its interaction with a series of heptad repeats (AGTTAAG) within the** *agr* **promoter**[4]**. Subsequent DNA-binding studies have revealed that SarA binds readily to multiple AT-rich sequences of variable lengths**[4-11]**. Here we describe the crystal structure of SarA and a SarA–DNA complex at resolutions of 2.50 Å and 2.95 Å, respectively. SarA has a fold consisting of a four-helix core region and 'inducible**
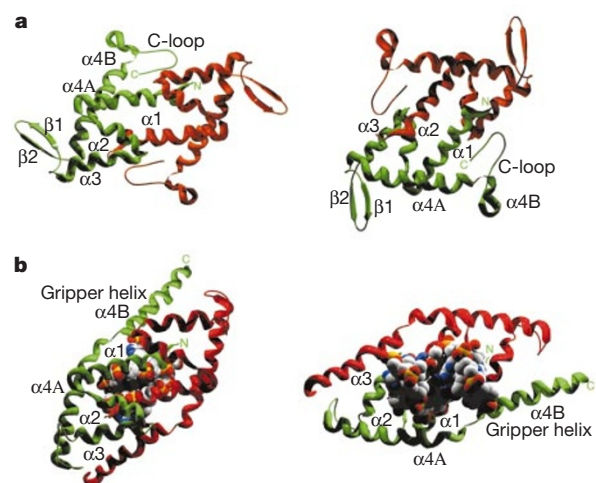

Figure 1 The SarA dimer and SarA–DNA complex. **a**, Views of SarA looking directly into the DNA-binding pocket or from the 'back' side. The secondary structures, β-hairpin and C-terminal loop of one monomer are labelled. Each monomer is red or green. **b**, The SarA–DNA complex in the identical orientation of the corresponding apo SarA directly above. The DNA duplex is shown as CPK atoms, with carbon, nitrogen, oxygen and phosphates coloured white, blue, red and yellow, respectively. The α4B gripper helix of one monomer is also labelled. The narrow and deep minor groove and major groove are seen on the left and right respectively.

regions' comprising a β-hairpin and a carboxy-terminal loop. On binding DNA, the inducible regions undergo marked conformational changes, becoming part of extended and distorted α-helices, which encase the DNA. SarA recognizes an AT-rich site in which the DNA is highly overwound and adopts a D-DNA-like conformation by indirect readout. These structures thus provide insight into SarA-mediated transcription regulation.

We determined the crystal structure of full-length SarA (125 residues) from *S. aureus* strain DB by multiple-wavelength anomalous dispersion (MAD). Notably, on cryo-protection and freezing, crystals of SarA changed space group from $P2_12_12_1$ to $P2_12_12$ (Table 1), and the diffraction limit increased from about 3.0 Å to beyond 2.5 Å. The structure of SarA (Fig. 1a) has a new protein fold as ascertained by searches with the DALI server[12]. The SarA monomer contains five α-helices, a short β-hairpin and a long C-terminal loop (C-loop). The topology is α1, (residues 10–30); α2, (residues 32–40); α3, (residues 44–56); β1, (residues 62–64); β2, (residues 68–70); α4A, (residues 74–94); α4B, (residues 97–103); and the C-loop, (residues 104–121). Two notable features are the protrusion of the β-hairpin and the C-loop from opposite ends of the mono-mer and the irregular conformation of helix α4B, which immediately precedes the C-loop (Fig. 1a). We only observed weak density for residues 116–121, which are located at the very end of the C-loop. These residues are not present in functional SarA protein

from some *S. aureus* strains, suggesting that they are dispensable for function[13].

The physiologically relevant SarA dimer[11] is readily identified in the structure (Fig. 1a). The formation of this dimer leads to the creation of a hydrophobic core forged mainly from extensive contacts between α1 and α1′ (where the prime indicates the other subunit), which pack in an antiparallel fashion. This extensive dimerization interface buries 720 Å^2 of accessible surface area per monomer and remains essentially unchanged in the protein–DNA complex. The contribution of Phe 80 to the nonpolar core of the dimer is notable, as it is the sole hydrophobic residue found on α4A, which provides a critical hydrophobic attachment point for anchoring this helix to the rest of the protein. In addition to this contact, a salt bridge between residues Glu 29 from α1 and Lys 83 from α4A provides a second contact point between α4A and the body of the protein.

The β-hairpin and C-loop, which project out from opposite ends of the protein, are tethered to the core by few contacts (Fig. 1a). The C-loop is anchored by residues 111–116, which loop back and pack against the amino terminus and residues from α1 of the same subunit. The β-hairpin is fastened to the body of the protein by stacking interactions between β-hairpin residue Tyr 62 and core residues, Phe 30 and Phe 80, and a hydrogen bond between β-hairpin residue Gln 64 and Tyr 92 from α4A. The main stabilization of these extended structures, however, is from crystal packing contacts in which the β-hairpin of one subunit packs against the C-loop of another subunit in the crystal. Such ordering is consistent with the marked increase in diffraction observed when data are collected under cryo-conditions, which may serve to 'freeze out' a specific conformation.

Examination of the SarA structure reveals no known DNA-binding motif or obvious DNA-binding pocket, with the possible exception of an unusual cleft that is formed at the dimer interface. This shallow cavity is formed by the α1 and α1′ helices, which would form the base, and the α4A and α4A′ helices, which would form the sides (Fig. 1a). However, attempts to model a fragment of duplex B-DNA into this cleft were unsuccessful as the cleft is too narrow. We therefore used co-crystallization studies to determine the mechanism of DNA binding by SarA. Because there is no known SarA DNA consensus site, we carried out co-crystallization trials of SarA with numerous deoxynucleotides that encompassed the AT-

Table 1 SarA and SarA-DNA crystallographic data

SarA

	λ1	λ2	λ3	λ4
λ	0.9798	0.9795	0.9250	1.0630
No. of total reflections	22,932	22,878	23,121	21,507
No. of unique reflections	5,757	5,763	5,740	5,660
Per cent complete	90.4	90.5	90.0	89.1
R_{sym} (%)*	8.5	8.7	8.7	8.2
$I/\sigma(I)$	4.7	5.0	4.9	4.9
Phasing power†	1.84/1.14	1.82/2.34	0.43/1.69	
R_{cullis}‡	0.605	0.625	0.798	
Figure of merit§	0.688			

Refinement

Resolution limit (Å)	2.5
R/R_{free} (%)‖	19.9/26.7

r.m.s. deviation

Bond angles (°)	2.10
Bond lengths (Å)	0.013
B values (Å^2)	4.09

SarA–DNA

	Native	Thimerosal
Resolution	2.95	3.5
No. of total reflections	14,850	10,297
No. of unique reflections	10,951	8,623
Per cent complete	89	80
R_{sym} (%)*	7.8	10.8
$I/\sigma(I)$	6.5	4.5
Phasing power†		1.99
R_{cullis}‡		0.505
§ Figure of merit		0.408

Refinement

Resolution range	10.0–2.95
R/R_{free} (%)‖	22.0/30.8

r.m.s. deviation

Bond angles (°)	1.98
Bond lengths (Å)	0.011
B values (Å^2)	4.12

*$R_{sym} = \Sigma|I_o - \langle I\rangle|/I_o$, where I_o is the observed intensity and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry related reflections.
† Phasing power = r.m.s.($|F_h|/E$), where $|F_h|$ is the heavy atom structure factor amplitude and E is residual lack of closure error.
‡ $R_{cullis} = \Sigma||F_{h.obs}| - |F_{h.calc}||/\Sigma|F_{h.obs}|$ for centric reflections where $|F_{h.obs}|$ is the observed heavy atom structure factor amplitudes and $|F_{h.calc}|$ is the calculated heavy atom structure factor amplitude.
§ Figure of merit = $\langle\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)\rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.
‖ $R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$. The lower resolution limit of refinement for all structures was 10.0 Å and $R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$; where all reflections belong to a test set of 10% randomly selected data.
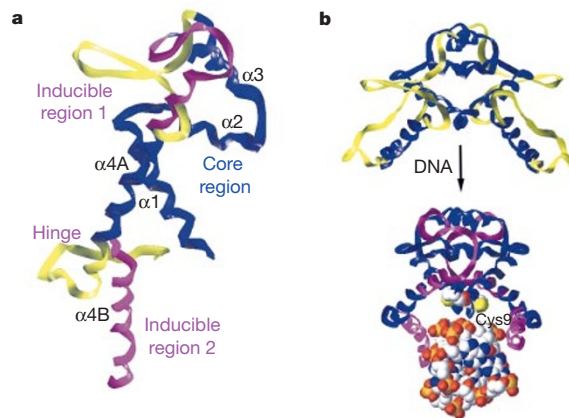

