and trimethylamine dehydrogenase[19].

Our results show that structural diversity prevails within the NADP-dependent enzyme family, even when function is closely related. On the other hand, it emphasizes the versatility of the α/β-barrel scaffold, which often appears even in functionally unrelated proteins. Indeed, it is becoming evident that the number of stable folds used to achieve biological diversity is limited.

Preliminary clinical studies strongly support the value of aldose reductase inhibitors in the treatment of diabetic complications[2]. Knowledge of the three-dimensional structure of aldose reductase will enable more specific drugs to be designed so that therapy can eventually be improved. □

Received 7 August; accepted 6 November 1991.

1. Burg, M. B. *Kidney Int.* **33**, 635-641 (1988).
2. Dvornik, D. in *Aldose Reductase Inhibition. An Approach to the Prevention of Diabetic Complications* (McGraw-Hill, New York, 1987).
3. Banner, D. W. *et al. Nature* **255**, 609-614 (1975).
4. Rossmann, M. G., Moras, D. & Olsen, K. W. *Nature* **250**, 194-199 (1974).
5. Wermuth, B. *Prog. clin. biol. Res.* **174**, 209-230 (1985).
6. Bohren, K. M., Bullock, B., Wermuth, B. & Gabbay, K. H. *J. biol. Chem.* **264**, 9547-9551 (1989).
7. Chung, S. & LaMendola, J. *J. biol. Chem.* **264**, 14775-14777 (1989).
8. Nishimura, C., Wistow, G. & Carper, D. *Prog. clin. biol. Res.* **290**, 211-220 (1989).
9. Schade, S. Z. *et al. J. biol. Chem.* **265**, 3628-3635 (1990).
10. Watanabe, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 11-15 (1988).
11. Anderson, S. *et al. Science* **230**, 144-149 (1985).
12. Fujii, Y. *et al. J. biol. Chem.* **265**, 9914-9923 (1990).
13. Winters, C. J., Molowa, D. T. & Guzelian, P. S. *Biochemistry* **29**, 1080-1087 (1990).
14. Oechsner, U., Magdolen, V. & Bandlow, W. *FEBS Lett.* **238**, 123-128 (1988).
15. Wierenga, R. K., Drenth, J. & Schulz, G. E. *J. molec. Biol.* **167**, 725-739 (1983).
16. Bystroff, C., Oatley, S. J. & Kraut, J. *Biochemistry* **29**, 3263-3277 (1990).
17. Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. *Biochemistry* **24**, 1346-1357 (1985).
18. Farber, G. K. & Petsko, G. A. *Trends biol. Sci.* **15**, 228-234 (1990).
19. Lindqvist, Y., Brändén, C.-I., Mathews, F. S. & Lederer, F. *J. biol. Chem.* **266**, 3198-3207 (1991).
20. Rondeau, J. M. *et al. J. molec. Biol.* **195**, 945-948 (1987).
21. Kabsch, W. *J. appl. Crystallogr.* **21**, 916-924 (1988).
22. Terwilliger, T. C. & Kim, S.-H. *Acta crystallogr.* A**43**, 1-5 (1987).
23. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* A**39**, 813-817 (1983).
24. Jones, T. A. in *Computational Crystallography* 303-317 (Clarendon, Oxford, 1982).
25. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **35**, 458-460 (1987).
26. Priestle, J. *J. appl. Crystallogr.* **21**, 572-576 (1988).
27. Grimshaw, C. E. *et al. Biochemistry* **28**, 5343-5353 (1989).
28. Vander Jagt, D. L., Robinson, B., Taylor, K. K. & Hunsaker, L. A. *J. biol. Chem.* **265**, 20982-20987 (1990).
29. Del Corso, A. *et al. Arch. biochem. biophys.* **270**, 604-610 (1989).
30. Del Corso, A. *et al. Arch. biochem. biophys.* **283**, 512-518 (1990).

ACKNOWLEDGEMENTS. We thank A. Van Dorsselaer and M. Jaquinod for unpublished sequence data, B. Rees for his contribution to the phase problem and Y. S. Babu (of Biocryst) for discussions.

Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures

Axel T. Brünger

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, 06511, USA

THE determination of macromolecular structure by crystallography involves fitting atomic models to the observed diffraction data[1]. The traditional measure of the quality of this fit, and presumably the accuracy of the model, is the *R* value. Despite stereochemical restraints[2], it is possible to overfit or 'misfit' the diffraction data: an incorrect model can be refined to fairly good *R* values as several recent examples have shown[3]. Here I propose a reliable and unbiased indicator of the accuracy of such models. By analogy with the cross-validation method[4,5] of testing statistical models I define a statistical quantity (R_T^{free}) that measures the agreement between observed and computed structure factor amplitudes for a 'test' set of reflections that is omitted in the modelling and refinement process. As examples show, there is a high correlation between R_T^{free} and the accuracy of the atomic model phases. This is useful because experimental phase information is usually inaccurate, incomplete or unavailable. I expect that R_T^{free} will provide a measure of the information content of recently proposed models of thermal motion and disorder[6–8], time-averaging[9] and bulk solvent[10].

The most common measure for the quality of a crystal structure is the R value[11],

$$R = \frac{\sum_{h,k,l} \| F_{obs}(h,k,l) - |k| F_{calc}(h,k,l) \|}{\sum_{h,k,l} |F_{obs}(h,k,l)|} \quad (1)$$

where h, k, l are the reciprocal lattice points of the crystal, $|F_{obs}(h,k,l)|$ and $|F_{calc}(h,k,l)|$ are the observed and calculated structure factor amplitudes, respectively. R is closely related to the crystallographic residual[11]

$$R' = \sum_{h,k,l} (|F_{obs}(h,k,l)| - k|F_{calc}(h,k,l)|)^2 \quad (2)$$

which is a linear function of the negative logarithm of the likelihood of the atomic model assuming that all observations are independent and normally distributed[12]. R can be made arbitrarily small by increasing the number of model parameters and subsequent refinement against R' (ref. 13), that is the diffraction data can be overfit without improvement or even worsening of the information content of the atomic model.

Crystallographic diffraction data are redundant to some degree, for example, refinement of the penicillopepsin crystal structure from *Penicillium janthinellum*[14,15] at 1.8 Å resolution with 50% of the diffraction data randomly omitted only results in a 0.3 Å root-mean-square (r.m.s.) difference to the atomic structure refined against the full data set (Fig. 1). In analogy to cross-validation[4,5] I thus propose to partition a unique set of the observed reflections into a 'test' set T and a 'working' set A, that is, T and A are disjoint and their conjunction is the full

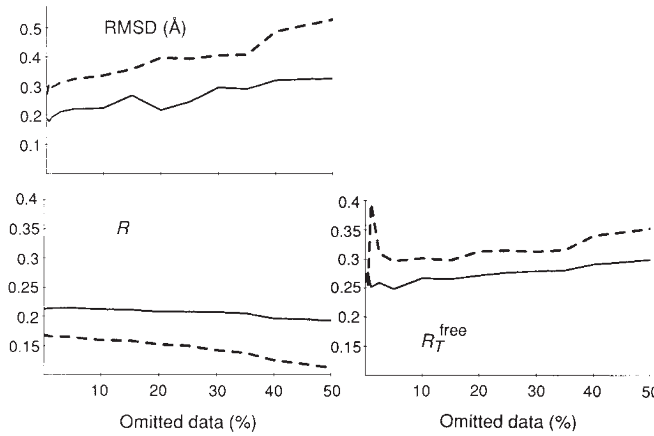

FIG. 1 SA-refinements of penicillopepsin[14,15] at 6–1.8 (solid lines) and 6–2.8 Å (dashed lines) resolution as a function of the percentage of omitted data. T was obtained by random selection from a unique set of all observed reflections. R is computed for the A set of reflections. The r.m.s. differences (RMSD) are computed between the structures refined against A and a unique set of all observed reflections. The RMSD is unequal to zero for 0% of the data omitted; this reflects the r.m.s. difference between two independently refined structures[25]. The penicillopepsin crystal structure[14,15] without water molecules and unit occupancy values was used as the starting point. Each refinement consisted of a slow-cooling protocol[26] using the program X-PLOR[16,27] starting at 1,000 K, overall B-factor refinement, and restrained individual B-factor refinement with the target values for the temperature factor deviations[2] of 1.5, 2, 2, 2.5 for bonded backbone, angle-related backbone, bonded sidechain, and angle-related sidechain atoms, respectively.

