

Isotropic Reconstruction of 3D Fluorescence Microscopy Images Using Convolutional Neural Networks

Martin Weigert^{1,2}(✉), Loic Royer^{1,2}, Florian Jug^{1,2}, and Gene Myers^{1,2}

¹ Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
mweigert@mpi-cbg.de

² Center for Systems Biology Dresden, Dresden, Germany

Abstract. Fluorescence microscopy images usually show severe anisotropy in axial versus lateral resolution. This hampers downstream processing, i.e. the automatic extraction of quantitative biological data. While deconvolution methods and other techniques to address this problem exist, they are either time consuming to apply or limited in their ability to remove anisotropy. We propose a method to recover isotropic resolution from readily acquired anisotropic data. We achieve this using a convolutional neural network that is trained end-to-end from the same anisotropic body of data we later apply the network to. The network effectively learns to restore the full isotropic resolution by restoring the image under a trained, sample specific image prior. We apply our method to 3 synthetic and 3 real datasets and show that our results improve on results from deconvolution and state-of-the-art super-resolution techniques. Finally, we demonstrate that a standard 3D segmentation pipeline performs on the output of our network with comparable accuracy as on the full isotropic data.

1 Introduction

Fluorescence microscopy is a standard tool for imaging biological samples [15]. Acquired images of confocal microscopes [3] as well as light-sheet microscopes [4], however, are inherently anisotropic, owing to a 3D optical point-spread function (PSF) that is elongated along the axial (z) direction which typically leads to a 2 to 4-fold lower resolution along this axis. Furthermore, due to the mechanical plane-by-plane acquisition modality of most microscopes, the axial sampling is reduced as well, further reducing the overall resolution by a factor of 4 to 8. These effects later render downstream data analysis, e.g. cell segmentation, difficult. To circumvent this problem, multiple techniques are known and used: Classical deconvolution methods [9, 12] are arguably the most common of these. They can be applied on already acquired data, however, their performance is typically inferior to other more complex techniques. Some confocal systems, e.g. when using two-photon excitation with high numerical aperture objectives and an isotropic axial sampling, can acquire almost isotropic volumes [3, 10] (cf. Fig. 3). Downsides are low acquisition speed, high photo toxicity/bleaching, and large file

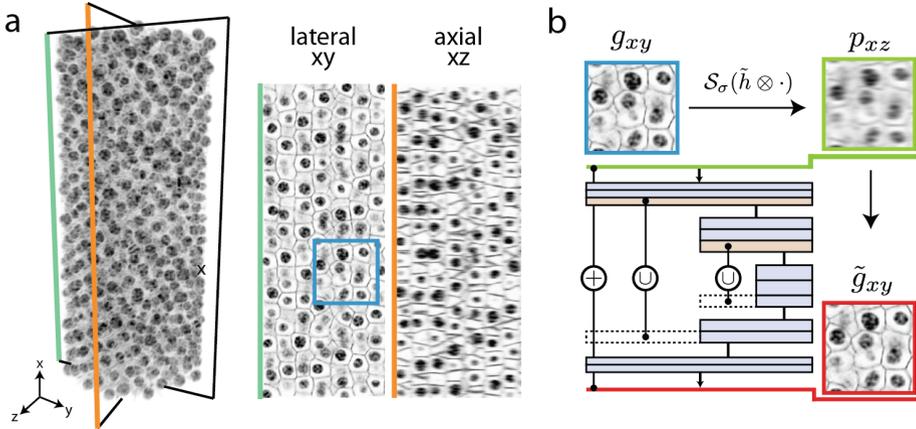


Fig. 1. (a) 3D images acquired on light microscopes are notoriously anisotropic due to axial undersampling and optical point spread function (PSF) anisotropy. (b) The IsoNet-2 architecture has a U-net [13] like topology and is trained to restore anisotropically blurred/downsampled lateral patches. After training it is applied to the axial views.

sizes. Light-sheet microscopes, instead, can improve axial resolution by imaging the sample from multiple sides (views). These views can then be registered and jointly deconvolved [11]. The disadvantage is the reduced effective acquisition speed and the need for a complex optical setup. A method that would allow to recover isotropic resolution from a single, anisotropic acquired microscopic 3D volume is therefore highly desirable and would likely impact the life-sciences in fundamental ways.