Figure 2 SarA inducible regions. **a**, Superimposed SarA monomers of the DNA-bound and apo protein. The core region is blue and the inducible regions are yellow and magenta for the apo and DNA bound proteins, respectively. The hinge is labelled. **b**, Conformational changes induced by DNA binding. The apo form is shown on top (inducible regions in yellow) and the DNA-bound form below (inducible regions in magenta). The DNA duplex and Cys 9 side chains, which line the DNA binding channel, are shown as CPK atoms, with carbon, nitrogen, oxygen and phosphate/sulphur coloured white, blue, red and yellow, respectively.

rich heptad repeat (AGTTAAG) located within the *agr* promoter[4]. Synthetic DNA containing multiple heptad repeats was shown to bind SarA and could be used to purify the protein from *S. aureus* extracts[9]. Because the minimal SarA-binding site is unclear, we used deoxyoligonucleotides duplexes ranging from 7 to 42 bp (base pairs). Data quality crystals were obtained only with an 8-bp duplex (8-mer) with 5′A and T overhangs, (5′AGTTAAGA3′)·(3′CAATTCTT5′).

We determined the structure of the full-length SarA−8-mer complex by single isomorphous replacement, which uses phases from a thimerosal derivatization of the highly reactive cysteine, Cys 9 (Table 1). Individual substitution of each thymine with 5-iodouracil failed to reveal any iodine sites, indicating that the DNA was disordered. Despite this disorder, clear electron density was observed for the phosphodiester backbone of 6 bp and the 3′ nucleotide at the end of each strand, as well as the SarA dimer (Fig. 1b).

Comparison of the SarA structure of the DNA bound form with that of the apo form shows that the four-helix core region, α1, α2, α3 and α4A, are essentially identical with root mean square (r.m.s.) deviations of 0.6 Å for the corresponding Cα atoms of residues 10–45 and 76–90 (Fig. 2a). However, the extended regions comprised of the β-hairpin and C-loop regions, which lie on either side of α4A, have undergone marked conformational changes. Because of their obvious structural flexibility, we called these the 'inducible regions' (Fig. 2a). Structural changes in inducible region 1 include the conversions of β2 of the β-hairpin and five additional residues (71–75) to a helical conformation, which extend α4A by three turns at its N terminus (Fig. 2a, b). Structural changes in inducible region 2 involve the incorporation of residues 103–114 of the C-loop and residues 94–95 into helix α4B, which increases to six full turns. The net result of these large conformational changes is the formation of an extended and distorted α-helix 4 (residues 67–114), which thoroughly encases the DNA. The main distortion in α4 occurs at residues 91–93, which seem to act as a flexible hinge that allows the α4B helices to re-orient and encase the DNA, a process impossible with a single, long helix. Such flexibility is consistent with the different orientations of these 'gripper helices' in each monomer, whereby one is more proximal and the other more distal to the DNA (Fig. 1b). The inducible regions seem to undergo their structural transitions to avoid C-loop−DNA steric clash and to maximize electrostatic and hydrophobic interactions with the DNA.

The DNA is bound in the long channel between α4A and α4A′ and the helices of the inducible regions of the SarA dimer (Figs 1b and 2b). As noted, this pocket is too narrow to accommodate a duplex B-DNA structure. Instead the DNA adopts an unusual conformation that is highly overwound and marked by a very narrow minor groove. In contrast to canonical B-DNA, which has an average helical twist of 34.3°, a minor-groove width of 5.7 Å and 10.5 bp per turn, the SarA-bound DNA displays an average helical twist of 41.8°, a minor-groove width of 2.8 Å and 8.6 bp per turn (Fig. 3)[14]. Notably, there is a similar DNA conformation in fibre diffraction studies on DNA with AT and AATT repeats[15,16]. This DNA, called D-DNA, is also characterized by an extremely narrow minor groove (1.7 Å), a large helical twist of 44.8° and an over-wound helix with 8 bp per turn (Fig. 3). Fibre diffraction studies on such AT-rich repeats show that these polymers do not adopt or transit through the A-DNA conformation, and polymers of (dA-dA-dT)·(dA-dT-dT) adopt only a B-DNA or D-DNA conformation depending on the environment. Specifically, divalent cations and low humidity favour D-DNA[15].

The 'relaxed sequence' binding of SarA is consistent with the lack of base-specific contacts observed in the SarA−DNA complex, and suggests that indirect readout is the primary mode of DNA recognition. SarA is anchored to the DNA by phosphate contacts of residues from helices α1/α1′ and α4A/α4A′. All contacts are made to the minor groove of the DNA, which faces the protein. A key set of SarA−phosphate contacts is provided by an extended network of hydrogen bonds and ionic interactions at either end of the dimer dyad consisting of P_5−Tyr 18/Arg 84−Glu 11′−Ca^{2+}−$P_{7'}$ (Fig. 4). This network both requires and buttresses the narrow minor groove, which is an intrinsic property of the D-DNA conformation (Fig. 4)[15].

In addition to these phosphate backbone contacts, SarA makes a number of hydrophobic contacts to the DNA minor groove. These interactions are made possible by the structural changes that allow SarA to encase the DNA and which result in the creation of an environment favouring D-DNA formation[15]. Specifically, there is a
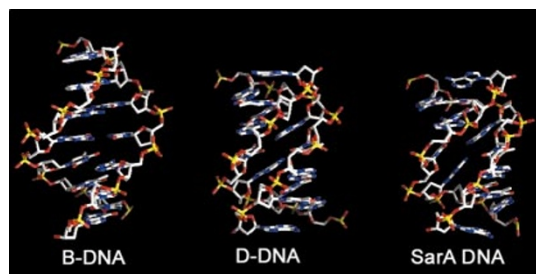

Figure 3 Conformation of SarA-bound DNA. B-DNA, D-DNA and DNA from the SarA−DNA complex are shown. Each 6-bp fragment is shown in the same orientation. We note the extremely narrow minor groove and overwound helices of the D-DNA and SarA-bound DNA structures, which are in sharp contrast to those B-DNA conformational features. Carbon, nitrogen, oxygen and phosphorous atoms are coloured white, blue, red and yellow, respectively.
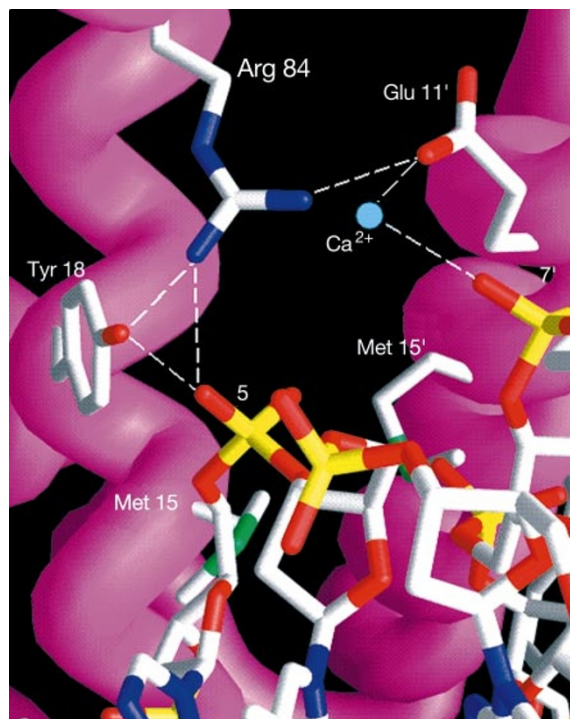