set of observed reflections. I refer to

$$R_T^{\text{free}} = \frac{\sum_{(h,k,l)\in T} \| F_{\text{obs}}(h,k,l)| - k|F_{\text{calc}}(h,k,l)\|}{\sum_{(h,k,l)\in T} |F_{\text{obs}}(h,k,l)|} \qquad (3)$$

as the free R value computed for the T set of reflections. T is omitted in the modelling process, for example in the case of crystallographic refinement[2] the residual to be minimized is given by

$$R'_A = \sum_{(h,k,l)\in A} (|F_{\text{obs}}(h,k,l)| - k|F_{\text{calc}}(h,k,l)|)^2. \qquad (4)$$

One would expect that R_T^{free} is less prone to overfitting than R. This concept can be applied to other statistical quantities, such as the standard linear correlation coefficient[11]. It can even be applied to crystal structures which have already been refined with all diffraction data included: refinement by simulated annealing (SA)[16] with T omitted will remove some of the memory towards T.

R_T^{free} reflects the information content of the atomic model. Suppose both the atomic model and diffraction data are perfect, resulting in $R = 0$. Refinement against A as opposed to all data will not change the atomic model and thus $R_T^{\text{free}} = 0$. Suppose the data contain small errors and an atomic model is overfit to a very low R value by introducing a large number of free parameters. As the noise is independent among different reflections, overfitting against A will not bias R_T^{free}. A similar argument applies to the case of partially incomplete or incorrect atomic models where the agreement with the diffraction data is improved by fitting noise.

The enhanced sensitivity of R_T^{free} with respect to model errors is illustrated in Fig. 2 which compares a portion of the correct[17] and incorrect[18] crystal structures of the plant ribulose-1,5-biphosphate carboxylase oxygenase (RuBisCO). Although the R difference between the correct and incorrect model is only 4% for comparable geometry, the R_T^{free} difference is 13%, suggesting that the incorrect model had been overfit. This is corroborated by the electron density maps in Fig. 2 which show a poorer agreement for the incorrect model. An alternative way to detect the errors in the RuBisCO crystal structure is provided by computing 'omit maps' with simulated annealing (A. Hodel, D. Eisenberg, S.-H. Kim and A.T.B., manuscript in preparation) which essentially is the real-space analogue to R_T^{free}.

Both R_T^{free} and the r.m.s. difference between the model refined against the complete data set and against A increase more or less monotonically as a function of the percentage of omitted data (Fig. 1). This is to be expected for terms that truly monitor the validity of a model. R decreases, which is paradoxical and misleading behaviour for an indicator of the models accuracy. As a compromise between avoiding fluctuations of R_T^{free} and maintaining small r.m.s. differences between refined models, I suggest T is obtained from a random selection of 10% of the observed reflections. The definition of R_T^{free} implies $R_T^{\text{free}} > R$; the difference between R_T^{free} and R is uniformly distributed as a function of resolution (not shown).