Here we propose a method to restore isotropic image volumes from anisotropic light-optical acquisitions with the help of convolutional networks without the need for additional ground truth training data. This can be understood as a combination of a super-resolution problem on subsampled data, and a deconvolution problem to counteract the microscope induced optical PSF. Our method takes two things into account: (i) the 3D image formation process in fluorescence microscopes, and (ii) the 3D structure of the optical PSF. We use and compare two convolutional network architectures that are trained end-to-end from the same anisotropic body of data we later apply the network to. During training, the network effectively learns a sample specific image prior it uses to deconvolve the images and restore full isotropic resolution.

Recently, neural networks have been shown to achieve remarkable results for super-resolution and image restoration on 2D natural images where sufficient ground truth data is available [2, 6]. For fluorescence microscopy data there is, unfortunately, no ground truth (GT) data available because it would essentially require to build an ideal and physically impossible microscope. Currently there is no network approach for recovering isotropic resolution from fluorescence microscopy images. Our work uses familiar network architectures [2, 13],

and then applies the concept of *self super-resolution* [5] by learning from the very same dataset for which we restore isotropic resolution.

2 Methods

Given the true fluorophore distribution $f(x, y, z)$ the acquired volumetric image g of a microscope can be approximated by the following process

$$g = \mathcal{P}[\mathcal{S}_\sigma(h \otimes f)] + \eta \quad (1)$$

where $h = h(x, y, z)$ is the point spread function (PSF) of the microscope, \otimes is the 3D convolution operation, \mathcal{S}_σ is the axial downsampling/slicing operator by a factor σ , \mathcal{P} is the signal dependent noise operator (e.g. poisson noise) and η is the detector noise. As the PSF is typically elongated along z and $\sigma > 1$, the lateral slices g_{xy} of the resulting volumetric images show a significant higher resolution and structural contrast compared to the axial slices g_{xz} and g_{yz} (cf. Fig. 1a).

2.1 Restoration via Convolutional Neural Networks

The predominant approach to invert the image formation process (1) is, in cases where it is possible, to acquire multiple viewing angles of the sample, and register and deconvolve these images by iterative methods without any sample specific image priors [9, 11, 12]. In contrast to these classical methods for image restoration, we here try to directly learn the mapping between blurred and downsampled images and its true underlying signal. As no ground truth for the true signal is available, we make use of the resolution anisotropy between lateral and axial slices and aim to restore lateral resolution along the axial direction. To this end, we apply an adapted version of the image formation model (1) to the lateral slices g_{xy} of a given volumetric image

$$p_{xy} = \mathcal{S}_\sigma(\tilde{h} \otimes g_{xy}) \quad (2)$$

with a suitable chosen 3d *rotated PSF* \tilde{h} . To learn the inverse mapping $p_{xy} \rightarrow g_{xy}$ we assemble lateral patches $(g_{xy}^n, p_{xy}^n)_{n \in \mathbb{N}}$ and train a fully convolutional neural network [8] to minimize the pixel wise PSNR loss

$$\mathcal{L} = \sum_n -[20 \log_{10} \max g_{xy}^n - 10 \log_{10} |g_{xy}^n - \tilde{g}_{xy}^n|^2] \quad (3)$$

where \tilde{g}_{xy}^n is the output of the network when applied to p_{xy}^n . For choosing the best \tilde{h} we consider the two choices (i) *full*: $\tilde{h} = h_{rot}$ where h_{rot} is a rotated version of the original PSF that is aligned with the lateral planes, and (ii) *split*: $\tilde{h} = h_{split}$ which is the solution to the deconvolution problem $h_{rot} = h_{iso} \otimes h_{split}$ and h_{iso} is the isotropic average of h . The later choice is motivated by the observation that convolving lateral slices with h_{split} leads to images with a resolution comparable to the axially ones. After training we apply the network on the unseen, anisotropically blurred, bicubic upsampled axial slices g_{xz} of the whole volume to get the final estimation output.