Figure 4 SarA−DNA contacts. One end of the network of protein−DNA interactions that stabilize the narrow minor groove. The side chains of Tyr 18, Met 15, Arg 84, Glu 11′ and Met 15′ and the DNA are shown as sticks; carbon (white), oxygen (red), nitrogen (blue), sulphur (green) and phosphorus (yellow) atoms are also shown. The calcium ion is depicted as a light blue sphere and interactions are shown as dashed lines. The six base pairs and 3′ nucleotide overhangs of the DNA duplex are labelled as in Methods. Figures 3 and 4 were made with GRASP[29].

large nonpolar face at the centre of the DNA-binding site created by residues Met 15 and Met 15′. These residues abut directly and interact with the deoxyribose phosphate backbone of base pairs 5 and 6 and 5′ and 6′, respectively (Fig. 4). Furthermore, the α4A/α4B hinge and the N terminus of α4B, including residues Val 92, Leu 93, Ile 94, Leu 95 and Val 96, contribute to the hydrophobic environment of the binding pocket and may be potentially involved in van der Waals contacts to the bases. Indeed, the side chain of Leu 95 from the DNA proximal gripper helix is close to Ade 5. More extensive contacts with these aliphatic residues would require a longer DNA site that is bent towards the protein.

Other residues of the α4B gripper helices are also likely to be involved in binding longer DNA sites, which, given their ability to adopt several orientations, could wrap around these sites. These helices contain a highly positively charged face, comprising Lys 102, Lys 103, Arg 110, Lys 113 and Arg 114, and their two-fold equivalents, which line the far ends of the DNA-binding cleft and make long-range electrostatic contacts to the DNA. The α4B helix is highly analogous in structure, dynamics and function to the DNA-binding regions of the basic leucine zipper and basic helix–loop–helix proteins, as all three undergo a coil to helix transition on DNA binding, and contain numerous basic residues involved, or probably involved, in DNA binding[17]. The structures of SarA and the SarA–DNA complex seem to have captured for the first time the
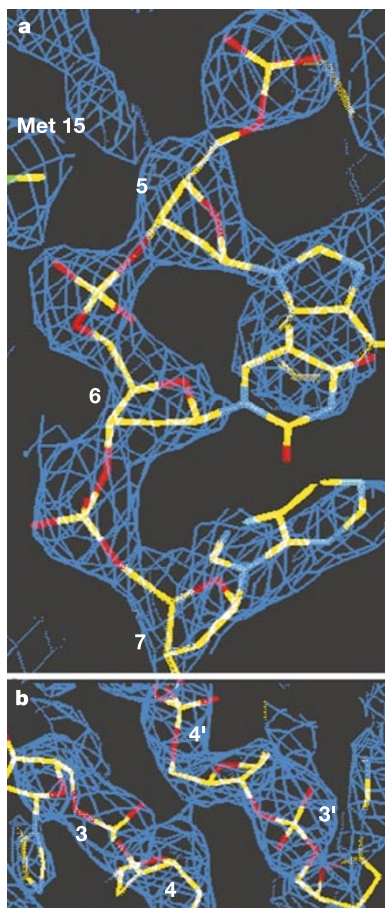

Figure 5 $2F_{obs} - F_{calc}$ omit electron density maps in which the DNA has been omitted. **a**, View of the omitted electron density map showing the region about the bases and phosphodiester backbone of nucleotides 5–7. **b**, View of the omitted electron density map highlighting the very close approach of the phosphodiester backbones across the minor groove. Both **a** and **b** are contoured at 1.2σ. The DNA is depicted as sticks, in which carbon, nitrogen, oxygen and phosphorous atoms are coloured yellow, blue, red and orange, respectively.

conformational end states of a DNA-induced coil-to-helix folding transition.

SarA both represses and activates transcription from different promoters. Moreover, activation of the *agr* locus is only partially responsible for the global effect of SarA on virulence gene expression, as both *sar*-dependent repression of genes encoding surface proteins and *sar*-dependent activation of genes encoding exoproteins occur[5–8]. What could explain the basis for such an apparent dichotomy in function? The SarA–DNA structure shows the marked overwinding of the DNA binding site that results in a significant decrease in the number of base pairs per turn of helix. This altered conformation could directly explain the observation that SarA activates the *agr* P3 promoter, in which there are about 20 bp between the −35 and −10 boxes[4], and represses transcription of the collagen-binding protein gene (*cna*), where the 17-bp spacer between −35 and −10 is optimal[7]. In the former site, the overtwisting would reduce the −35/−10 spacer to 17 bp, and in the latter the 'optimal' spacer would shrink to 14 bp. However, the overall regulatory effects of SarA are likely to be more complex, as indicated by the ability of this protein to bind numerous DNA sites in the *S. aureus* genome[4–11]. SarA, like the *Escherichia coli* proteins Fis, IHF and HU, might therefore function as a global, architectural DNA-binding protein that influences DNA superhelicity and, in turn, transcription in a manner dependent on binding-site position[18–20].

SarA regulates virulence genes in a growth-phase-dependent manner, a feature critical to the progression of infection in *S. aureus*. Therefore, the temporal regulation of SarA binding becomes critical. In this regard, we note that the *sar* locus encodes three overlapping transcripts all of which encode the SarA protein. Although these transcripts are initiated by different σ-factors at different times in the *S. aureus* life cycle, the total amount of SarA protein seems to remain constant[7]. This suggests that the activity of SarA is regulated by either its binding to other proteins or its post-translational modification. In support of the latter hypothesis, SarA activity is significantly reduced under oxygen-replete conditions[21]. This is particularly intriguing because Cys 9, the single cysteine in SarA, is highly reactive and is located in the DNA binding channel (Fig. 2b). Oxidation of the thiol side chain to a sulphone or cysteic acid would interfere directly with DNA binding.

Given the structural uniqueness and critical function of SarA in *S. aureus* virulence, the structures described here provide scaffolds from which to initiate the design of new and highly specific anti-staphylococcal chemotherapeutics. □

Methods

SarA purification and crystallization

We purified full-length recombinant SarA from strain DB as described[11]. We grew 523 l of bacteria to obtain enough protein to determine the structures of SarA and the SarA–DNA complex. We grew crystals of SarA by hanging drop/vapour diffusion at 4 °C using the protein at 10 mg ml⁻¹ and a precipitant of 6% isopropanol, in 50 mM Tris, pH 7.6 and 10 mM MgCl₂. We characterized the space group at room temperature as orthorhombic, $P2_12_12_1$ with $a = 44.8$ Å, $b = 84.4$ Å, $c = 141$ Å; however, at room temperature the crystals decay rapidly. Only glycerol worked as a cryoprotectant, yet typically produced a large increase in mosaic spread (0.3° to ∼ 2.2°). Cryoprotection extends the diffraction limit from 3.0 Å to beyond 2.5 Å resolution and, notably, changes the space group to $P2_12_12$ with $a = 44.8$ Å, $b = 84.4$ Å, $c = 27.6$ Å. The cryoprotected crystals contain a monomer per asymmetric unit and a solvent content of 25%.

Structure determination and refinement

We determined the structure by MAD[22], as individual frozen crystals display severe nonisomorphism. Because the wild-type protein contains only one methionine, a mutant, L95M, was constructed and selenomethionine L95M SarA was purified and isomorphously crystallized. We collected MAD data at the Stanford synchrotron radiation laboratory at line 1–5 using an ADSC 180 mm charged coupled device (CCD) detector. We processed the data with MOSFLM[23]. Only one site, corresponding to Met 95, was located in Patterson maps; the second site was found by difference Fourier techniques.

