

## Contribution of hydrophobic interactions to protein stability

James T. Kellis Jr, Kerstin Nyberg, Daša Šali  
& Alan R. Fersht\*

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AZ, UK

A major factor in the folding of proteins is the burying of hydrophobic side chains. A specific example is the packing of  $\alpha$ -helices on  $\beta$ -sheets by interdigititation of nonpolar side chains. The contributions of these interactions to the energetics of protein stability may be measured by simple protein engineering experiments. We have used site-directed mutagenesis to truncate hydrophobic side chains at an  $\alpha$ -helix/ $\beta$ -sheet interface in the small ribonuclease from *Bacillus amyloliquefaciens* (barnase). The decreases in stability of the mutant proteins were measured by their susceptibility to urea denaturation. Creation of a cavity the size of a  $-\text{CH}_2-$  group destabilizes the enzyme by 1.1 kcal mol<sup>-1</sup>, and a cavity the size of three such groups by 4.0 kcal mol<sup>-1</sup>.

Protein design *de novo* requires a quantitative understanding of the rules for the folding of proteins (see ref. 1 for recent review). One approach is the direct application of computational methods. Important advances are being made as methods improve and computational power increases. A complementary approach is descriptive, surveying known structures and formulating general rules about motifs. An important example is the classification of packing of helices and sheets in proteins<sup>2–7</sup>. There is, however, no direct evidence on the strengths and specificity of the interactions involved and all estimates have been made from model compounds. Direct, simple experimental data on proteins are required in general to test and refine predictive methods.

Significant experimental contributions are now being made using protein engineering (see, for example, refs 8, 9). Work so far has primarily used the selection of temperature-sensitive mutants after random mutagenesis and the construction of large numbers of mutations at the temperature-sensitive loci. We are using an alternative strategy: adopting methods previously employed to dissect the structure and activity of the tyrosyl-tRNA synthetase<sup>10</sup> to provide a quantitative experimental basis to the descriptive approach of refs 2–7. Mutations that remove defined interactions without introducing new or unfavourable interactions, that is, nondisruptive deletions<sup>11</sup> are made on the basis of the crystal structure of the wild-type enzyme. It is especially important to use nondisruptive deletions in analysing protein folding because the free energy of unfolding depends on the difference in free energy between the folded and denatured states and so both should be perturbed as little as possible. Empirical measurements of the energetics of interactions are obtained that may be used directly for experimental design and provide raw data for refinement of computational methods and theory.

The enzyme we have chosen for these studies is barnase, the RNase from *B. amyloliquefaciens*<sup>12</sup>. This enzyme has almost ideal characteristics as a paradigm for protein folding for the following reasons. It is a monomer of 110 residues of  $M_r$  12,382, one of the smallest known enzymes. The crystal structure has been solved at high resolution<sup>13</sup>; the small size and favourable crystals enabling the resolution to be increased to 1.4 Å. In addition, barnase is small enough for sequence specific two-dimensional NMR assignments to be made and to be tractable to analysis by computational methods. Yet it is large enough to have significant secondary structure. It has, for example, the packing of an  $\alpha$ -helix on a  $\beta$ -sheet<sup>13</sup>: the helix formed from

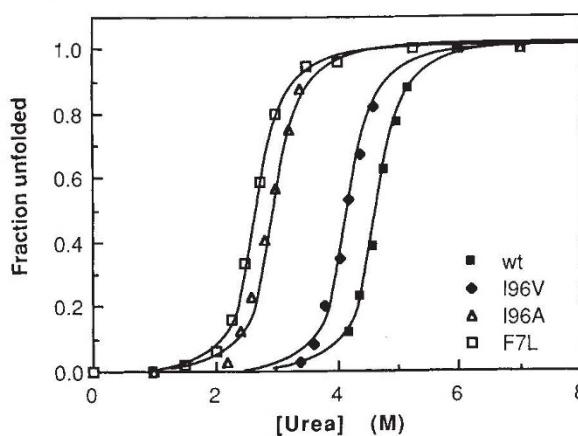


Fig. 1 Urea denaturation of recombinant barnase and engineered mutants. Denaturation was observed spectroscopically using difference spectroscopy at 286 and 270 nm. Experiments were performed at 25 °C in a buffer containing 19.3 mM MES acid and 30.7 mM base (MES = 2-(*N*-morpholino)ethanesulphonic acid) at ionic strength 0.05 M and pH 6.30 (except for Phe → Leu 7 which was at pH 6.15). The stability of the protein hardly varies between pH 6.0 and 6.5 under these conditions. The concentration of barnase was between 9 and 12 µM. There is a decrease of 13.5% in absorbance at 286 nm and an increase of 4.5% at 270 nm on unfolding ( $\Delta A_{286} - \Delta A_{270}$ ) was measured since this both amplifies the signal and minimizes dilution errors as  $A_{286} \sim A_{270}$ . The solid curves are just for visual aid.

