Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Co-crystallization and preliminary crystallographic analysis of the high mobility group domain of HMG-D bound to DNA

Frank V. Murphy IV,[a,b] Jonathan V. Sehy,[c] Linda K. Dow,[d] Yi-Gui Gao[d] and Mair E. A. Churchill[a,e]*†

[a]Department of Pharmacology, The University of Colorado Health Sciences Center, 4200 E. Ninth Avenue, Denver, CO 80262 USA, [b]Department of Biochemistry, University of Illinois at Urbana-Champaign, USA, [c]Program in Molecular Cell Biology, Washington University at St Louis, USA, [d]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, USA, and [e]The Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, USA

† Presently at: Pharmaceutical Product Development at Eli Lilly and Company.

Correspondence e-mail: mair.churchill@uchsc.edu

Received 12 May 1999
Accepted 17 June 1999

Structural studies are essential to understand mechanisms of non-sequence-specific DNA binding used by chromosomal proteins. A non-histone high-mobility group (HMG) chromosomal protein from *Drosophila melanogaster*, HMG-D, binds duplex DNA in a non-sequence-specific fashion. The DNA-binding domain of HMG-D has been co-crystallized with a duplex DNA fragment in the primitive orthorhombic space group $P2_12_12_1$, with unit-cell dimensions $a = 43.74$, $b = 53.80$, $c = 86.84$ Å. Data have been collected to 2.20 Å at 99 K, with diffraction observed to at least 2.0 Å. Heavy-atom derivative crystals have been obtained by co-crystallization with oligonucleotides halogenated at major-groove positions near the end of the DNA.

1. Introduction

Chromosomal proteins are essential for correct patterns of gene expression; however, the current understanding of their structures and interactions with DNA lags substantially behind that of their sequence-specific counterparts. The structure of the nucleosome core particle was the first eukaryotic chromosomal protein–DNA complex to be determined (Luger *et al.*, 1997). Structures of DNA complexes of members of the Sac7d class of archaeal chromosomal proteins have also been determined recently using X-ray crystallography (Robinson *et al.*, 1998). These novel structures reveal different mechanisms of non-sequence-specific DNA binding. In the nucleosome, the core histone proteins contact the DNA, making primarily electrostatic interactions on the inside of the DNA bend. In contrast, hydrophobic interactions dominate the Sac7d DNA interface, which is on the outside of a DNA bend. The high-mobility group (HMG) proteins are important in the structure and function of chromatin (Bustin & Reeves, 1996), but their interactions with DNA in chromatin are poorly defined. There are currently no structures of the non-sequence-specific HMG proteins bound to linear DNA.

Members of the HMG-domain superfamily, typified by HMG1/2, bind to DNA with different degrees of specificity using the same DNA-binding motif, known as the HMG-box (Bianchi *et al.*, 1992; Bustin & Reeves, 1996; Ferrari *et al.*, 1992; Giese *et al.*, 1992; Grosschedl *et al.*, 1994; Wolffe, 1994). Analysis of the HMG-box sequences reveals that several residues correlate with the specificity of the domain. In fact, the identity of these residues can define novel HMG-box proteins as belonging to the sequence-specific transcription-factor class or to the non-sequence-specific chromosomal class (Balaeff *et al.*, 1998; Baxevanis & Landsman, 1995; Ner, 1992). HMG-D is a chromosomal protein (Churchill *et al.*, 1995) and, like other HMG proteins, it preferentially binds to DNA structures which are pre-bent and underwound (Churchill *et al.*, 1995, 1999; Payet & Travers, 1997; Wolfe *et al.*, 1995). HMG-D contains only a single copy of the HMG-box DNA-binding domain (Ner & Travers, 1994; Wagner *et al.*, 1992) at the N-terminus, which is followed by a 'tail' region which has 'basic motifs' similar to the C-terminal domain of histone H1 and a C-terminal acidic stretch similar to those in HMG1/2. The structures of the HMG-domain of HMG-D and HMG1 have been determined by NMR (Hardman *et al.*, 1995; Jones *et al.*, 1994; Read *et al.*, 1993; Weir *et al.*, 1993), revealing an L-shaped fold comprised of three helices held together by two hydrophobic cores. The structures of complexes of sequence-specific HMG-domains, LEF-1 and SRY, bound to DNA have also been determined by NMR and reveal a common protein fold and severe protein-induced DNA bending toward the major groove (Love *et al.*, 1995; Werner, Bianchi *et al.*, 1995; Werner, Huth *et al.*, 1995). A structure of a chromosomal HMG-box protein bound to DNA is not yet known, but the structure of the complex of HMG-D74 bound to DNA has been predicted using molecular-dynamics simulations (Balaeff *et al.*, 1998). This model has many features which are similar to those observed in the LEF-1 and SRY-DNA complexes, but also has some additional features, including a side-