We refined the selenium atomic positions by treating the data as a special case of MIR[24]. The resulting map revealed clear density for helices and the β-hairpin. Model building[25] was aided by phase combination. We refined the structure with TNT[26]. Although the data extend beyond 2.5 Å resolution, the refinement was limited to 2.5 Å owing to problems

with the crystal mosaicity and anisotropy of the data. The current structure consists of residues 2–121 and 82 waters with 92% of the residues in favoured regions of the Ramachandran plot, 6% in generously allowed regions and 2% in unfavoured regions[27]. Residues Ile 104 and Ala 118 are the two outliers; Ile 104 lies in the tight turn between α4B and the C-loop; and Ala 118 has weak density along with residues 116–121

Crystallization of the SarA–DNA complex

We used a number of deoxyoligonucleotides in crystallization trials with SarA, and only those with an 8-mer duplex yielded data-quality crystals. We grew these crystals using hanging drop/vapour diffusion with a crystallization solution of 25% PEG 8000, 200 mM $CaCl_2$, 100 mM cacodylate, pH 6.5. They are monoclinic, space group $P2_1$, with $a = 54.5$ Å, $b = 65.2$ Å, $c = 57.8$ Å and $\beta = 118.0°$ and contain one SarA dimer and 8-mer duplex per asymmetric unit. All data were collected at room temperature with an ADSC multiwire area detector[28] and a RIGAKU RU200-H rotating anode X-ray generator with graphite monochromator operating at 40 kV and 150 mA.

Structure determination and refinement

Because of DNA disorder, iodinated deoxyoligonucleotides failed as derivatives, therefore phases were derived from a thimerosal derivative in conjunction with twofold averaging and solvent flattening[24]. Density was evident for the α1 and α4A helices and the backbone of helices α2 and α3. However, in averaged maps, the density for α4B was smeared and subsequently phase-combined; unaveraged maps revealed different orientations of these helices. The density for the phosphate backbone of 6 bp in the centre of the DNA-binding channel was also evident (Fig. 5). Good electron density was also observed for the bases but their identities were ambiguous and therefore, fitted with alternating A·T base pairs $(5'-A_1T_2A_3T_4A_5T_6A_7-3')$. We then subjected the structure to positional and tightly restrained B-factor refinement[26]. The current model includes residues 2–113 of one subunit and 2–121 of the other, and a 6-bp duplex with 3'-adenylate nucleotide overhangs. Analysis of the Ramachandran plot shows that 94% of the residues are in the most favoured regions, 5.5% in generously allowed regions and 0.5% in unfavoured regions (Phe 30 from one subunit is the outlier)[27]. Two large solvent peaks were located in the DNA minor-groove binding site. Their thermal parameters indicated that they are Ca^{2+} ions, which is consistent with the observation that either magnesium or calcium is required for crystallization of the SarA–DNA complex. Residual density in the DNA-binding channel indicates additional disordered DNA in the binding site. Although the gripper helices α4B and α4B′ take different orientations, residues 5–95 of each subunit overlay with r.m.s. deviations of 0.55 Å.

1. Projan, S. L. & Novickn, R. P. in *Staphylococci in Human Disease* (eds Cross, K. B. & Archer, G. L.) 55–81 (Churchill Livingstone, New York, 1997).
2. Cheung, A. L., Koomey, J. M., Butler, C. A., Projan, S. J. & Fischetti, V. A. Regulation of exoprotein expression in *Staphylococcus aureus* by a locus (*sar*) distinct from *agr*. *Proc. Natl Acad. Sci. USA* **89**, 6462–6466 (1992).
3. Cheung, A. L. & Projan, S. J. Cloning and sequencing of *sarA* of *Staphylococcus aureus*, a gene required for expression of *agr*. *J. Bacteriol.* **176**, 4168–4172 (1994).
4. Morfeldt, E., Tegmark, K. & Arvidson, S. Transcriptional control of the *agr*-dependent virulence gene regulator, RNAIII, in *Staphylococcus aureus*. *Mol. Microbiol.* **21**, 1227–1237 (1996).
5. Wolz, C. *et al. Agr*-independent regulation of fibronectin-binding protein(s) by the regulatory locus *sar* in *Staphylococcus aureus*. *Mol. Microbiol.* **36**, 230–243 (2000).
6. Cheung, A. & Ying, P. Regulation of α- and β-hemolysins by the *sar* locus of *Staphylococcus aureus*. *J. Bacteriol.* **176**, 580–585 (1994).
7. Blevins, J. S., Gillaspy, A. F., Rechtin, T. M., Hurlbert, B. K. & Smeltzer, M. S. The staphylococcal accessory gene regulator (*sar*) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (*cna*) in an *agr*-independent manner. *Mol. Microbiol.* **33**, 317–326 (1999).
8. Gaskill, M. E. & Khan, S. A. Regulation of the enterotoxin B gene in *Staphylococcus aureus*. *J. Biol. Chem.* **263**, 6276–6280 (1988).
9. Manna, A. C., Bayer, M. G. & Cheung, A. L. Transcriptional analysis of different promoters in the *sar* locus in *Staphylococcus aureus*. *J. Bacteriol.* **180**, 3828–3836 (1998).
10. Chien, Y.-t., Manna, A., Projan, S. J. & Cheung, A. L. SarA, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for *sar*-dependent gene regulation. *J. Biol. Chem.* **274**, 37169–37176 (1999).
11. Rechtin, T. M. *et al.* Characterization of the SarA virulence gene regulator of *Staphylococcus aureus*. *Mol. Microbiol.* **33**, 307–316 (1999).
12. Holm, L. & Sander, C. Protein structure comparisons by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
13. Heinrichs, J. H., Bayer, M. G. & Cheung, A. L. Characterization of the *sar* locus and its interactions with *agr* in *Staphylococcus aureus*. *J. Bacteriol.* **178**, 418–423 (1996).
14. Lavery, R. & Sklenar, H. J. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *Biomol. Struct. Dynam.* **6**, 63–91 (1988).
15. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
16. Mahendrasingam, A. *et al.* Time-resolved X-ray diffraction studies of the B ↔ D structural transition in the DNA double helix. *Science* **233**, 195–197 (1986).
17. Baxevanis, A. D. & Vinson, C. R. Interactions of coiled coils in transcription factors: where is the specificity? *Curr. Opin. Genet. Dev.* **3**, 278–285 (1993).
18. Pruss, G. & Drlica, K. DNA supercoiling and prokaryotic transcription. *Cell* **56**, 521–523 (1989).
19. Record, M. T. Jr, Reznikoff, W. S., Craig, M. L., McQuade, K. L. & Schlax, P. J. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 792–820 (ASM, Washington DC, 1996).
20. Oberto, J., Drlica, K. & Rouviere-Yaniv, J. Histones, HMG, HU, IHF: Meme combat. *Biochimie* **76**, 901–908 (1994).
21. Chan, P. F. & Foster, S. J. Role of SarA in virulence determinant production and environmental signal transduction in *Staphylococcus aureus*. *J. Bacteriol.* **180**, 6232–6241 (1998).
22. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
23. Collaborative computational project number 4. *J. Appl. Crystallogr. D* **50**, 760 (1994).
24. Furey, W. B. & Swaminathan, S. PHASE-95: a program package for the processing and analysis of diffraction data from macromolecules. In ;*Methods in Enzymol.* **Vol. 277** (eds Carter, C. W. & Sweet, R. M.) 590–620 (Academic, Orlando, 1997).
25. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A* **47**, 110–119 (1991).
26. Tronrud, D. E., TenEyck, L. F. & Matthews, B. W. An efficient general purpose least-squares refinement program for macromolecular structures. *Acta. Crystallogr. A* **43**, 489–501 (1985).
27. Laskowski, R. A., MacArthur, M. W. & Thorton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
28. Howard, A. J., Nielson, C. & Xuong, N. H. Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 452–472 (1985).
29. Nicholls, A., Sharp, K. & Honig, B. H. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Projection structure of a ClC-type chloride channel at 6.5 Å resolution

Joseph A. Mindell*, Merritt Maduke*, Christopher Miller* & Nikolaus Grigorieff*†

* Department of Biochemistry, Howard Hughes Medical Institute and † Rosenstiel Basic Medical Science Research Center and W. M. Keck Institute for Cellular Visualization, Brandeis University, Waltham, Massachusetts 02454, USA

Virtually all cells in all eukaryotic organisms express ion channels of the ClC type, the only known molecular family of chloride-ion-selective channels. The diversity of ClC channels highlights the multitude and range of functions served by gated chloride-ion conduction in biological membranes, such as controlling electrical excitability in skeletal muscle, maintaining systemic blood pressure, acidifying endosomal compartments, and regulating electrical responses of GABA (γ-aminobutyric acid)-containing interneurons in the central nervous system[1]. Previously, we expressed and purified a prokaryotic ClC channel homologue[2]. Here we report the formation of two-dimensional crystals of this ClC channel protein reconstituted into phospholipid bilayer membranes. Cryo-electron microscopic analysis of these crystals yields a projection structure at 6.5 Å resolution, which shows off-axis water-filled pores within the dimeric channel complex.