2.2 Network Architecture and Training

We propose and compare two learning strategies, *IsoNet-1* and *IsoNet-2*, that implement two different established network topologies. The notation for the specific layers is as follows: $C_{n,w,h}$ for a convolutional layer with n filters of size (w, h) , $M_{p,q}$ for max pooling by a factor of (p, q) , and $U_{p,q}$ for upsampling by a factor of (p, q) . In conjunction with the two different methods of training data generation (*full*, *split*), the specific topologies are:

IsoNet-1. Which is the proposed network architecture of [1] used for super resolution: $C_{64,9,9} - C_{32,5,5} - C_{1,5,5} - C_{1,1,1}$. Here the first layer acts as a feature extractor whose output is mapped nonlinearly to the resulting image estimate by the subsequent layers. After each convolutional layer a rectifying activation function (ReLU) is applied.

IsoNet-2. Which is similar to the proposed network architecture of [13] for segmentation which consists of a contractive part: $C_{16,7,7} - M_{2,2} - C_{32,7,7} - M_{2,2} - C_{64,7,7} - U_{2,2} - C_{32,7,7} - U_{2,2} - C_{16,7,7} - C_{1,1,1}$ and symmetric skip connections. The contractive part of the network learns sparse representations of the input whereas skip connections are sensitive to image details (cf. Fig. 1b). In contrast to [13], however, the network learns the residual to the input.

For all datasets, both architectures were trained for 100 epochs with the Adam optimizer [7] and a learning rate $5 \cdot 10^{-3}$. We furthermore use a dropout of 20% throughout and apply data augmentation (flipped and rotated images) where it is compatible with the symmetries of the PSF (i.e. whenever the latter commutes with the augmentation symmetry).

3 Results

3.1 Synthetic Data

We use 3 different types of synthetic datasets of size 512^3 that resemble typical biological structures, as shown in Fig. 2: The uppermost row shows small axial crops from a volume containing about 1500 simulated nuclei. The middle row shows crops of membrane structures as they are frequently seen in tightly packed cell epithelia. The last row shows both, simulated cell nuclei and surrounding labeled membranes. All volumes were created in-silico by combining plausible structure distributions, perlin-noise based textures and realistic camera noise. Note that the first column shows the ground truth images that were used to generate the isotropic ground truth, by convolving with the isotropic PSF, and the blurred images that were subsampled and convolved with realistic PSFs in order to resemble microscopic data. This third column (blurred) is then used as the input to all our and other tested methods. The subsequent 6 columns show the results of (i) Richardson-Lucy deconvolution [9], (ii) pure SRCNN [1], i.e. disregarding the PSF, (iii) the IsoNet-1 using the full PSF, (iv) the IsoNet-1 using the anisotropic component of the PSF h_{split} , (v) the IsoNet-2 using the

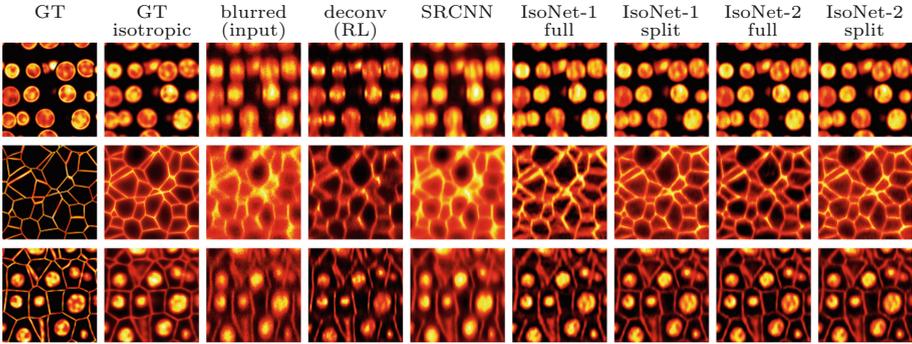


Fig. 2. Comparison of results on synthetic data. Rows show axial slices of 3D nuclei data, membrane data, and a combined dataset, respectively. The columns are: (i) ground truth phantom fluorophore densities, (ii) the same ground truth convolved with an isotropic PSF, (iii) anisotropically blurred isotropic GT image (the input images to all remaining columns), (iv) deconvolved images using Richardson-Lucy [9, 12], (v) SRCNN [1], (vi) IsoNet-1 with one (full) PSF, (vii) IsoNet-1 making use of the split PSFs, (viii/ix) IsoNet-2 with full PSF and split PSFs, respectively.