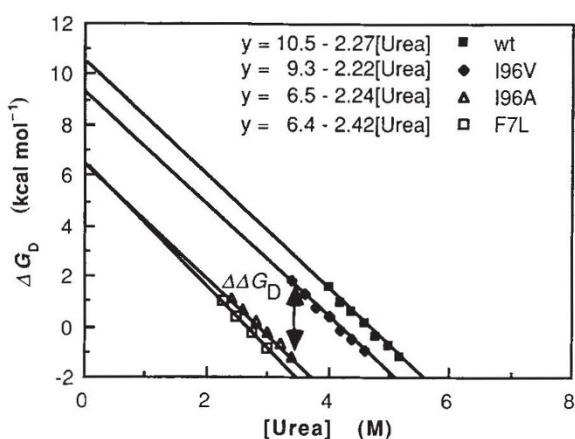
**Methods.** Wild-type enzyme was secreted from *E. coli* harboring the plasmid pMT410, a derivative of pUC9 (ref. 16 and R. W. Hartley, personal communication). This construction contains the structural gene for barnase fused to the alkaline phosphatase promoter and signal sequence and the gene for barstar (the polypeptide inhibitor of barnase) under the control of its own promoter contained on a 1.4 kilobase (kb) EcoRI-HindIII fragment. For mutagenesis, this fragment was subcloned into M13. The mutated genes were subsequently cloned back into pUC9 for expression. Site-directed mutagenesis was carried out using the double primer method<sup>20</sup>. The mutagenic primers were used together with an M13 universal primer. To improve the yield of mutants, a strain defective in mismatch repair, *E. coli* BMH 71–18 mut L (ref. 27) was used in the transfection. Mutant bacteriophages were identified by oligonucleotide hybridization<sup>28</sup> and the mutations verified by dideoxy sequencing of the entire genome. The following primers were used to direct the mutations: 5'TTTTGTAG\*AC\*CAGCCAG3', Ile → Val96; 5'TTTTGTAAG\*C\*CAGCCAG3' Ile → Ala96; Phe → Leu7 5'CCCCGTCAAG\*CGTGTG3' (asterisks follow mismatches). Cells were grown in a low phosphate medium<sup>29</sup>. Secreted wild-type and mutant enzymes were purified to homogeneity (D.

Mossakowska, K.N. and A.R.F., in preparation).

residues 6–18 packs against the anti-parallel  $\beta$ -sheet formed from residues 50–55, 70–75, 85–91, 94–101 and 106–108. The protein is composed of one domain and undergoes a reversible thermal or urea-induced denaturation, approximating to a two-state process<sup>14,15</sup>. This enables thermodynamic measurements to be made on the factors that influence protein folding and stability. Equilibrium and kinetic measurements of denaturation may readily be made by difference spectroscopy. Barnase has no disulphide bridges. This reduces structural constraints in the denatured protein and so decreases complications in analysing denaturation. The protein has been cloned and expressed<sup>16</sup>.

The mutations chosen in this study are the truncations Ile96 → Val, Ile96 → Ala and Phe7 → Leu. The first two mutations are, in theory, true nondisruptive deletions since simple truncations are involved. Phe → Leu is a conventional conservative mutation, but there is a change in geometry about the  $\gamma$ -carbon ( $sp^2$  in Phe and  $sp^3$  in Leu). The mutations remove hydrophobic inter-

\* To whom correspondence should be addressed.