Of the approximately ten ClC isoforms found in eukaryotes, the muscle subtypes ClC-0 and ClC-1 are by far the most thoroughly studied. These voltage-gated channels are homodimers of polypeptides with a relative molecular mass of 90,000 that contain 10–12 putative transmembrane segments[3]. Detailed single-channel analysis indicates that these channels are double-barrelled homodimers in which the functional dimer contains two identical pores devoid of axial symmetry, each with its own voltage-dependent gate[4–8]. It is not known, however, whether this unusual molecular architecture applies to all ClC channels or only to the muscle subfamily. Moreover, this picture is not universally accepted; recently, a conventional

18. Miteva, T., Palmer, L., Kloppenburg, L., Neher, D. & Bunz, U. H. F. Interplay of thermochromicity and liquid crystalline behavior in poly(*p*-phenyleneethynylene): π-π interactions or planarization of the conjugated backbone? *Macromolecules* **33**, 652–654 (2000).

19. McQuade, D. T., Kim, J. & Swager, T. M. Two-dimensional conjugated polymer assemblies: Interchain spacing for control of photophysics. *J. Am. Chem. Soc.* **122**, 5885–5886 (2000).

20. Kim, J., McQuade, D. T., McHugh, S. K. & Swager, T. M. Ion-specific aggregation in conjugated polymers: Highly sensitive and selective fluorescent ion chemosensors. *Angew. Chem. Int. Edn Engl.* **39**, 3868–3872 (2000).

21. Cronwell, E. Excimer formation and luminescence in conducting polymers. *Trends Polym. Sci.* **5**, 218–222 (1997).

....

Biodegradation of oil in uplifted basins prevented by deep-burial sterilization

A. Wilhelms*, **S. R. Larter**†, **I. Head**†, **P. Farrimond**†, **R. di-Primio*** **& C. Zwach***

* *Norsk Hydro Research Centre, PO 7190, 5020 Bergen, Norway*
† *Fossil Fuels and Environmental Geochemistry Postgraduate Institute (NRG), University of Newcastle, Newcastle upon Tyne NE1 7RU, UK*

....

Biodegradation of crude oil by bacterial activity—which has occurred in the majority of the Earth's oil reserves[1]—is known to reduce greatly the quality of petroleum in reservoirs[2]. For economically successful prospecting for oil, it is therefore important to understand the processes and conditions in geological formations that lead to oil biodegradation. Although recent studies speculate that bacterial activity can potentially occur up to temperatures as high as 150 °C (refs 3, 4), it is generally accepted that effective petroleum biodegradation over geological time-scales generally occurs in reservoirs with temperatures below 80 °C (ref. 2). This appears, however, to be at odds with the observation that non-degraded oils can still be found in reservoirs below this temperature. Here we compile data regarding the extent of oil biodegradation in several oil reservoirs, and find that the extensive occurrence of non-biodegraded oil in shallow, cool basins is restricted to those that have been uplifted from deeper, hotter regions of the Earth. We suggest that these petroleum reservoirs were sterilized by heating to a temperature around 80–90 °C during deep burial, inactivating hydrocarbon-degrading organisms that occur in the deep biosphere. Even when such reservoirs are subsequently uplifted to much cooler regions and filled with oil, degradation does not occur, implying that the sterilized sediments are not recolonized by hydrocarbon-degrading bacteria.

Biodegradation of oil leads to a systematic decrease in paraffin content and an increase in oil density, sulphur content, acidity and viscosity[2], with negative economic consequences for oil production and refining operations. So prediction of the degree of biodegradation of oil is important for assessing the risk of finding degraded oils in an exploration target before drilling. The details and rates of the processes involved in crude-oil degradation are still poorly understood, but the central role of bacteria in subsurface petroleum degradation is accepted. Although there is abundant evidence for the presence of active bacteria deep (>1 km) in the Earth's crust[3,5–7], and a view among biologists that life in deep sediments may occur even up to 150 °C (refs 3, 4), there is a general view today among petroleum geoscientists that biodegradation in reservoirs ceases around 75–80 °C (refs 2, 8) in the zone of thermophilic organisms.

Hyperthermophilic organisms grow best above 80 °C, and have been reported to live at temperatures of up to 113 °C (ref. 9), but often display no growth below 60 °C (ref. 4). Most reports of hyperthermophilic activity are limited to environments rich in reduced electron donors, electron acceptors and inorganic nutrients necessary for biosynthesis—environments typically associated with shallow sediment or near-surface hydrothermal activity[10], where high levels of metabolic activity can be supported. Although petroleum reservoirs are rich in reduced organic electron donors (hydrocarbons), most petroleum reservoirs and deep aquifers are nutrient depleted[11] and recent studies indicate that biodegradation of petroleum in deep reservoirs is occurring at very slow net rates, typically consuming around 10^{-6} mmol oil per litre per day (ref. 12). These are similar to the respiration rates suggested for other deep sediments, and several orders of magnitude slower than respiration rates in anaerobic laboratory or near-surface sediment environments[11,13,14]. As electron donor (oil) supply is not limiting, degradation is presumably nutrient limited. Under these stressed, nutrient-limited conditions, it is unlikely that the temperature extremes survived by organisms in nutrient-rich hydrothermal environments are relevant. Furthermore, the empirical upper temperature limit of 80 °C observed for petroleum biodegradation in deep reservoirs is considerably lower than temperatures quoted for the denaturation and decomposition of many bacterial and archaeal proteins and nucleic acids[9,15]; this implies that active hyperthermophilic hydrocarbon-degraders are probably not present in petroleum reservoirs. Indeed, the single report of a hyperthermophilic archaeon that grew in enrichment cultures with crude oil as sole source of carbon[16] did not provide unequivocal evidence that the hyperthermophiles isolated could degrade hydrocarbons.

Although the occurrence of hyperthermophiles in petroleum reservoirs is now well documented[17–21], only four isolates capable of growth at greater than 90 °C have been reported[16–19]. In the most extreme of these cases (growth at 102 °C), evidence was presented that suggested that the organisms concerned may have originated from hydrothermal vent systems and were introduced with injected sea water[16]. The other cases where hyperthermophiles have been isolated are also sea-water-flooded reservoirs[19]. For the most part, microbial enrichments from petroleum reservoir samples conducted at temperatures above 85 °C have been unsuccessful[19,21], with increasing success of enrichment cultures with decreasing growth temperature (reaching 77% ($n = 26$) at temperatures below 60 °C)[19]. We also note that the only thermophilic anaerobic bacterium for which hydrocarbon degradation has been unequivocally demonstrated (an anaerobic sulphate-reducer) has an optimum growth temperature of only 60 °C (ref. 22). It has also been found that enrichment cultures supplied with crude oil as the sole source of organic carbon at temperatures greater than 85 °C failed to produce cultures of putative hydrocarbon-degrading Archaea although hyperthermophiles that grow at temperatures up to 102 °C could be isolated[16]. Taken together, data from most reports of Bacteria and Archaea isolated from petroleum reservoirs support the notion that the organisms which inhabit these reservoirs do not thrive in the upper temperature ranges for hyperthermophiles (up to 113 °C), and show little indication of hydrocarbon degrading capacity above 80 °C. We realize, however, that any interpretation must be viewed cautiously (as most reports provide little information on failed enrichments, and recent experience in defining the limits of life warn of inherent difficulty in definitively excluding life processes from even severe environments).