Table 1. Computed PSNR values against isotropic GT (upper rows), and against GT (lower rows). PSF types are: gaussian ($\sigma_{xy}/\sigma_z = 2/8$); confocal with numerical aperture $NA = 1.1$; light-sheet with $NA_{\text{detect}} = 0.8$ and $NA_{\text{illum}} = 0.1$. Bold values indicate best. Standard deviation in brackets ($n = 10$).

Volume (PSF/scale)	Blurred (input)	Deconv (RL)	SRCNN	IsoNet-1		IsoNet-2	
				Full	Split	Full	Split
Nuclei (gaussian/8)	25.84(0.17)	27.48(0.16)	25.89(0.19)	32.18(0.18)	32.47(0.18)	35.11(0.18)	35.61(0.18)
	24.09(0.10)	25.88(0.11)	24.15(0.10)	27.74(0.10)	27.53(0.10)	29.51(0.10)	28.84(0.10)
Membranes (confocal/4)	21.83(0.13)	18.52(0.14)	21.69(0.13)	19.55(0.12)	26.19(0.13)	19.14(0.12)	27.33(0.14)
	15.95(0.01)	16.48(0.01)	15.94(0.01)	16.84(0.01)	16.49(0.01)	17.09(0.01)	16.62(0.01)
Nuclei+memb. (light-sheet/6)	28.13(0.37)	25.00(0.37)	28.69(0.39)	25.59(0.38)	30.23(0.37)	25.40(0.39)	30.95(0.36)
	24.61(0.52)	26.57(0.51)	24.59(0.52)	26.86(0.51)	26.07(0.51)	27.85(0.51)	26.66(0.51)

full PSF, and (vi) the IsoNet-2 using the split PSF. In addition to the visuals given in the figure, Table 1 compares the PSNR of the full volumes with the two ground truth versions, averaged over 10 different randomly created stacks per dataset type. As can be seen, our method performs significantly ($p < 0.01$) best in all cases. Note that failing to incorporate the PSF (as with pure SRCNN) results in an inferior reconstruction.

Simple 3D Segmentation. To provide a simple example of how the improved image quality helps downstream processing we applied a standard 3D segmentation pipeline on the simulated nuclei data (cf. Fig. 2), consisting of 3 simple steps:

First, we apply a global threshold using the *intermodes* method [14]. Then, holes in thresholded image regions are closed. Finally, cells that clump together are separated by applying a 3D watershed algorithm on the 3D euclidian distance transform. This pipeline is freely available to a large audience in tools like Fiji or KNIME. We applied this pipeline to the isotropic ground truth data, the blurred and subsampled input data, and the result produced by the *IsoNet-2*. As evaluation metric we used SEG (ISBI Tracking Challenge 2013), the average intersection over union of matching cells when compared to the ground truth labels, that takes values in $[0, 1]$, where 1 corresponds to a perfect voxel-wise matching. The results for the different conditions $SEG_{GT} = 0.923$ (isotropic ground truth), $SEG_{blurred} = 0.742$ (blurred input), and $SEG_{IsoNet-2} = 0.913$ (network output), demonstrate the effectiveness of our approach.