chain intercalation which was not previously observed and which may be important for non-sequence-specific DNA recognition.

Crystallographic studies with HMG-D bound to duplex DNA were initiated in order to obtain a high-resolution view of an HMG-domain–DNA complex and to test the predictions made by the molecular-dynamics model. The crystallization and preliminary characterization of the complex between HMG-D and a duplex DNA decamer is described.

Table 1
Summary of derivative crystal diffraction and data.

Name	Unit-cell dimensions (Å)			Diffraction	Completeness (%)	R_{sym} (%)†
Native (293 K)	46.30	54.35	87.62	2.8	>96	10.0
Native (99 K)	43.74	53.80	86.84	2.2	>98	5.2
Iodo-C2	44.84	53.96	86.98	2.4	98.3	4.3
Bromo-C2	43.07	53.03	85.92	2.5	n.d.	n.d.
Iodo-C8	44.30	55.01	85.47	2.8	n.d.	n.d.
Bromo-C8	39.36	53.79	85.48	2.5	n.d.	n.d.
Iodo-T7				≥5.0		

† R_{sym} (on intensity) = $\sum\sum|I(h)_i - \langle I(h)\rangle|/\sum\sum I(h)_i$, where $I(h)_i$ is the observed intensity and $\langle I(h)\rangle$ is the mean intensity of reflection h over all measurements of $I(h)$.

2. Materials and methods

2.1. Protein and DNA purification

The first 74 residues of HMG-D, which comprise the HMG-box domain, were expressed and purified as described previously (Dow *et al.*, 1997; Jones *et al.*, 1994). To increase yields of the unoxidized protein, chemical reduction of the methionine oxide was performed by incubating the protein at 298 K with 0.725 *M* *n*-methyl mercaptoacetamide buffered at pH 7.0 for 72 h (Houghten & Li, 1979). The native protein was then separated from the oxidized form by isocratic reverse-phase HPLC in phosphate-buffered (pH 2.2) 28% acetonitrile (μBondapak C18 column, Waters; Dow *et al.*, 1997). The organic solvent was removed using SP Sepharose Fast Flow (Pharmacia) chromatography before the protein was concentrated and dialyzed by ultrafiltration (Amicon, Centricon). Electrospray ionization mass spectrometry confirmed the presence of pure homogeneous native protein. The DNA oligonucleotides were synthesized (Operon Technologies) and purified by size-exclusion chromatography on a Sephacryl HR 200 (Pharmacia) column with a mobile phase of 27 m*M* ammonium hydroxide. The DNA was then lyophilized, resuspended, annealed and finally diafiltered and concentrated (Amicon). The final concentrations of protein and DNA were determined by UV-absorbance spectroscopy.

2.2. Co-crystallization

Small crystals were first obtained within a week using the vapor-diffusion method with a sitting drop consisting of 4 μl 4.7 m*M* HMG-D74, 4 μl 10 m*M* Tris–HCl pH 7.5, 3 μl 30% PEG 4K, 2 μl 10 m*M* GCGATATCGC duplex equilibrated against 30 ml 30% PEG 4K at room temperature. Crystals were improved through screening; the drop, which contained 4 μl 4.7 m*M* HMG-D74, 4 μl well solution, 0.5 μl water, 1.5 μl 10 m*M* GCGATATCGC duplex, was equilibrated against 1 ml of well solution [23%(*w*/*v*) PEG 4K (Fluka), 20 m*M* Tris–HCl pH 7.85]. The best crystals were obtained by streak-seeding 48–60 h after the initial setup (Stura & Wilson, 1991).