Biodegraded oils are common in cool (<80 °C), shallow reservoirs in currently subsiding basins such as those in the North Sea,

the Gulf of Mexico and the Niger delta, or where very near surface exposure of oil occurs—as with tar sands[1] or oil seeps. However, some cool, shallow reservoirs contain non-degraded petroleum, and these are of particular interest for petroleum exploration because of their greater commercial value and ease of exploitation. In our experience, such non-degraded shallow oils are chiefly found either in reservoirs with a very recent oil charge or, more interestingly, in reservoirs that have been uplifted from their maximum burial depth. Examples of basins containing light non-degraded oils in currently shallow cool reservoirs that have been exposed to substantially higher temperatures in the past occur in the Barents Sea (Norway), Wessex and East Midlands basins (UK), Michigan, San Juan, Anadarko and Illinois basins (USA), Murzuk basin (Libya), Zagros folded zone (Iraq), Cooper-Eromanga basins (Australia) and the Papua fold and thrust belt (see Supplementary Information). Although late oil recharge into these uplifted reservoirs may explain the observed low levels of net biodegradation, we consider this process unlikely to have been significant in most cases. This is because the hydrocarbon-generating source rocks cool when the basins are uplifted, and primary oil charging ceases.

Comparison of levels of biodegradation in similar petroleum systems that have very different reservoir temperature histories before oil charging indicates that uplifted and continuously subsiding basins show quite different biodegradation characteristics (Fig. 1). In the uplifted Barents Sea (Norway) and Wessex (UK) basins, oils are currently found in reservoirs over depth ranges of 1,000–2,000 m and 500–1,600 m, respectively, with some of these oils being found at temperatures below 40 °C. These oils are uniformly non-degraded (Fig. 2). In contrast, oils in the currently subsiding Viking graben found in reservoirs between 1,000 and 2,000 m depth show variable pristane/n-heptadecane ratios ranging from <1.0 (non-degraded) to over 10 (Fig. 2). This indicates a wide range of degradation, related to the relative timing of fresh oil charge, and variable remixing of fresh and degraded oils in the reservoirs[8,23]. All three petroleum systems are similar in that the source rocks are Jurassic marine shales of similar mineralogy, with

similar type II kerogen source organic matter (Draupne Formation[24], Hekkingen Formation[25] and the Lias[26]) producing broadly similar initial gross oil compositions in the reservoirs.

The Viking graben reservoirs have had a relatively continuous subsidence history, marked by only minor uplift events. All the reservoirs are at present at maximum temperature, and the reservoir temperature in these shallow degraded Tertiary reservoirs has never exceeded 80 °C (ref. 27). Basin modelling indicates that oil charging of the traps occurred from the Late Cretaceous, but most traps have had oil charges as recent as 5 Myr ago (ref. 28). Generally well defined trends of depth versus degradation observed regionally in the Viking graben[8] suggest that degradation has occurred close to current reservoir depths. Figure 2 also shows data for oils from Jurassic reservoirs in the Tampen Spur area of the North Sea which contain essentially the same type of oil as found in the Tertiary reservoirs of the Viking graben but are at their maximum depths of greater than 2,000 m and at temperatures of over 80 °C. These oils are non-degraded.

The reservoirs in the Barents Sea and Wessex basins have had a more turbulent burial history, including extensive uplift. In the case of the Barents Sea, major oil charging occurred in the Cretaceous followed by uplift of the reservoirs by >1.5–2 km in the early Tertiary[25,28]. Maximum temperatures experienced by the currently shallowest reservoirs were around 85 °C, with the deeper reservoirs encountering higher temperatures. The Wessex basin has a similar history. Oil charging from the Lias occurred in the Cretaceous[26], and was followed in late Cretaceous time by uplift of approximately 2 km (refs 29, 30). None of the reservoired oils at the Wytch Farm or Kimmeridge Bay oilfields in this basin show any signs of degradation, despite the oils being in cool reservoirs as shallow as 500 m. Maximum temperatures of the reservoirs at Wytch Farm and Kimmeridge Bay oilfields are indicated to have been around 100 °C and 90 °C (ref. 29) respectively. It is probable that the current oil charge at Wytch Farm has been there since the Cretaceous[26], and that at Kimmeridge Bay oilfield since soon after the Tertiary uplift[30].

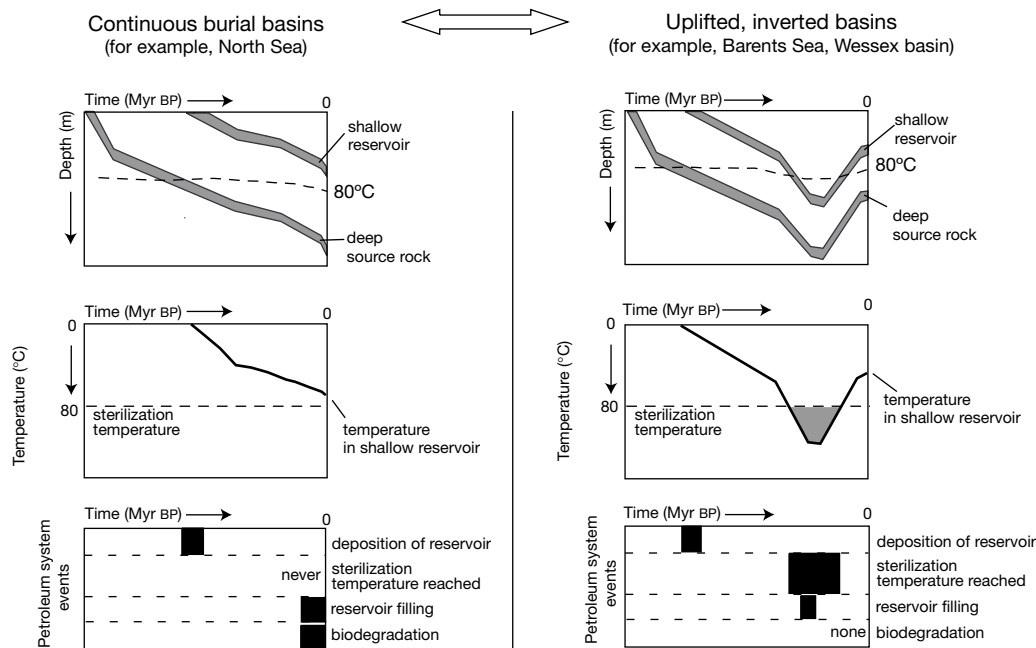Accordingly, reservoir sequences that have been exposed to

Figure 1 Comparison of continuously subsiding and uplifted sedimentary basins. The former are represented by the Viking graben, Tampen Spur and North Sea basins, and the latter by the Barents Sea and Wessex basins. Charts show schematic burial history (top), reservoir temperature history (middle), and petroleum system events (bottom), illustrating key differences with respect to biodegradation in petroleum reservoirs. Biodegradation of petroleum occurs only in sedimentary units that have not been exposed to temperatures exceeding 80 °C ('palaeosterilization') before oil charging.

temperatures in excess of ~90 °C before the current oil charge seem to be no longer capable of supporting petroleum biodegradation, even though the reservoir is far from chemical equilibrium and thus chemical energy to support life is present (that is, both hydrocarbons and potential oxidants including excess water[13] coexist). It seems most unlikely that all of the Triassic to Upper Jurassic fluvial to marine reservoir sediment systems of the Wessex and Barents Sea basins were devoid of hydrocarbon-degrading bacteria initially, because comparable reservoir settings in the Viking graben do have degraded oils. We suggest that the uplifted reservoirs do not contain degraded oils because the hydrocarbon-degrading bacteria of the reservoirs were inactivated by deep burial before the oil charge entered—'palaeosterilization'.