**Fig. 2** Linear dependence of  $\Delta G_D$  on [urea] over experimental ranges allows simple determination of  $\Delta G_{D,0}^{H_2O}$  and  $\Delta\Delta G_D$ . The value of  $\Delta\Delta G_D$  between, for example, Ile  $\rightarrow$  Val96 and Ile  $\rightarrow$  Ala96, may readily be read off the graph at an intermediate concentration of urea (such as 3.55 M) to be 2.9 kcal mol $^{-1}$ . Even more precise data are obtained when the two denaturation curves overlap as there is no extrapolation and values of  $\Delta G_D$  for each protein can be made under identical conditions (for example, wild type and Ile  $\rightarrow$  Val96).

actions where the helix composed of residues 6 to 18 docks on to the  $\beta$ -sheet, effectively introducing cavities into the enzyme. The change in stability on mutation is then measured by the change in free energy of the reversible denaturation of the enzyme, determined by UV difference spectroscopy. We have chosen urea-induced denaturation rather than thermally induced denaturation for two reasons. First, it is not a simple matter to apply thermal denaturation for quantitative comparison of different proteins with significantly different values of  $T_m$ . The enthalpy of unfolding of proteins varies with temperature and the van't Hoff plots are not linear $^{17}$ . Thermodynamic quantities are consequently not easily extrapolated to different temperatures. Urea denaturation, however, may be performed at a single temperature for a whole series of mutants (for example, 25 °C as in Fig. 1) and analysed as below $^{18}$ . Second, urea denaturation is thought to give a more completely unfolded protein than does thermal denaturation $^{17,18}$ .

Wild-type and engineered mutants were expressed in *Escherichia coli*. The yield of mutant enzymes decreases as the mutations become more radical, although specific activity towards hydrolysis of RNA drops by no more than 30%. Mutants Ile96  $\rightarrow$  Ala and Phe7  $\rightarrow$  Leu are expressed at levels of only some 10% of wild type. The decrease in yield presumably reflects a greater susceptibility to proteolytic degradation as the folded structure becomes less stable. Urea denaturation confirms that mutation lowers stability. It can be seen from Fig. 1 that the concentration of urea required to unfold 50% of the protein is 4.63 M for wild-type barnase, 4.19 M for Ile96  $\rightarrow$  Val, 2.90 M for Ile96  $\rightarrow$  Ala and 2.66 M for Phe7  $\rightarrow$  Leu at pH 6.3.

The free energy of unfolding in the absence of urea ( $\Delta G_{D,0}^{H_2O}$ ) is conventionally estimated by extrapolating the free energy of unfolding at each individual concentration of urea ( $\Delta G_D$ ) to zero concentration assuming that they are linearly related (equation (1), ref. 18). This procedure gives  $10.5 \pm 0.2$  kcal mol $^{-1}$

$$\Delta G_D = \Delta G_{D,0}^{H_2O} - m[\text{urea}], \quad (1)$$

for  $\Delta G_{D,0}^{H_2O}$  for wild-type barnase,  $9.3 \pm 0.2$  kcal mol $^{-1}$  for Ile96  $\rightarrow$  Val,  $6.5 \pm 0.1$  kcal mol $^{-1}$  for Ile96  $\rightarrow$  Ala and  $6.44 \pm 0.04$  kcal mol $^{-1}$  for Phe7  $\rightarrow$  Leu. (About 2 parts in  $10^5$  of the latter two mutants are unfolded in the absence of urea.) The slopes of the plots are quite similar for wild type and mutants

( $m = 2.27 \pm 0.04$ ;  $2.22 \pm 0.05$ ,  $2.24 \pm 0.04$  and  $2.42 \pm 0.01$ , respectively). Wild-type barnase is significantly more stable than T1 ribonuclease ( $\Delta G_{D,0}^{H_2O} = 4$  kcal mol $^{-1}$ , ref. 18) and staphylococcal nuclease ( $\Delta G_{D,0}^{H_2O} = 5.5$  kcal mol $^{-1}$ , ref. 8), so allowing more radically destabilizing mutations to be made and analysed.

While the above procedure is adequate for estimating the approximate value of  $\Delta G_{D,0}^{H_2O}$ , the extrapolations are over too wide a range of concentrations of urea to measure reliably small differences in stability of mutants (errors in the slope are magnified and the assumption about precise linearity between free energy of unfolding and [urea] is questionable, although the slope  $m$  is reasonably constant between 2.5 and 5 M urea). We have used two alternative procedures for measuring differences more directly. The first is the better of the two and makes no assumptions about the dependence of unfolding energy on urea concentration. Wild-type and mutant enzymes of suitably similar stability are denatured in parallel at the same concentrations of urea. Then, at each concentration of urea, the difference in  $\Delta G_D$  between wild-type and mutant ( $= \Delta\Delta G_D = \Delta G_{D,(\text{wt})} - \Delta G_{D,(\text{mut})}$ ) is given directly by:

$$\Delta\Delta G_D = RT \ln \frac{\frac{[\text{unfolded}]}{[\text{folded}]}_{\text{wt}}}{\frac{[\text{unfolded}]}{[\text{folded}]}_{\text{mut}}} \quad (2)$$