DNA sequences for derivative crystals were iodo-T7 (GCGATA$^{5\text{-I}}$UCGC), iodo- and bromo-C2 (G$^{5\text{-Br}}$CGATATCGC) and iodo- and bromo-C8 (GCGATAT$^{5\text{-Br}}$CGC). Derivative crystals grew under similar conditions to the native crystals.

2.3. Data collection and analysis

Diffraction data were collected on an R-AXIS IIC image-plate detector using a Rigaku rotating-anode generator (λ = 1.5418 Å, 50 kV, 80 mA). Data collection was performed at room temperature and at 103 K. Cryo-cooling required that the crystal be immersed in cryoprotectant [27%(*w*/*v*) PEG 4K, 7%(*v*/*v*) glycerol] for several seconds prior to mounting and freezing (Rogers, 1994). A typical crystal-to-image plate distance was 100 mm, with oscillations of 1.5°. Images were auto-indexed and the data processed and reduced with the programs *DENZO* and *SCALEPACK* implemented in *HKL* using a Silicon Graphics workstation (fewer than 0.3% of the reflections were rejected for poor agreement; Otwinowski & Minor, 1997). The intensities were then truncated to amplitudes using the *TRUNCATE* program in the *CCP*4 suite of crystallographic programs (Collaborative Computational Project, Number 4, 1994).

3. Crystallization and X-ray crystallographic analysis

Previous attempts to co-crystallize HMG-D with DNA were unsuccessful because of protein inhomogeneity arising from methionine oxidation (Churchill *et al.*, 1995; Dow *et al.*, 1997). The oxidation is not easily prevented, but a method was developed to separate the two species of protein (Dow *et al.*, 1997). A combination of chemical reduction of the methionine sulfoxide and reverse-phase chromatographic separation of the two proteins was used to obtain sufficient quantities of pure protein for crystallization trials.

Co-crystals were obtained by screening different DNA duplexes of suitable length and sequence in crystallization trials. Crystals of a complex of HMG-D74 and the DNA fragment containing the sequence GCGATATCGC were grown from a 5 m*M* solution by vapor diffusion at 293 K against a well solution of 23% PEG 4K and 10 m*M* HEPES pH 7.85 (Fig. 1*a*). Analysis using

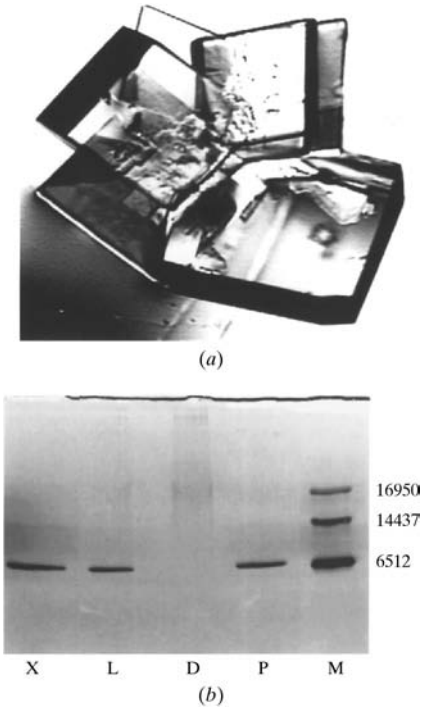

Figure 1
Typical native crystals (*a*) contain HMG-D and DNA (*b*). (*a*) Photomicrograph of HMG-domain–DNA complex crystals, showing chunky plates of dimensions 0.8 × 0.8 × 0.4 mm. (*b*) Silver-stained 18% SDS–PAGE of contents of a single dissolved well rinsed crystal (X) and control lanes of mother liquor (L), DNA (D), pure HMG-D74 (P) and molecular-weight markers (M). The DNA is visualized by UV shadowing (not shown).