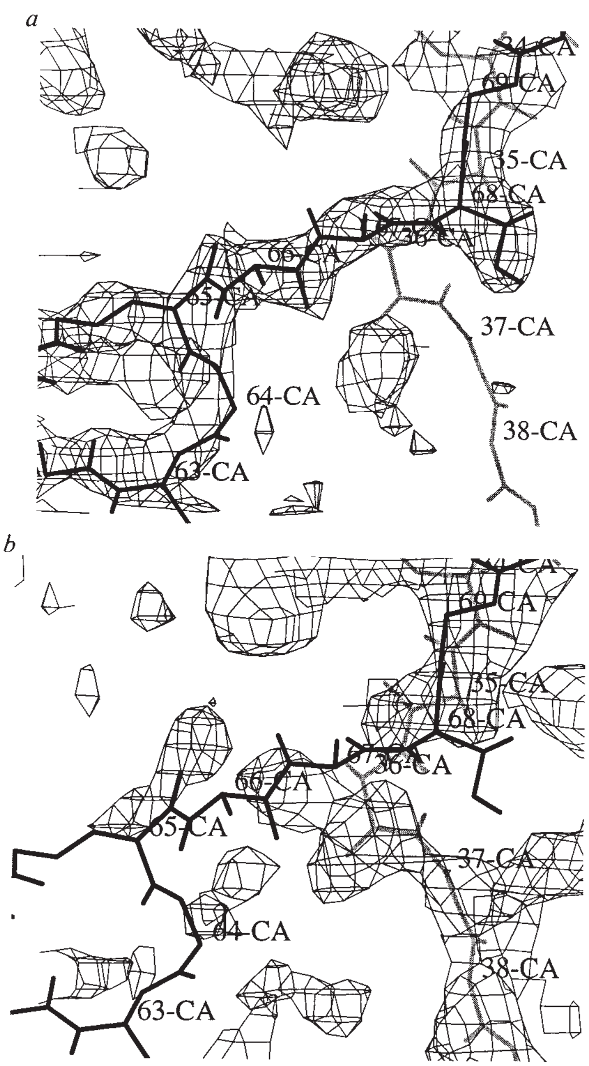

FIG. 2 The region around residue 66 of the small subunit of RuBisCO. The correct[17] structure is shown in black, whereas the incorrect[18] structure, which involved the nearly backwards tracing of the polypeptide chain of the small subunit, is shown in grey. Superimposed are σ_A-weighted[28] $2F_o - F_c$ electron density maps with phases computed from the correct model (*a*, $R = 0.16$, $R_T^{\text{free}} = 0.34$) and incorrect model (*b*, $R = 0.2$, $R_T^{\text{free}} = 0.47$) shown at 2.5 Å resolution for a contour level of 1σ. The maps are ordinary omit maps (for a review of omit map techniques, see A. Hodel, D. Eisenberg, S.-H. Kim and A.T.B., manuscript in preparation), that is residues 36–47 and all residues within 5 Å of this loop were removed in the phase calculation. T was obtained by a 10% random selection from the observed reflections. SA-refinements and restrained B-factor refinements were done at 2.5 Å resolution using A. The r.m.s. deviations of bond lengths and bond angles from ideal were 0.02 Å and 4°, respectively, for the correct structure whereas they were 0.03 Å and 5° for the incorrect structure.