This procedure clearly works well for small differences in stability where denaturation curves overlap and the ratio of [folded]/[unfolded] for wild-type and for mutant can be measured under the same conditions. This cannot be applied to situations where there are large differences in stability between two mutants because the ratio of [folded]/[unfolded] can be measured accurately only in the range 0.1 to 10. Consequently, the ratio will be too extreme for an unstable mutant when the ratio is measurable for wild-type. Eventually, the procedure will be scaled over a series of mutants, each one differing from the next in stability by less than a kcal mol $^{-1}$ . Until then, the second procedure may be applied: a small extrapolation is made to a concentration of urea intermediate between the values required for denaturation of wild type and mutant (or between two mutants, c.f. ref. 19, Fig. 2). This assumes linearity in the dependence of  $\Delta G_D$  on [urea] only over a small range beyond the experimental data. Values of  $\Delta\Delta G_D$  are listed in Table 1. Truncating Ile to Val destabilizes wild type by 1.1 kcal mol $^{-1}$ , and Ile to Ala by 4.0 kcal mol $^{-1}$ . Mutation of Phe to Leu destabilizes by 4.6 kcal mol $^{-1}$ . The experiments thus give direct measure-

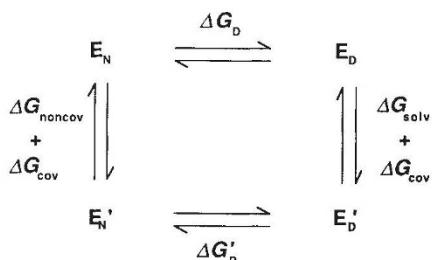
Table 1

Mutant	$\Delta\Delta G_D^*$ (in urea)	$\Delta G_{D,0}^{H_2O} \dagger$ (at 0 M urea) (kcal mol $^{-1}$ )	$\Delta G_{\text{solv}} \ddagger$
Ile $\rightarrow$ Val96	1.1	1.2	-0.12
Ile $\rightarrow$ Ala96	4.0	4.0	-0.15
Phe $\rightarrow$ Leu7	4.6	4.1	2.24

\*  $\Delta G_{D,(\text{wt})} - \Delta G_{D,(\text{mut})}$ , calculated by direct comparison at the same concentrations of urea (wild type and Ile  $\rightarrow$  Val96) or interpolation to an intermediate concentration of urea (Ile  $\rightarrow$  Val96 and Ile  $\rightarrow$  Ala96 at 3.5 M urea, Ile  $\rightarrow$  Ala96 and Phe  $\rightarrow$  Leu7 at 2.8 M urea, see Fig. 2 and text). The similarities in the slopes of the denaturation curves renders the interpolation insensitive to the choice of [urea] within  $\pm 1$  M.

$\dagger \Delta G_{D,0}^{H_2O} - \Delta G_{D,(\text{wt})}$ , from extrapolation to [urea] = 0 M (equation (1)). Note that there is expected to be a small discrepancy between this and the previous column because of the different stabilities of the exposed side chains in the denatured state when exposed to water and urea. For complete exposure of the side chain, this is negligible for Ile  $\rightarrow$  Val, +0.15 kcal mol $^{-1}$  for Ile  $\rightarrow$  Ala and is +0.18 kcal mol $^{-1}$  for Phe  $\rightarrow$  Leu at 4 M urea (the larger side chains being more stable in urea) $^{25}$ .

$\ddagger$  Data from ref. 23 for the change in free energy of solvation of side chain ( $G_{\text{solv}}(\text{mut}) - G_{\text{solv}}(\text{wt})$ ).