SDS–PAGE of the rinsed and dissolved crystals reveals by silver-staining and UV-shadowing techniques (data not shown) that they contain both protein and DNA (Fig. 1*b*). DNA-binding experiments using competitive electrophoretic mobility shift assays had previously shown that this DNA length would bind to HMG-D (Churchill *et al.*, 1995).

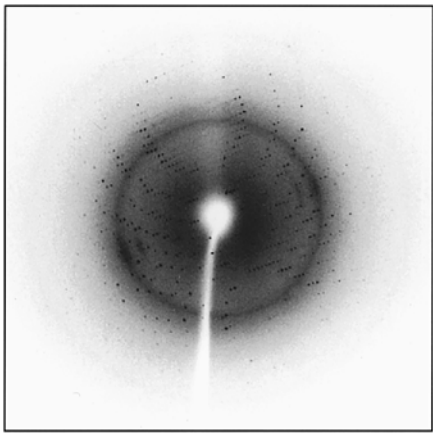

Figure 2
Crystallographic analysis of native HMG-D–DNA co-crystals. Typical diffraction image from data collection on an R-AXIS IIC. Data were collected under cryo-cooling conditions (99 K nitrogen stream) after being dunked in a cryoprotectant consisting of the mother liquor containing 7% glycerol. Exposure to Cu $K\alpha$ X-rays for each 1.5° oscillation image was 15 min; the edge of the detector is 1.9 Å at a distance of 100 mm. Ice rings typical of frozen crystals are observable, but do not interfere with very many reflections.

The crystals diffract to approximately 2.0 Å and data were collected under room temperature and cryo-cooling conditions. The unit-cell dimensions shrink by 1–5% in each dimension on freezing, as shown in Table 1. The effective resolution increases from 2.8 to 2.2 Å and the quality of the data collected increases substantially with cryo-cooling (Rogers, 1994). Reflections are visible to the edge of the detector, which is 1.8 Å at a distance of 100 mm, as shown in Fig. 2. Table 2 shows the data-collection statistics of the best native data set. It is 98% complete, with an $R_{\rm sym}$ value of 5.2% for all data to 2.20 Å and sufficient redundancy (average of eight measured per unique reflection). There appears to be one complex per asymmetric unit, with a solvent content of ∼51%.

Derivatives for multiple isomorphous replacement (MIR) or multiple-wavelength anomalous dispersion (MAD) phasing techniques can be prepared simply by substituting cytosine positions with 5-bromo or 5-iodo cytosine and thymine positions with 5-bromo or 5-iodo uridine (Hendrickson & Ogata, 1997). Excluding the end base pair, this approach produces two heavy-atom sites in the palindromic DNA fragment for each incorporated nucleotide. Co-crystallization trials were conducted with the series of halogen-substituted DNA fragments, and crystals of varying qualities grew for several of the complexes. The crystals and diffraction patterns are illustrated in Fig. 3. The crystals have slightly different morphologies compared with the native crystals. The best diffraction was seen for the cytosine substitutions, which are at the ends of the binding site, whereas the thymines are at the center of the binding site. The increased charge of the halogen may interfere with crystallization when it is in the center of the binding site, because the major groove would be most compressed in this region (Wolfe *et al.*, 1995). All of the derivative crystals obtained were non-isomorphous with the native crystals. Therefore, the best approach for solution of the structure will be to obtain multi-wavelength data from one of the bromine-containing crystals and solve the structure using MAD phasing techniques.