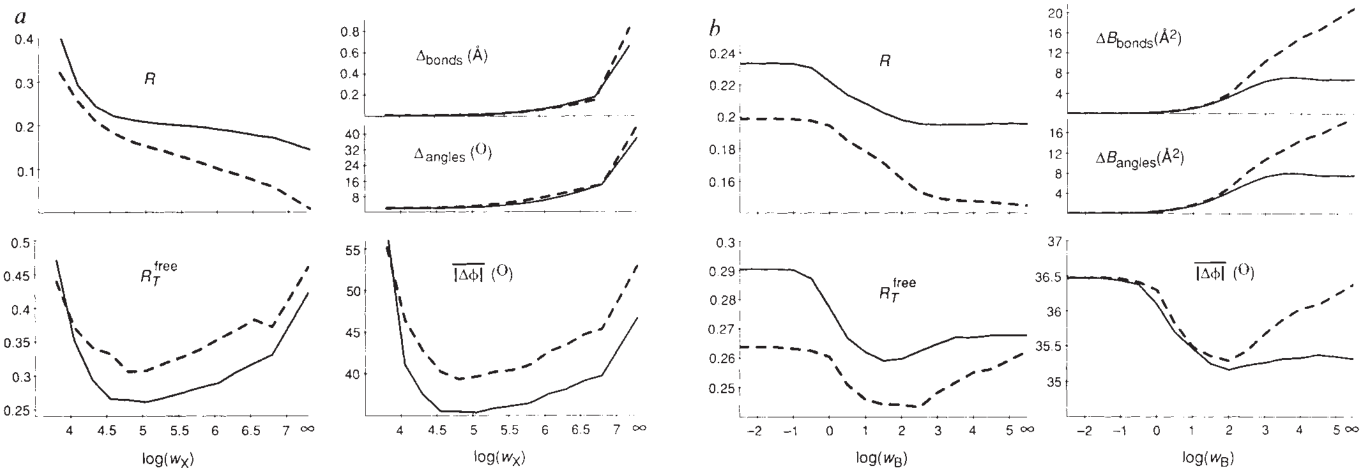

FIG. 3 *a*, SA-refinements of penicillopepsin[14,15] at 6–1.8 (solid lines) and 6–2.8 Å (dashed lines) resolution as a function of w_x (equation (5)) with R' replaced by R'_A (equation (4)). T was obtained by a 10% random selection. R was computed for A. Δ_{bonds} and Δ_{angles} are the r.m.s. deviations of bond lengths and bond angles from their ideal values. $\overline{|\Delta\Phi|}$ is the figure-of-merit weighted mean phase difference between model phases and the most probable MIR phases at 6–2.8 Å resolution. Details of the penicillopepsin model and refinement procedure are the same as in Fig. 1. The standard linear correlation coefficient between the R_T^{free} and $\overline{|\Delta\Phi|}$ graphs is 0.98 for both resolution ranges. *b*, Restrained B-factor refinements of penicillopepsin as a function of w_B (equation (6)) with R' replaced by R'_A (equation (4)). ΔB_{bonds} and ΔB_{angles} are the r.m.s. deviations between B-factors of atoms sharing a covalent bond or bond angle, respectively. $w_B = \infty$, represents the completely unrestrained case whereas $w_B \rightarrow -\infty$ represents refinement of a single overall B-factor.

Diffraction data and prior knowledge are often combined as is the case in restrained least-squares refinement of atomic positions[2] that can be viewed as minimization of the atomic coordinates against a cost function[19,20]

$$C = w_x R' + E_{chemical} \quad (5)$$

where w_x is a weight and $E_{chemical}$ is a geometric[19] or empirical[21] energy function which has been made unitless by multiplication of a conversion factor. If w_x chosen is too small, too much emphasis is put on the geometry as provided in $E_{chemical}$, which results in an inaccurate R value. If w_x is chosen too large, the structure will be overfit to a very good R value, but the geometry of the structure becomes severely distorted. The optimal choice of w_x cannot be obtained by linear hypothesis tests[13] because of the presence of nonlinear restraints, such as repulsive contact functions[2]. R_T^{free} is not subject to such limitations.

A series of positional refinements of the penicillopepsin structure[14,15] produced a minimum for R_T^{free} at $\log(w_x) = 5$, independent of the resolution range used (Fig. 3*a*). At this minimal value the r.m.s. deviation of bond lengths and bond angles from ideality are 0.013 Å and 2.5°, respectively. As an independent determination of the optimal w_x I used the multiple isomorphous replacement (MIR) phases at 6–2.8 Å resolution[14,15]; these phases were of exceptional quality with a figure of merit of 0.9. Experimental phase information is normally less accurate, incomplete or missing. R_T^{free} is highly correlated with the mean difference between the model and the MIR phases ($\overline{|\Delta\Phi|}$) (Fig. 3*a*). R_T^{free} thus yields the optimal choice for w_x without reference to experimental phase information or to expected deviations of the geometry from ideality. The resulting relatively tight geometry is a consequence of the diffraction data, not of the geometric or empirical energy function.