It now seems likely that most sediments cooler than the maximum temperature boundary of the deep biosphere contain microorganisms[3,7,31]. If 'palaeosterilization' occurs, it appears that once destroyed, effective bacterial and archaeal communities (or at the very least hydrocarbon-degrading bacterial and archaeal communities) are not re-established in deep reservoirs. This is true even in the Wessex and Barents Sea basins that had large surface exposure of terrains ostensibly suitable for recharge of meteoric waters, nutrients and microorganisms. These observations indicate that both surface recharge of fluids or organisms, and migration of organisms in the deep subsurface environment of petroleum basins, may be insignificant. Thus, in deep reservoirs where degradation does occur, microorganisms must have already been in place during slow burial, and have survived and probably evolved on geological timescales. This implies that the microbial flora of deep reservoirs may be isolated biospheres, descendants of ancient lineages responding in an isolated manner to evolution over millions of years, as suggested for other sediments[31].

Our reservoir palaeo-temperature estimates suggest that the base of the hydrocarbon-degrading biosphere is at temperatures lower than 90 °C, in agreement with much oilfield data and the empirical rule used by petroleum geologists that biodegraded oils are only found in reservoirs cooler than 80 °C. Our proposed 'palaeosterilization' model does not provide information before drilling about the degree of oil biodegradation to be expected in continuously subsiding basins. However, it does explain extensive, shallow occurrences of non-degraded oil in many uplifted reservoirs that are at equal or lower temperatures than those at which degraded oils are found in other basins (see Supplementary Information). This simple model, if further validated, implies that some very shallow reservoirs, which have been avoided because of the risk of petroleum biodegradation, would now represent much lower risk, low cost, exploration targets if they have been buried to temperatures in excess of ~80–90 °C. □

1. Roadifer, R. E. in *Exploration for Heavy Crude Oil and Natural Bitumen* (ed. Meyer, R. F.) 3–23 (AAPG Studies in Geology 25, American Association of Petroleum Geologists, Tulsa, 1987).
2. Connan, J. in *Advances in Petroleum Geochemistry* Vol. 1 (eds Brooks, J. & Welte, D. H.) 299–335 (Academic Press, London, 1984).
3. Parkes, R. J. *et al.* Deep bacterial biosphere in Pacific Ocean sediments. *Nature* **371,** 410–413 (1994).
4. Stetter, K. O., Fiala, G., Huber, G., Huber, R. & Segerer, A. Hyperthermophilic microorganisms. *FEMS Microbiol. Rev.* **75,** 117–124 (1990).
5. ZoBell, C. E. The role of bacteria in the formation and transformation of petroleum hydrocarbons. *Science* **102,** 364–369 (1945).
6. Jobson, A., Cook, F. D. & Westlake, D. W. S. Microbial utilization of crude oil. *Appl. Microbiol.* **23,** 1082–1089 (1972).
7. Pedersen, K. Exploration of deep intraterrestrial microbial life; current perspectives. *FEMS Microbiol. Lett.* **185,** 9–16 (2000).
8. Barnard, P. C. & Bastow, M. A. in *Petroleum Migration* (eds England, W. A. & Fleet, A. J.) 167–190 (Special Publication No. 59, The Geological Society, London, 1991).
9. Blöchl, E. *et al.* Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1,** 14–21 (1997).
10. Stetter, K. O. Hyperthermophilic prokaryotes. *FEMS Microbiol. Rev.* **18,** 149–158 (1996).
11. Murphy, E. M. & Schramke, J. A. Estimation of microbial respiration rates in groundwater by geochemical modelling constrained with stable isotopes. *Geochim. Cosmochim. Acta* **62,** 3395–3406 (1998).
12. Larter, S. R. *et al.* in *Conference Abstracts, Proc. GeoCanada 2000* (eds Bachu, S. and Esslinger, P.) Abstract 120, 1–4 (CSPG/GAC, Calgary, Alberta, 2000).
13. Zengler, K., Richnow, H. H., Rossella-Mora, R., Michaelis, W. & Widdel, F. Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401,** 266–269 (1999).
14. Chapelle, F. H. & Lovely, D. R. Rates of microbial-metabolism in deep coastal-plain aquifers. *Appl. Environ. Microbiol.* **56,** 1865–1874 (1990).
15. Daniel, R. M. & Cowan, D. A. Biomolecular stability and life at high temperatures. *Cell Mol. Life Sci.* **57,** 250–264 (2000).
16. Stetter, K. O. *et al.* Hyperthermophilic Archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365,** 743–745 (1993).
17. Takahata, Y., Nishijima, M., Hoaki, T. & Maruyama, T. Distribution and physiological characteristics of hyperthermophiles in the Kubiki oil reservoir in Niigata, Japan. *Appl. Environ. Microbiol.* **66,** 73–79 (2000).
18. L'Haridon, S., Reysenbach, A. L., Glénat, P., Prieur, D. & Jeanthon, C. Hot subterranean biosphere in a continental oil reservoir. *Nature* **377,** 223–224 (1995).
19. Grassia, G. S., McLean, K. M., Glénat, P., Bauld, J. & Sheehy, A. J. A systematic survey for thermophilic fermentative bacteria and archaea in high temperature petroleum reservoirs. *FEMS Microbiol. Ecol.* **21,** 47–58 (1996).
20. Nilsen, R. K., Beeder, J., Thorstenson, T. & Torsvik, T. Distribution of thermophilic sulfate reducers in North Sea oil field reservoir waters and oil reservoirs. *Appl. Environ. Microbiol.* **62,** 1793–1798 (1996).
21. Orphan, V. J., Taylor, L. T., Hafenbradl, D. & Delong, E. F. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **65,** 700–711 (1999).
22. Rueter, P. *et al.* Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature* **372,** 455–458 (1994).
23. Horstad, I. & Larter, S. R. Petroleum migration, alteration, and remigration within Troll Field, Norwegian North Sea. *Bull. Am. Assoc. Petrol. Geol.* **81,** 222–248 (1997).
24. Cornford, C. in *Introduction to the Petroleum Geology of the North Sea* (ed. Glennie, K. W.) 294–361 (Blackwell Scientific Publications, Oxford, 1990).
25. Linjordet, A. & Grung Olsen, R. in *Giant Oil and Gas Fields of the Decade 1978–1988* (ed. Halbouty, M. T.) 349–370 (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1992).
26. Cornford, C., Christie, O., Endresen, U., Jensen, P. & Myhr, M. B. Source rock and seep oil maturity in Dorset, southern England. *Org. Geochem.* **13,** 399–409 (1988).
27. Schroeder, F. W. & Sylta, Ø. in *Basin Modelling: Advances and Applications* (eds Doré, A. G. *et al.*) 469–484 (Vol. 3, NPF Special Publication, Elsevier, Amsterdam, 1993).
28. Theis, N. J., Nielsen, H. H., Sales, J. K. & Gail, G. J. in *Basin Modelling: Advances and Applications* (eds Doré, A. G. *et al.*) 433–444 (Vol. 3, NPF Special Publication, Elsevier, Amsterdam, 1993).
29. Bray, R. J., Duddy, I. R. & Green, P. F. in *Development, Evolution and Petroleum Geology of the Wessex Basin* (ed. Underhill, J. R.) 199–213 (Special Publication 133, Geological Society, London, 1998).
30. Evans, J., Jenkins, D. & Gluyas, J. in *Development, Evolution and Petroleum Geology of the Wessex Basin* (ed. Underhill, J. R.) 407–413 (Special Publication 133, Geological Society, London, 1998).
31. Frederickson, J. K. *et al.* Microbial community structure and biogeochemistry of Miocene subsurface sediments—implications for long-term microbial survival. *Mol. Ecol.* **4,** 619–626 (1995).
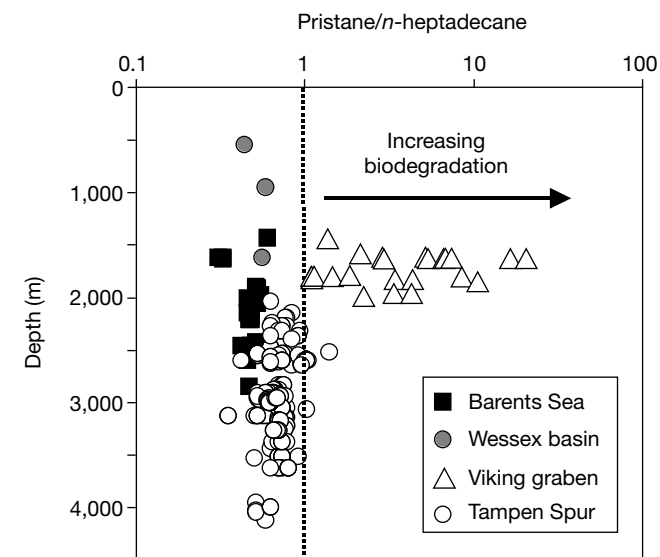