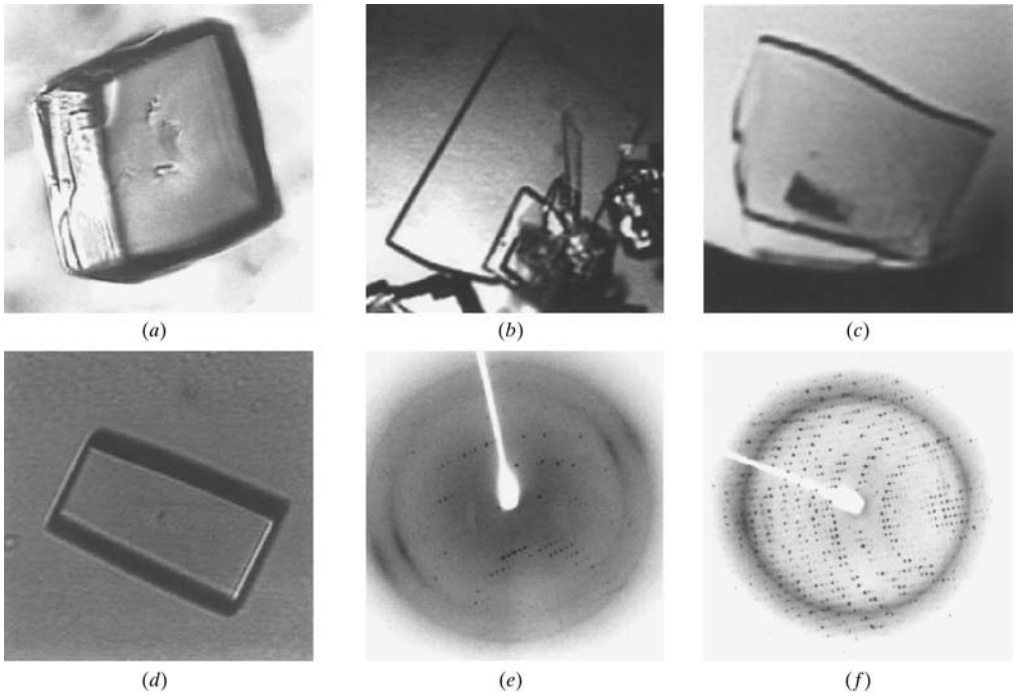

Figure 3
Typical crystals and diffraction patterns from crystals of HMG-D bound to different halogenated DNA fragments. Crystals are grown under similar conditions to the native crystals and data were collected under comparable conditions. (*a*) Iodo-T7 crystals (plates 0.2 × 0.2 × 0.1 mm). (*b*) Bromo-C8 (chunky plates of dimensions 0.4 × 0.25 × 0.05 mm). (*c*) Iodo-C2 (plates, 0.45 × 0.2 × 0.05 mm). (*d*) Bromo-C2 (blocks, 0.1 × 0.05 × 0.05 mm). Diffraction patterns from oscillation photographs of derivative crystals were obtained with an R-AXIS IIC imaging-plate area-detector system with cryocooling in a ∼123 K nitrogen stream. (*e*) Bromo-C8 oscillation picture (2° oscillation; 40 min exposure). (*f*) Iodo-C2 (1.5° oscillation, 25 min exposure).

4. Conclusions

HMG-D has been co-crystallized with DNA. Complete removal of the methionine-oxidized form of the protein was critical for successful co-crystallization. Substitution with halogenated bases resulted in several co-crystals of sufficient quality to be used as derivatives. Interestingly, the base substitutions near the ends of the DNA fragment and not those at positions in the middle of the binding site gave crystals which diffracted well. However, even these crystals were non-isomorphous with the native. MAD is the method of choice for obtaining phases for this structure, because good bromine-derivative crystals were obtained. Solution of this structure will reveal the detailed interactions at the protein–DNA interface of the HMG-D–DNA complex and possibly suggest which interactions allow these proteins to bind to a wide range of DNA sequences with moderate affinity.

We thank Dr A. Joachimiak for use of the R-AXIS IIC with

Table 2
Data-collection statistics of native data.

Resolution (Å)	Measured reflections	Unique reflections	Completeness (%)	$I > 3\sigma(I)$ (%)	R_{sym} (%)†
20–4.72		1164	97.1	97.0	3.5
4.72–3.76		1105	99.6	97.6	3.6
3.76–3.28		1090	98.9	96.4	4.3
3.28–2.98		1072	98.8	93.5	5.5
2.98–2.77		1049	98.0	87.8	7.7
2.77–2.61		1074	97.6	82.8	9.6
2.61–2.48		1017	97.1	75.6	12.0
2.48–2.37		1050	97.5	72.8	14.7
2.37–2.28		1038	97.0	66.0	18.2
2.28–2.20		1044	97.9	64.1	19.8
20–2.20	94446	10703	98.0	83.6	5.2

