Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 16 March 2004
Accepted 19 May 2004

Application of a modified Oszlányi and Süto *ab initio* charge-flipping algorithm to experimental data

J. S. Wu,[a]* J. C. H. Spence,[a] M. O'Keeffe[b] and T. L. Groy[b]

[a]Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504, USA, and [b]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA. Correspondence e-mail: jinsong.wu@asu.edu

The structures of two crystals have been solved using a new iterative phasing method. The iterative phasing algorithm is developed from the 'charge-flipping' method proposed by Oszlányi & Süto [*Acta Cryst.* (2004), A**60**, 134–141]. Positivity and point-atom constraints are incorporated within this extremely simple and effective algorithm by flipping (sign reversal) of less-positive density values during the iterations. Convergence is reliably achieved and the two structures were solved. This structure solution method does not require information on atomic scattering factors or symmetry. Heavy atoms can be distinguished from light ones by their charge-density values.

1. Introduction

The phase problem of X-ray crystallography can usually be solved because phase information, which cannot be directly measured, is encoded in the diffracted intensity or Fourier moduli. For diffraction from crystals, the positivity constraint (known sign of charge density) and atomicity constraint (that the density consists of a sum of point-like atoms, between which the density is approximately zero) form the basis of *ab initio* (direct methods) phase-determination methods, *e.g.* Karle–Hauptman determinants for structure factors, which are Fourier transforms of a positive electron density (Karle & Hauptman, 1950). An important step in utilizing atomicity was taken by Sayre (1952) and Hauptman & Karle (1953) in their probabilistic analysis of structure factors. For non-periodic scatterers, the iterative Gerchberg–Saxton–Fienup hybrid input–output algorithm, (referred to here as *HiO*) (Gerchberg & Saxton, 1972; Fienup, 1982) provides another practical method for phase retrieval from 'over-sampled' diffraction data. (In its original form, the Gerchberg–Saxton algorithm required knowledge of both image and diffraction pattern intensities, from which complex image and diffraction pattern were sought by iteration). Oversampling, which is not possible for three-dimensional crystals, normally requires a known support, or estimate of the object boundary, outside which the density is known to be zero. For weakly scattering (real) objects (without multiple scattering or absorption), the Fourier modulus constraint in reciprocal space and compact support constraint in real space have been found sufficient for convergence in simulations and experiments with X-rays (Miao *et al.*, 1999; Miao & Sayre, 2000; He *et al.*, 2003), laser light and coherent electrons in transmission electron microscopy (TEM) (Weierstall *et al.*, 2001). The first atomic resolution image of a carbon nanotube was recently obtained by applying this method to the electron nanodiffraction pattern from a single nanotube (Zuo *et al.*, 2003). More recently, the need for knowledge of the object shape has been eliminated with the development of the '*Shrinkwrap*' variant of the *HiO* algorithm, which obtains the support from the known autocorrelation function of the object, and improves this iteratively within *HiO* (Marchesini *et al.*, 2003). All these methods have much in common with the density-modification and solvent-flattening (Wang, 1985) approaches of crystallography. The direct application of the *HiO* iterative algorithm in crystallography is possible only if it is accompanied by other known constraints, such as non-crystallographic symmetry (Millane, 1990). It has been shown, however, that *HiO* can be used for phase extension for two-dimensional projections (Wu & Spence, 2003). The *HiO* algorithm has also been applied to two-dimensional monolayer crystals, by oversampling along the reciprocal-lattice rods, and so greatly improving the convergence of TEM cryomicroscopy data from proteins (Spence *et al.*, 2003). The atomicity constraint has also been used by Elser (2003) in an iterative algorithm that picks up atoms from the current density map.