**Fig. 3** Thermodynamic cycle relating the unfolding of wild-type (E) and mutant (E') enzymes. On mutation, there is a change in the free energy of the protein ( $\Delta G_{\text{cov}}$ ) because of covalent changes such as the change in chemical bonds on going from Ile to Ala).  $\Delta G_{\text{cov}}$  is the same for both the folded and unfolded states and so cancels out when comparing the right and left hand sides of the cycle. There are also changes in the noncovalent interactions,  $\Delta G_{\text{noncov}} (= G_{E_N} - G_{E'_N})$ , on mutation. In the unfolded enzyme, the side chains of the amino acids are more, if not completely exposed to solvent and so  $\Delta G_{\text{noncov}(D)}$  represents predominantly the changes in the solvation energy of the side chain,  $\Delta G_{\text{solv}}$  ( $= G_{E'_D} - G_{E_D}$ ), on mutation.  $\Delta G_{\text{noncov}}$  for the folded state represents predominantly the loss of interaction energy between the enzyme and those parts of the side chain that are deleted on mutation.  $\Delta G_{\text{noncov}}$  contains, in addition, the energy terms associated with any reorganization of enzyme and associated solvent on mutation. The experimentally determined quantity in our experiments is the difference in unfolding energy of the wild-type and mutant enzymes,  $\Delta G_D - \Delta G'_D$ . It is seen from the cycle that  $\Delta G_D - \Delta G'_D = \Delta G_{\text{noncov}} - \Delta G_{\text{solv}}$ .

ments of the destabilizing effects of cavities in enzymes. This points to a way of stabilizing protein structures: make substitutions to fill in holes that occur in the native structure.

Calculation of the free energy of folding (or unfolding) of proteins *ab initio* involves determining the free energies of the folded and unfolded states (Fig. 3,  $\Delta G_D = G_{E_D} - G_{E_N}$ , where  $E_N$  and  $E_D$  are the native and denatured forms of the enzyme). This requires calculating all the noncovalent interaction energies in the folded and unfolded proteins and the solvation energies. One fundamental unknown quantity is the noncovalent interaction energy in the unfolded state. This is probably not a single state but consists of a mixture of many different states of similar energies, in which the exposure of side chains to solvent is unknown and variable. This presents a stumbling block to calculation. It is important to relate the empirical data from the protein engineering experiments,  $\Delta\Delta G_D$ , ( $= \Delta G_D - \Delta G'_D$ ) to the energies used in calculations. This is done by a simple thermodynamic cycle (Fig. 3) which considers the native and denatured states for wild-type ( $E_N$  and  $E_D$ ) and mutant ( $E'_N$  and  $E'_D$ ) enzymes. In Fig. 3, the change in the noncovalent interaction energy in the native enzyme on removal of a side chain is defined by  $\Delta G_{\text{noncov}} = G_{\text{noncov(mut)}} - G_{\text{noncov(wt)}}$  (where the subscripts refer to the noncovalent interactions in the folded mutant and wild-type enzymes). The equivalent change in noncovalent energy in the denatured state is termed  $\Delta G_{\text{solv}}$  from Fig. 3:

$$\Delta G_D^{\text{H}_2\text{O}} - \Delta G'_D^{\text{H}_2\text{O}} = \Delta G_{\text{noncov}} - \Delta G_{\text{solv}} \quad (3)$$

First, it can be seen from equation (3) that the changes in free energy of unfolding are a function of the interactions in folded and denatured states. Second, equation (3) may be used to gather information about interactions in the denatured state as follows.  $\Delta G_{\text{noncov}}$  may be calculated from the native structures of the wild-type and mutant enzymes using conventional computational methods. Recent developments in computational methods allow the direct calculation of the values of  $\Delta G_{\text{noncov}}$  (and  $\Delta G_{\text{solv}}$ ) by the mathematical equivalent of mutating the side chains in slow increments<sup>20-22</sup>. If the side chain in the unfolded state is completely exposed to solvent, then  $\Delta G_{\text{noncov}}$  is simply calculated from equation (3) using relative solvation energies of the amino acid side chains that are available from

experiments on model compounds<sup>23,24</sup>. For example,  $\Delta G_{\text{solv}}$  for Ile  $\rightarrow$  Val is  $-0.12$ , Ile  $\rightarrow$  Ala is  $-0.15$  and Phe  $\rightarrow$  Leu is  $2.24$  kcal mol $^{-1}$  (ref. 23). Thus, if the side chain of Phe7 in denatured wild-type enzyme is completely exposed to solvent and so also is Leu7 in the mutant, then  $\Delta G_{\text{noncov}}$  for Phe  $\rightarrow$  Leu would be  $4.6 + 2.2 = 6.8$  kcal mol $^{-1}$ . Discrepancies between calculated and measured values of  $\Delta G_{\text{noncov}}$  will give clues about side chain interactions in the unfolded states. Further mutational experiments will provide additional direct data on protein stability and more general information on interactions in native and denatured states.

We thank Dr R. W. Hartley for his generous gifts of DNA constructs and encouragement. This work was supported by the MRC and EEC (BAP). D.S. is supported by funds from The Biochemical Society and IJS.

Received 20 April; accepted 11 May 1988.