† R_{sym} (on intensity) = $\sum\sum|I(h)_i - \langle I(h)\rangle|/\sum\sum I(h)_i$, where $I(h)_i$ is the observed intensity and $\langle I(h)\rangle$ is the mean intensity of reflection h over all measurements of $I(h)$.

cryocooling prior to the installation of our cryocooling system and Dr A. Wang for use of oligonucleotides used in crystal screening. We appreciate the support from the NIH (Shannon Award to MEAC), American Cancer Society (MEAC), American Heart Association Grant-In-Aid (MEAC), Molecular Biophysics NIH pre-doctoral training grant (FVM and LKD) and a Colgate–Palmolive Research Award (JVS).

References

Balaeff, A., Churchill, M. E. A. & Schulten, K. (1998). *Proteins Struct. Funct. Genet.* **30**, 113–135.
Baxevanis, A. D. & Landsman, D. (1995). *Nucleic Acids Res.* **23**, 1604–1613.
Bianchi, M. E., Falciola, L., Ferrari, S. & Lilley, D. (1992). *EMBO J.* **11**, 1055–1063.
Bustin, M. & Reeves, R. (1996). *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 35–100.
Churchill, M. E. A., Changela, A., Dow, L. K. & Krieg, A. J. (1999). *Methods Enzymol.* **304**, 99–133.
Churchill, M. E. A., Jones, D. N. M., Glaser, T., Hefner, H., Searles, M. A. & Travers, A. A. (1995). *EMBO J.* **14**, 1264–1275.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dow, L., Changela, A., Hefner, H. & Churchill, M. (1997). *FEBS Lett.* **414**, 514–520.
Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovellbadge, R. & Bianchi, M. E. (1992). *EMBO J.* **11**, 4497–4506.
Giese, K., Cox, J. & Grosschedl, R. (1992). *Cell*, **69**, 185–195.
Grosschedl, R., Giese, K. & Pagel, J. (1994). *Trends Genet.* **10**, 94–100.
Hardman, C. H., Broadhurst, R. W., Raine, A. R. C., Grasser, K. D., Thomas, J. O. & Laue, E. D. (1995). *Biochemistry*, **34**, 16596–16607.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–522.
Houghten, R. A. & Li, C. H. (1979). *Anal. Biochem.* **98**, 36–46.
Jones, D. N. M., Searles, A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A. & Neuhaus, D. (1994). *Structure*, **2**, 609–627.
Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). *Nature (London)*, **376**, 791–795.
Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). *Nature (London)*, **389**, 251–60.
Ner, S. S. (1992). *Curr. Biol.* **2**, 208–210.
Ner, S. S. & Travers, A. A. (1994). *EMBO J.* **13**, 1817–1822.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Payet, D. & Travers, A. A. (1997). *J. Mol. Biol.* **266**, 66–75.
Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C. & Norman, D. G. (1993). *Nucleic Acids Res.* **21**, 3427–3436.
Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. & Wang, A. H. (1998). *Nature (London)*, **392**, 202–205.
Rogers, D. W. (1994). *Structure*, **2**, 1135–1140.
Stura, E. A. & Wilson, I. A. (1991). *J. Cryst. Growth*, **110**, 270–282.
Wagner, C. R., Hamana, K. & Elgin, S. (1992). *Mol. Cell. Biol.* **12**, 1915–1923.
Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A., Laue, E. D. & Thomas, J. O. (1993). *EMBO J.* **12**, 1311–1319.
Werner, M. H., Bianchi, M. E., Gronenborn, A. M. & Clore, G. M. (1995). *Biochemistry*, **34**, 11998–12004.
Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995). *Cell*, **81**, 705–714.
Wolfe, S. A., Ferentz, A. E., Grantcharova, V., Churchill, M. E. A. & Verdine, G. L. (1995). *Chem. Biol.* **2**, 213–221.
Wolffe, A. (1994). *Science*, **264**, 1100–1101.