Recently, a remarkably simple and effective method for adding the atomicity constraint to an iterative algorithm was developed by Oszlányi & Süto (2004). In that algorithm, by iteratively reversing (flipping) the sign of the charge density in regions where it lies below a small threshold, it was found that a 'support' function for point-like atoms can be found. This charge-flipping algorithm can be considered as a special case of the output–output variant of the *HiO* algorithm (Wu *et al.*, 2004). The algorithm assumes a real density in real space (without, for example, spatially dependent absorption or multiple scattering) and takes account of the small number of negative values in the charge density, which result from truncation of the structure factors in reciprocal space. It has been successfully used to find three-dimensional structures using simulated X-ray diffraction data (Oszlányi & Süto, 2004). In our work, we have applied a modified version of the algorithm to experimental X-ray diffraction data. The main modification we have applied to the algorithm is to change the way the threshold δ is set. Instead of using a fixed value of δ, we flip a certain fraction $1-\phi$. Pixels of density are ranked in order of magnitude from most positive to most negative, and the $1-\phi$ with the smallest density are identified as falling within our threshold. The sign of the density for these pixels, which includes all negative values, is reversed ('flipped').

In the following, a brief description of the charge-flipping algorithm is given, followed by a report of our experience with structure solution of two crystals using experimental X-ray diffraction data.

2. Phase-retrieval procedures

Given a set of experimentally observed moduli $|F_{\rm obs}(\mathbf{h})|$, lattice parameters and the space group of an unknown structure, the first step is to merge equivalent reflections according to the symmetry of the crystal. A set of random phases $\varphi(\mathbf{h})$ is then generated consistent with Friedel's law: $-\varphi(\mathbf{h}) = \varphi(-\mathbf{h})$. Combining $F(\mathbf{h})$ and $\varphi(\mathbf{h})$ and applying a Fourier transform to this current set of structure factors, we can generate the first estimate of the charge density $\rho(\mathbf{r})$. This estimate of $\rho(\mathbf{r})$ is a real function with both positive and negative values and with symmetry $P1$. In the starting cycle, $F(0)$ is set to 0, the integral over the first estimate of $\rho(\mathbf{r})$ is therefore zero. This accounts for the appearance of the negative values, in addition to the effect of truncation of structure factors at the resolution limit. The iterative Fourier transform algorithm for the kth iteration is given below.

1. Sort $\rho_k(\mathbf{r})$ in descending order of magnitude and set a ratio $\phi < 1$. Select m pixels with largest positive values, where $m = n\phi$ and n is the total number of pixels. These m pixels form the current support estimate $S_k(\mathbf{r})$, the boundaries of the atoms. Charge flipping is applied by retaining the values of these m pixels and flipping (reversing the sign of) all the rest:

$$\rho_{k+1}(\mathbf{r}) = \begin{cases} \rho_k(\mathbf{r}) & \text{if } \mathbf{r} \in S_k(\mathbf{r}) \\ -\rho_k(\mathbf{r}) & \text{if } \mathbf{r} \notin S_k(\mathbf{r}). \end{cases} \tag{1}$$

2. Fourier transform $\rho_{k+1}(\mathbf{r})$ to obtain $F_{k+1}(\mathbf{h})$.

3. Replace resulting structure-factor moduli with measured moduli $|F_{\rm obs}(\mathbf{r})|$. Three types of reflections are distinguished here: observed reflections $\mathbf{h} \in \mathbf{H}_{\rm obs}$; high-frequency reflections $\mathbf{h} \notin \mathbf{H}_{\rm obs}$ beyond the resolution limit; and the 000 reflection:

$$F_{k+1}(\mathbf{h}) = \begin{cases} |F_{\rm obs}(\mathbf{h})| \exp(i\varphi_k) & \text{if } \mathbf{h} \in \mathbf{H}_{\rm obs} \\ F_k(0) & \text{if } \mathbf{h} = 0 \\ 0 & \text{if } \mathbf{h} \notin \mathbf{H}_{\rm obs} \text{ and } \mathbf{h} \neq 0. \end{cases} \tag{2}$$

4. Inverse Fourier transform of $\mathbf{F}_{k+1}(\mathbf{h})$ to obtain $\rho_{k+1}(\mathbf{r})$.

5. Go to step 1 with k replaced by $(k+1)$.

The best fraction ϕ is found to lie around 0.2 for three-dimensional crystals. The progress of the iterations can be followed using a conventional residual R, calculated using the moduli of the current estimate of the charge density and the observed moduli:

$$R = \sum_{\mathbf{h}\in\mathbf{H}_{\rm obs}} ||F_k(\mathbf{h})| - |F_{\rm obs}(\mathbf{h})|| \Big/ \sum_{\mathbf{h}\in\mathbf{H}_{\rm obs}} |F_{\rm obs}(\mathbf{h})|. \tag{3}$$