Figure 2 Degree of oil biodegradation, assessed by the pristane/*n*-heptadecane ratio, as a function of current reservoir depth for oils from four different settings. The data derive from Jurassic and Triassic reservoirs of the Wessex basin, Jurassic reservoirs of the Barents Sea, Palaeocene and Eocene reservoirs of the Viking graben and Jurassic reservoirs of the Tampen Spur and Horda platform, and were obtained by gas chromatographic analysis of whole oils or separated saturated hydrocarbon fractions. Values greater than ~1 represent biodegraded oils. The Viking graben and Tampen Spur oils (North Sea) are from continuously subsided reservoirs (empty symbols), whilst the Barents Sea and Wessex basin reservoirs were uplifted and recharged with oil after deeper burial (filled symbols).

Mineral disequilibrium in lavas explained by convective self-mixing in open magma chambers

S. Couch*, R. S. J. Sparks* & M. R. Carroll†

* *Department of Earth Sciences, Bristol University, Bristol BS8 1RJ, UK*
† *INFM and Dipartimento di Scienze della Terra, Università di Camerino, 62032 Camerino, Italy*

Characteristic features of many porphyritic andesite and dacite lavas are that they are rich in crystals and display a range of disequilibrium features, including reversely zoned crystals, resorption surfaces, wide ranges of mineral compositions and minerals which are not in equilibrium with the surrounding rock matrix. These features are often interpreted as evidence of the mixing of magmas of contrasting composition, temperature and origin[1,2]. Here, however, we propose that such features can also be caused by convection within a magma body with a single composition, that is heated from below and cooled from above. We describe petrological observations of andesite lava erupted at the Soufrière Hills volcano, Montserrat, which indicate a heating event and the intermingling of crystals that have very different thermal histories. We present experimental data on a representative groundmass composition of this lava, which indicate that it is difficult to explain the calcic compositions of plagioclase overgrowth rims and microphenocrysts unless parts of the magma were at temperatures much higher than the inferred average temperature. The concept of convective self-mixing allows us to explain the occurrence of compositions of minerals that apparently cannot coexist under equilibrium conditions.

Here we summarize key features of the petrology of the Soufrière Hills andesite[3,4], and present new experimental data which further constrains magma genesis. The lava contains 45–55% phenocrysts of plagioclase, hornblende, orthopyroxene, oxide and minor quartz, and microphenocrysts (crystals 100 to 300 μm in length) of plagioclase, clinopyroxene, orthopyroxene and oxide. Variable amounts of microlites (<100 μm in length) are set in a high-silica rhyolitic glass. There are abundant features of disequilibrium and mixing of crystals with contrasted thermal histories. Plagioclases in particular have complex and variable zoning patterns, as is typical of calc-alkaline andesites[5,6]. Some plagioclase phenocrysts have strongly resorbed sieve-textured zones and rims 30–150 μm in thickness of more calcic composition (65–80% anorthite; An_{65-80}) in comparison to cores (An_{48-58}). Some orthopyroxenes have thin overgrowth rims (up to 25 μm) with high Ca content, a feature consistent with crystallization from higher-temperature magma than the core[4]. Hornblendes range from euhedral, to slightly resorbed, to those completely pseudomorphed by coarse pyroxene–plagioclase–oxide intergrowths[3,4]. Quartz crystals range from those with resorbed outlines to those with clinopyroxene reaction rims. All variants of texture and mineral composition can be observed in a single thin section. Microphenocrysts of plagioclase vary widely in composition

(cores of An_{48-80}, rims of An_{52-75}) with most being more calcic than the phenocryst cores[4]. Groundmass plagioclase microlites are characteristically more sodic (An_{40-75}) than either overgrowth rims or microphenocrysts.

Figure 1a shows a phase diagram of a representative melt composition of the rhyolitic groundmass found in the andesitic Montserrat bulk composition[7]. The diagram, contoured in An content of plagioclase, illustrates the difficulties of reconciling the observed mineralogy and textures with a simple magmatic evolution. Core compositions of the phenocrysts and the overall stability of amphibole are consistent with magma temperatures in the 840 to 870 °C range at $p_{H_2O} = 120$ to 140 MPa (refs 9 and 7). Water contents of melt inclusions are estimated to be 4–5 wt%, consistent with water partial pressures of 115–130 MPa (ref. 7). The experimental data (Fig. 1) indicate that the overgrowth rims on plagioclase phenocrysts and microphenocrysts are too calcic to be formed at the inferred magma temperatures, despite the fact that the estimated groundmass composition is relatively calcium-rich. Calcic plagioclase compositions also cannot be explained by isothermal decompression, because plagioclase compositions become increasingly sodic at lower p_{H_2O} values. Thus, we conclude that calcic plagioclase overgrowth rims and microlites ($>An_{60}$) formed at substantially higher temperatures (900–1,000 °C) and at still elevated p_{H_2O} values (>100 MPa). These inferences are consistent with evidence of high-temperature crystallization, up to 970 °C, such as
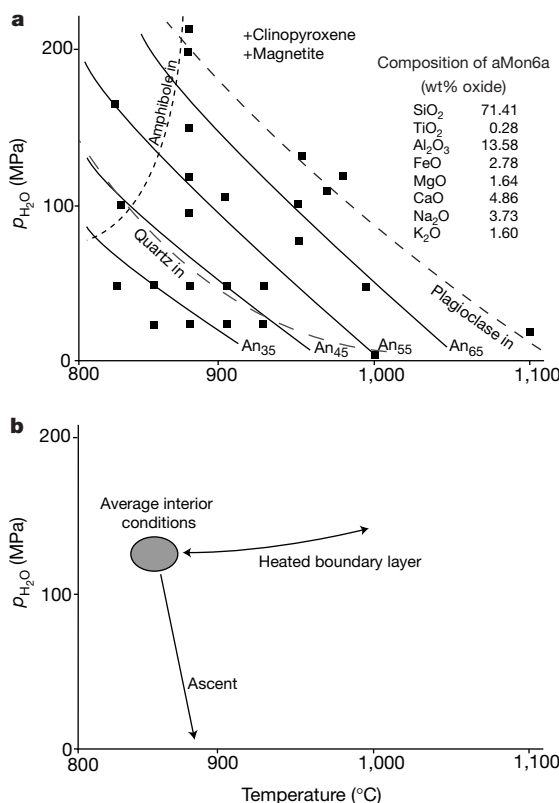

Figure 1 Pressure–temperature diagrams to show conditions of Soufrière Hills magma. **a**, Experimentally derived phase diagram of Montserrat groundmass (aMon6a). The composition was determined by averaging 200 rastered electron-probe analyses of the groundmass. Experiments were run in rapid-quench cold-seal pressure vessels and were water-saturated. Black squares denote experimental conditions; dashed lines denote the boundaries of given phases (plagioclase, quartz, amphibole); solid lines denote contours of plagioclase feldspar anorthite content. All experiments were saturated in clinopyroxene and magnetite. **b**, Schematic diagram of proposed pressure–temperature history of Montserrat magma before eruption. p_{H_2O}, partial pressure of water.