Restrained temperature factor refinement[22] poses a similar

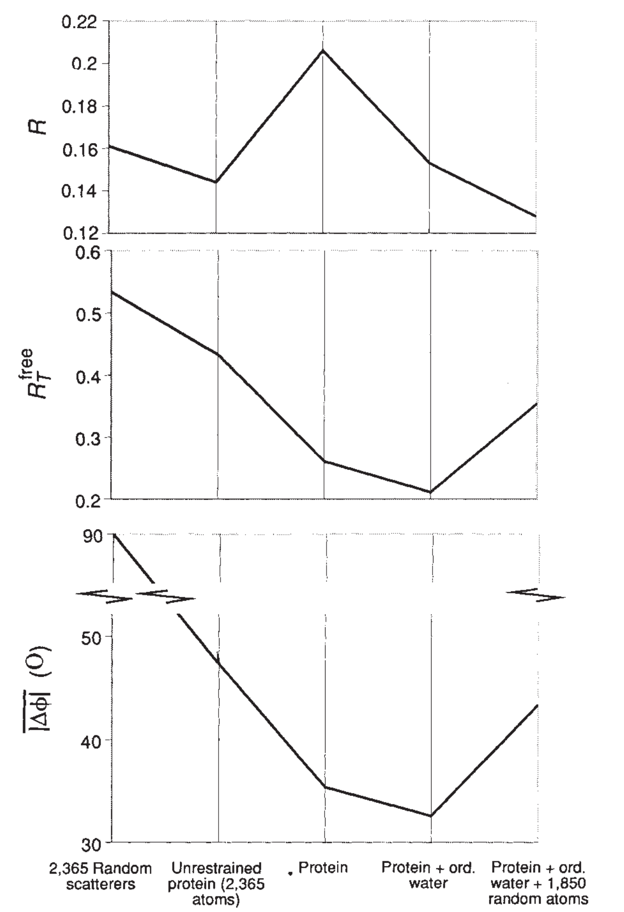

FIG. 4 '2,365 Random scatterers' consists of 2,365 oxygen atoms with a reduced van der Waals' radius of 1.57 Å randomly placed in the asymmetric unit of the crystal. 'Unrestrained protein', consists of the same scatterers placed near the non-hydrogen positions of the protein portion of the penicillopepsin structure. 'Protein' is the protein portion of the penicillopepsin structure refined with chemical restraints (equation (5)). 'Protein + ord.water' includes an additional 314 ordered water molecules. 'Protein + ord.water + 1,850 random atoms' includes an additional 1,850 oxygen atoms randomly placed in the bulk solvent region. The definition of T and $\overline{|\Delta\Phi|}$ is identical to Fig. 3. Each refinement consisted of a two iterations of SA-refinement and restrained B-factor refinement as detailed in Fig. 1. The standard linear correlation coefficient between the R_T^{free} and $\overline{|\Delta\Phi|}$ graphs is 0.98.

problem as positional refinement. It consists of minimization of a cost function

$$C = w_B R' + \sum_{(i,j)\,\text{bonds}} \frac{(B_i - B_j)^2}{\sigma^2_{\text{bonds}}} + \sum_{(i,j,k)\,\text{angles}} \frac{(B_i - B_k)^2}{\sigma^2_{\text{angles}}} \qquad (6)$$

where B_i is the temperature factor of atom i and the summations are carried out over all covalent bonds and bond angles[22]. R_T^{free} is highly correlated with $\overline{|\Delta\Phi|}$ and thus determines the optimal choice of w_B (Fig. 3*b*).

The information content of a random distribution of scatterers is obviously minimal, although it can be refined to a very low R value (Fig. 4); R_T^{free} stays at 54% which is close to the random limit of 59% for an acentric space group[11]. Unrestrained refinement with a model consisting of the same scattering starting at the positions of the non-hydrogen protein atoms yields $R_T^{\text{free}} =$ 43% (Fig. 4). Thus, R_T^{free} can distinguish between a distribution of scatterers that is close to the crystal structure and a random distribution, both of which can be refined to a very low R. Inclusion of chemical restraints increases R somewhat while greatly decreasing both R_T^{free} and $\overline{|\Delta\Phi|}$, thus improving the information content of the model (Fig. 4). Inclusion of ordered water molecules lowers R, R_T^{free} and $\overline{|\Delta\Phi|}$ (Fig. 4). Refinement of randomly placed scatterers in the bulk solvent region of the crystal lowers R while increasing both R_T^{free} and $\overline{|\Delta\Phi|}$, thus decreasing the information content.