We find that, with an appropriate ϕ, R decreases with iteration number, with the attainment of a plateau being a sign of convergence of the algorithm (see Fig. 1). The value of R can then be further reduced by using a larger ϕ, so that more pixels are not flipped, but note that this does not imply that a better solution is being achieved, only that a smaller amount of density is being flipped. Since $F(000)$ is usually not observed and the scale of the charge-density map may vary, the definition of a threshold using a percentage ϕ seems better than using a fixed parameter δ. Indeed, we find that R begins to decrease almost immediately using ϕ (see Fig. 1) rather than having an induction period of 100 or more cycles with fixed δ (Oszlányi & Süto, 2004).

Also shown in Fig. 1 is $F(000)$ as a function of iteration number. We find, as also observed by Oszlányi & Süto (2004), that achievement of the correct structure is signaled by a large drop of $F(000)$ to a constant value, and indeed the behavior of $F(000)$ might equally be used as a criterion for convergence to the correct structure.

3. Structure determination using the iterative algorithm

Single crystal X-ray diffraction data collected in our laboratory (C. F. Bonneau & T. L. Groy, to be published) from monoclinic hexabromobenzene (C_6Br_6) ($P2_1/n$, $a = 8.381$, $b = 4.0192$, $c = 15.3939$ Å, $\beta = 92.674°$, $Z = 2$) were used as input to the iterative phasing algorithm. 4387 observed reflections, in a sphere of resolution of about 0.8 Å ($-10 \le h \le 10$, $-5 \le k \le 5$ and $-19 \le l \le 19$) were used to form 1355 unique reflections

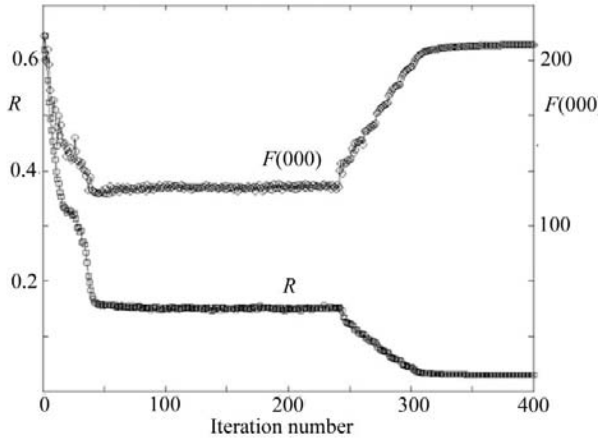

Figure 1
R and current estimation of $F(000)$ *versus* iteration number for the C_6Br_6 structure solution with ϕ initially equal to 0.185. After 240 iterations, ϕ is increased in increments of 0.05 to a final value of 0.535.

after merging using the given symmetry. We then imposed symmetry again to these 1355 reflections to generate altogether 4738 reflections. This was the starting set of experimental data used to fill in a three-dimensional array with dimension 21 × 11 × 39. All unobserved high-order reflections in the array but outside the sphere were set to zero, and held at zero in the iterations. $F(000)$ was also set to zero at the beginning but updated with the current estimate at each iteration. The density map was calculated using 0.4 Å pixels. The parameter ϕ used in the iteration was 0.185 until the reconstruction converged. Fig. 2(*a*) shows the [010] projection of the reconstructed charge-density map at the 240th iteration, when the calculation has converged (little change in R). The corresponding residual R is 0.145. Fig. 2(*b*) shows the same projection with the charge-density map masked by the 24 highest density peaks. The molecules are clearly visible, the peaks for the heavy atoms (Br) are about six times those for the lighter atoms (C) as expected from the ratio of atomic numbers (5.83) and interatomic distances are reasonable, so we have essentially finished at this point.

In our tests on this structure, the algorithm always converges quickly to the correct structure, usually in less than 50 iterations. Furthermore, every test converged to the correct structure, independent of the (random) starting phases. These random phases result in random origins of coordinates. However, the resulting map is noisy and a number of smaller peaks occur in the density map, so we seek a procedure for reducing the background noise, so that structures of unknown composition might be solved.