1. Janicke, R. *Prog. Biophys. molec. Biol.* **49**, 117-237 (1987).
2. Levitt, M. F. & Chothia, C. *Nature* **261**, 552-558 (1976).
3. Janin, J. & Chothia, C. *J. mol. Biol.* **143**, 95-128 (1980).
4. Chothia, C., Levitt, M. & Richardson, D. *J. molec. Biol.* **145**, 215-250 (1981).
5. Chou, K.-C., Nemethy, G., Rumsey, S., Tuttle, R. W. & Scheraga, H. A. *J. molec. Biol.* **186**, 591-609 (1985); **188**, 641-649 (1987).
6. Sternberg, M. J. E. in *Computing in Biological Science* (ed. Geisow, M. J. & Barrett, A. N.) 143-177 (1983).
7. Chothia, C. *Ann. Rev. Biochem.* **53**, 537-572 (1984).
8. Shortle, D., Meeker, A. K. *Proteins* **1**, 81-89 (1986).
9. Albert, T. *et al.* *Nature* **330**, 41-45 (1987).
10. Fersht, A. R. *Biochemistry* **26**, 8031-8037 (1987).
11. Fersht, A. R., Leatherbarrow, R. J. and Wells, T. N. C. *Biochemistry* **26**, 6030-6038 (1987).
12. Nishimura, S. & Nomura, M. *Biochim. biophys Acta* **30**, 430-433 (1958).
13. Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C. & Jack, A. *Nature* **297**, 162-164 (1982).
14. Hartley, R. W. *Biochemistry* **7**, 2401-2408 (1968).
15. Hartley, R. W. *Biochemistry* **8**, 2929-2932 (1969).
16. Paddon, C. J. & Hartley, R. W. *Gene* **53**, 11-19 (1987).
17. Creighton, T. E. *Proteins* (Freeman, New York 1983).
18. Pace, C. N. *Meth. Enzym.* **131**, 266-279 (1986).
19. Cupo, J. F. & Pace, C. N. *Biochemistry* **22**, 2654-2658 (1983).
20. Postma, J. P. M., Berendes, H. J. C. & Haak, J. R. *Faraday Symp. Chem. Soc.* **17**, 55-67 (1982).
21. Wong, C. F. & McCammon, J. A. *J. Am. chem. Soc.* **108**, 3830 (1986).
22. Bash, P. A., Singh, U. C., Langridge, R. & Kollman, P. A. *Science* **236**, 564-568 (1987).
23. Wolfenden, R. L., Andersson, L., Cullis, P. M. & Southgate, C. C. B. *Biochemistry* **20**, 849-855 (1981).
24. Ben Naim, A. & Marcus, Y. *J. chem. Phys.* **81**, 2106-2027.
25. Nozaki, Y. & Tanford, C. *J. biol. Chem.* **238**, 4074-4081 (1963).
26. Zoller, M. J. & Smith, M. *DNA* **3**, 479-488 (1984).
27. Kramer, B., Kramer, V. & Fritz, H.-J. *Cell* **38**, 879-887 (1984).
28. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. *Cell* **38**, 835-840 (1984).
29. Serpersku, E. H., Shortle, D. & Mildvan, A. S. *Biochemistry* **25**, 68-87 (1986).

## Errata

### A hypothetical model of the foreign antigen binding site of Class II histocompatibility molecules

J. H. Brown *et al.*

*Nature* **332**, 845-850 (1988).

In Fig. 3a, the I-A<sup>d</sup> sequence numbers 10, 12, 14 should be replaced by the numbers 11, 13, 15 respectively; the word "first" in the seventh line, fourth paragraph of page 846 should be deleted.

### Palaeoenvironmental trends in the history of trace fossils

D. J. Bottjer, M. L. Droser & D. Jablonski

*Nature* **333**, 252-255 (1988).

On page 253, left column, fourth paragraph down, a line of text is missing between the seventh and eighth lines. This line reads "features, and, most important to our study, avoid the circular".

### A cellular analogue of visual cortical plasticity

Y. Frégnac, D. Shultz, S. Thorpe & E. Bienenstock

*Nature* **333**, 367-370 (1988).

The bottom line shown in Fig. 2 should be read in the reverse order, "LEFT RIGHT", instead of "RIGHT LEFT", such as that peri-stimulus histograms presented in the left column correspond to stimulation through the left eye ( $S^+$ ), and those presented in the right column correspond to stimulation through the right eye ( $S^0$ ).