R_T^{free} represents a reliable and unbiased parameter by which to evaluate the information content of a model produced by X-ray crystallography. It is not restricted to high-resolution diffraction data: tests carried out both at 6–2.8 Å and at 6–1.8 Å resolution produce large correlations between R_T^{free} and $\overline{|\Delta\Phi|}$ (Fig. 3). The observation that R_T^{free} can distinguish between a random distribution of scatterers and distribution close to the protein suggests applications to *ab initio* phasing. The increase of R_T^{free} on modelling the bulk solvent region of the penicillopepsin structure with stationary atoms confirms the disordered character of bulk solvent. A similar approach might be useful for the three-dimensional structure determined by solution NMR[23,24,29] if sufficient redundance in the data and accuracy of the NOE (nuclear Overhauser effect) intensities can be achieved. □

Received 14 August; accepted 30 October 1991.

1. Jensen, L. H. *Meth. Enzym.* **115,** 227–234 (1985).
2. Hendrickson, W. A. *Meth. Enzym.* **115,** 252–270 (1985).
3. Brändén, C. I. & Jones, A. *Nature* **343,** 687–689 (1990).
4. Mosteller, F. & Tukey, J. W. *Data Analysis and Regression: a Second Course in Statistics* (Addison-Wesley, Reading, Massachusetts, 1977).
5. Efron, B. & Tibshirani, R. *Science* **253,** 390–395 (1991).
6. Diamond, R. *Acta crystallogr.* **A46,** 425–435 (1990).
7. Kuriyan, J. & Weis, W. I. *Proc. natn. Acad. Sci. U.S.A.* **88,** 2773–2777 (1991).
8. Kuriyan, J. *et al. Proteins* **10,** 340–358 (1991).
9. Gros, P., Gunsteren, W. F. & Hol, W. G. J. *Science* **249,** 1149–1152 (1990).
10. Badger, J. & Caspar, D. L. D. *Proc. natn. Acad. Sci. U.S.A.* **88,** 622–626 (1991).
11. Stout, G. H. & Jensen, L. H. in *X-ray Structure Determination, a Practical Guide* 2nd edn 343–378 (Wiley, New York, 1989).
12. Press, W. H., Flannery, B. P., Teukolosky, S. A. & Vetterling, W. T. in *Numerical Recipes* 498–504 (Cambridge University Press, 1986).
13. Hamilton, W. C. *Acta crystallogr.* **18,** 502–510 (1965).
14. Hsu, I.-N., Delbare, L. T. J., James, M. N. G. & Hofmann, T. *Nature* **266,** 140–145 (1977).
15. James, M. N. G. & Sielecki, A. R. *J. molec. Biol.* **163,** 299–361 (1983).
16. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
17. Curmi, P. A. M., Schreuder, H., Cascio, D., Sweet, R. M. & Eisenberg, D. *J. biol. Chem.* (in the press).
18. Chapman, M. S. *et al. Science* **241,** 71–74 (1988).
19. Brünger, A. T. *A. Rev. Phys. Chem.* **42,** 197–233 (1991).
20. Jack, A. & Levitt, M. *Acta crystallogr.* **A34,** 931–935 (1978).
21. Brooks, B. R. *et al. J. comp. Chem.* **4,** 187–217 (1983).
22. Konnert, J. H. & Hendrickson, W. A. *Acta crystallogr.* **A36,** 344–349 (1980).
23. Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon, Oxford, 1987).
24. Wüthrich, K., *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
25. Brünger, A. T. *J. molec. Biol.* **203,** 803–816 (1988).
26. Brünger, A. T., Krukowski, A. & Erickson, J. *Acta crystallogr.* **A46,** 585–593 (1900).
27. Brünger, A. T. X-PLOR, Version 2.1, Yale University (1990).
28. Read, R. *Acta crystallogr.* **A42,** 140–149 (1986).
29. Clore, G. M. & Gronenborn, A. M. *Science* **252,** 1390–1399 (1991).

ACKNOWLEDGEMENTS. I thank M. N. G. James and A. R. Sielecki for providing the diffraction data and coordinates of the penicillopepsin structure, D. Eisenberg for providing the data and coordinates of RuBisCO and the Pittsburgh Supercomputer Center for support.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief; grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value $5' to prevent the imposition of import duty and value-added tax (VAT).