The following procedure works well. After the initial plateau in R is reached, we increase ϕ by 5% every ten subsequent iterations until it reaches a value of 0.535. If we examine the peaks, we find that the positions and integral charge density of the Br and C atoms are similar before and after the change in ϕ but the noise level is reduced.

If the smaller electron-density peaks are merely noise, they should appear at random positions as the starting phases are varied, so we further reduced noise by averaging the results of eight refinements (with different random starting phases), referred to a common origin determined by maximizing the overlap of the strongest peaks, as in Fig. 2. The largest resulting noise peak in this average map was 60% of the smallest C peak, but can be identified as noise as follows. The magnitude of each peak is determined by integrating over a box of 3 × 3 × 3 pixels. For an atomic peak, the largest pixel in the box is 0.10–0.12 of the integral, but in a noise peak the largest pixel is about 0.05 of the integral (note that for a flat distribution the largest peak is $1/27 = 0.037$ of the integral).

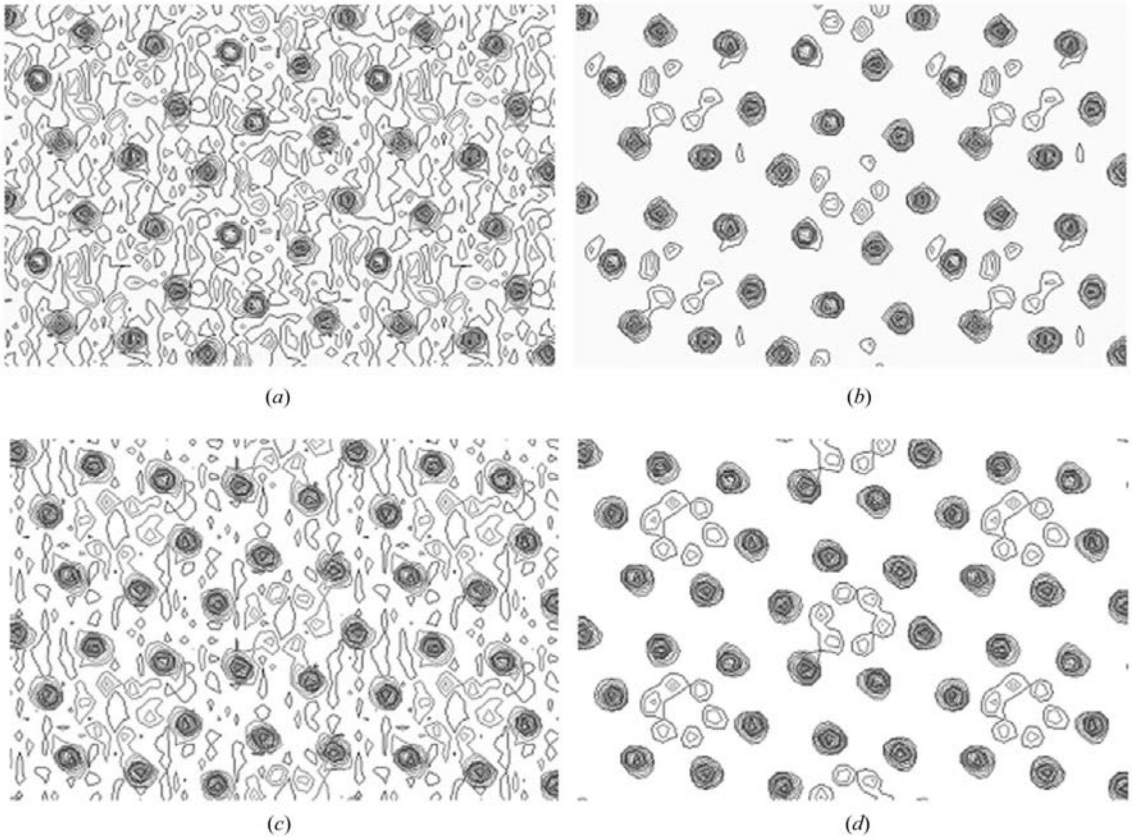

Figure 2
(*a*) [010] projection of the C_6Br_6 charge-density map by the iterative charge-flipping algorithm at the 240th iteration, when R is 0.145. (*b*) Application of an atomic mask to (*a*); only the 24 peaks with the highest density are selected and shown. (*c*) [010] projection of the charge-density map of C_6Br_6 obtained by averaging eight density maps starting from different random phases. (*d*) Application of an atomic mask to (*c*); only the 24 peaks with the highest density are selected and shown. The plane of the molecules is inclined at approximately 25° to (010).

Specifically, the largest pixel in the largest noise peak is only 0.3 times the largest pixel in the smallest C peak.

Of course in this case we know the symmetry ($P2_1/n$ is given uniquely by extinctions) and the atomic peaks found by the algorithm are observed to be closely related by symmetry. No such pattern is discerned for the noise. Thus, for crystals with symmetry higher than $P1$, the fact that symmetry is not used can be turned to good advantage. As we do know the symmetry in this case, we can finally use it to force the atomic coordinates to conform to the symmetry. Atomic parameters so determined are compared with the results of a conventional refinement (to $R_1 = 0.03$) in Table 1. Atomic positions agree within about 0.2 Å. Also listed are the relative peak heights.

The second crystal studied is orthorhombic $SnBr_4$–dioxane [$Pbcn$, $a = 8.7678$, $b = 11.2927$, $c = 11.1746$ Å, $Z = 4$, dioxane = $C_4H_8O_2$ (Bauer *et al.*, 2002)]. Altogether 9385 intensities within $-11 \leq h \leq 11$, $-14 \leq k \leq 14$ and $-14 \leq l \leq 14$ were measured, with the highest resolution at 0.8 Å. Data were merged according to the symmetry of the crystal. The 1411 reflections were then used to generate 9794 observed independent reflections. These reflections were then used to fill a three-dimensional structure-factor array with dimension 23 × 29 × 29. Again, all the high-order reflections outside the measured sphere and $F(000)$ were set to zero. The parameter ϕ was set to 0.2. The program was repeatedly executed and we found that about 80% of all runs converged, and then to the correct structure (20 runs), after a maximum of 200 iterations. Although the heavy and light atoms can be distinguished based on their charge-density values, it is difficult to distinguish light atoms such as C and O with similar atomic number.

4. Discussion

Charge flipping is a simple and easily implanted algorithm for *ab initio* structure determination for crystals based on our tests using experimental X-ray diffraction data. Unlike the *HiO* algorithm, it does not require knowledge of the support in real space, which (being the boundary of the atoms) is difficult to realize experimentally at atomic resolution. The *Shrinkwrap* algorithm similarly iteratively improves a support estimate based on a first guess from the autocorrelation function, which, as the Patterson function, could perhaps also be used in this case for a crystal, despite the lack of 'oversampled' data. In the charge-flipping iterative algorithm, the absence of the support constraint is compensated for by using atomicity (retention of large peaks). There have been previous attempts to add atomicity constraints to the iterative algorithm, see, for example, Elser (2003). Charge flipping does seem to be the best and simplest way to apply atomic constraints in an iterative algorithm. A large $F(000)$ in iterations is the sign of divergence, since the charge-flipping algorithm fails to find the large 'zero' area above which atomic peaks sit. Thus, a sharp decrease of $F(000)$ is also a good sign of convergence. Meanwhile, since the constraints in the charge-flipping algorithm (Fourier intensity and atomicity) are non-convex, no rigorous theorems regarding convergence can be established.

Table 1
Atomic coordinates in C_6Br_6 obtained by the charge-flipping algorithm and in parentheses from a conventional refinement.

d is the displacement between the two sets of coordinates in Å. *N* is the integrated density in a box of 1.2 × 1.2 × 1.2 Å (27 pixels)

Atom	*x*	*y*	*z*	*d*	*N*
Br(1)	0.76 (0.767)	0.83 (0.816)	0.128 (0.139)	0.18	6.5
Br(2)	0.39 (0.421)	0.58 (0.574)	0.204 (0.205)	0.26	6.3
Br(3)	0.15 (0.157)	0.22 (0.239)	0.059 (0.068)	0.17	6.3
C(1)	0.59 (0.614)	0.63 (0.635)	0.042 (0.058)	0.07	1.2
C(2)	0.46 (0.463)	0.50 (0.528)	0.096 (0.087)	0.18	1.2
C(3)	0.35 (0.354)	0.40 (0.391)	0.026 (0.029)	0.07	1.1

The flipping algorithm only works for data with atomic resolution (Oszlányi & Süto, 2004). We have explored this aspect further using the experimental $SnBr_4$–dioxane data. The program would still converge when the resolution was reduced to 1 Å. However, the sharp drop in the R curve, which is a sign of convergence, disappeared when the resolution was reduced to 1.4 Å. To our surprise, however, some features can still be seen in the reconstructed charge-density map, for example the heavy Sn atoms can still be distinguished in this lower-resolution charge-density map. Spurious peaks also appear, which make the determination of lighter atoms impossible. Recently, we have found that the charge-flipping algorithm can also be applied to image reconstruction of a non-periodic object, without atomic resolution (Wu *et al.*, 2004). In the present paper, the charge-flipping algorithm is applied to two relatively small structures. We have extended the calculations to significantly larger structures with some success. Although we used relative structure factors, we tried and confirmed that the usage of normalized structure factors is better since the atomic peaks become sharper.

Unlike direct methods, the charge-flipping algorithm does not require the scattering factors of the atoms to be known, or indeed any chemical information. (Here we refer to the algorithm we actually used, in which no normalized structure factors are used. Otherwise, calculation of the normalized structure factors requires cell content and scattering factors of the atoms). The symmetry was used only in merging the observed intensities and in reporting the final result, not within the algorithm. This means we only need to know the symmetry of the intensity, which is always an invariant, no matter how the origin moves. In this sense, compared to existing direct phasing methods, it is more '*ab initio*' than other algorithms in current use.

5. Conclusions

Two crystal structures have been solved *ab initio* using the iterative charge-flipping algorithm, which imposes atomic constraints in a simple way. Our modification to the original algorithm improves performance. In the reconstructed charge-density map, light atoms can be distinguished from heavy ones, which can be found at lower resolution. The operation of the algorithm is compared to the related *HiO* algorithm.

This work is supported by ARO award DAAD190010500 (JCHS) and NSF grant DMR 0103036 (MOK). JSW acknowledges helpful discussions with G. Oszlányi.

References

Bauer, M., Kouvetakis, J. & Groy, T. L. (2002). *Z. Kristallogr. New Cryst. Struct.* **217**, 421–422.

Elser, V. (2003). *Acta Cryst.* A**59**, 201–209.

Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.

Gerchberg, R. W. & Saxton, W. O. (1972). *Optik (Stuttgart)*, **35**, 237–246.

Hauptman, H. & Karle, J. (1953). *Am. Monograph* No. 3. Ann Arbor, USA: Edwards.

He, H., Marchesini, S., Howells, M., Weierstall, U., Hembree, G. & Spence, J. C. H. (2003). *Acta Cryst.* A**59**, 143–152.

Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.

Marchesini, S., He, H., Chapman, H., Hau-Riege, S., Noy, A., Howells, M., Weierstall, U. & Spence, J. C. H. (2003). *Phys Rev. B*, **68**, 140101(R).

Miao, J., Charalambous, C., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.

Miao, J. & Sayre, D. (2000). *Acta Cryst.* A**56**, 596–605.

Millane, R. P. (1990). *J. Opt. Soc. Am.* A**7**, 394–411.

Oszlányi, G. & Süto, A. (2004). *Acta Cryst.* A**60**, 134–141.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Spence, J., Weierstall, U., Fricke, T., Glaeser, R. & Downing, K. (2003) *J. Struct. Biol.* **144**, 209.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Weierstall, U., Chen, Q., Spence, J., Howells, M., Isaacson, M. & Panepucci, R. (2001). *Ultramicroscopy*, **90**, 171–195.

Wu, J. S. & Spence, J. C. H. (2003). *Acta Cryst.* A**59**, 577–583.

Wu, J. S., Weierstall, U., Spence, J. C. H. & Koch, C. T. (2004). *Opt. Lett.* Submitted.

Zuo, J. M., Vartanyants, I., Gao, M., Zhang, R. & Nagahara, L. A. (2003). *Science*, **300**, 1419–1421.