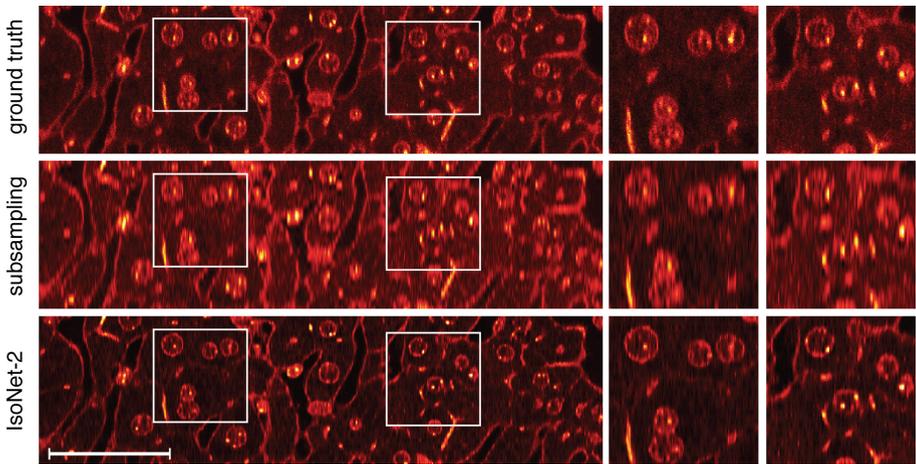


Fig. 3. Results on fluorescence microscopy images of liver tissue (data taken from [10]). Nuclei (DAPI) and membrane (Phalloidin) staining of hepatocytes, imaged with a two-photon confocal microscope (excitation wavelength 780 nm, NA = 1.3, oil immersion, $n = 1.49$). We start from an isotropic acquisition (ground truth), simulate an anisotropic acquisition (by taking every 8th slice), and compare the isotropic image to the *IsoNet-2* recovered image. Scalebar is $50 \mu\text{m}$.

3.2 Real Data

Furthermore, we validate our approach on confocal and light-sheet microscopy data and demonstrate the perceptual isotropy of the recovered stacks. First we show that artificially subsampled two-photon confocal acquisitions can be made isotropic using *IsoNet-2*. As can be seen in Fig. 3 the original isotropic data is nearly perfectly recovered from the 8-fold subsampled data (by taking every 8th axial slice). Second, we show that single view light-sheet acquisitions can be made isotropic. Figure 4 shows stacks from two different sample recordings

where we trained *IsoNet-2* to restore the raw axial (yz) slices. The final results exhibit perceptual sharpness close to that of the higher quality raw lateral (xy) slices, even when compared to multi deconvolution, demonstrating the ability to restore isotropic resolution from a single volume in different experimental settings.

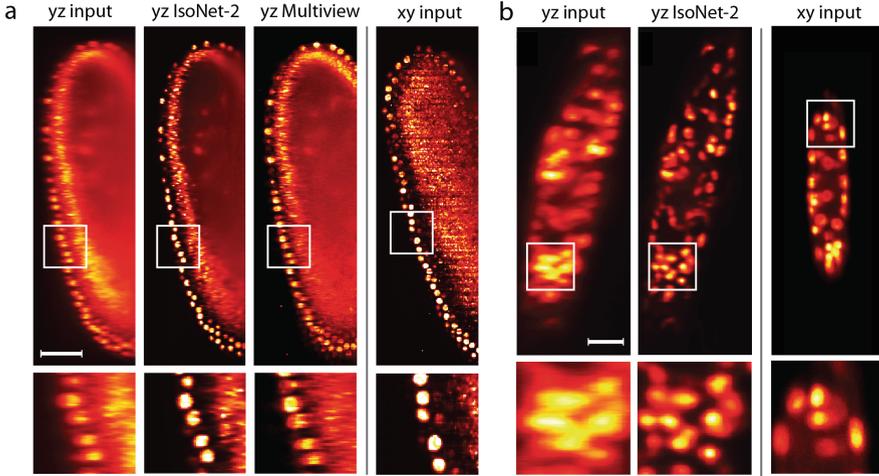


Fig. 4. IsoNet-2 applied to (a) *Drosophila* and (b) *C. elegans* volumes. The image quality of the recovered IsoNet-2 axial (yz) slices is significantly improved and shows similar isotropic resolution when compared to the lateral (xy) slices. In (b) we additionally compare to the result of multiview deconvolution [11]. Scalebar (a) $50\ \mu\text{m}$, (b) $10\ \mu\text{m}$.

4 Discussion

We presented a method to enhance the axial resolution in volumetric microscopy images by reconstructing isotropic 3D data from non-isotropic acquisitions with convolutional neural networks. Training is performed unsupervised and end-to-end, on the same anisotropic image data for which we recover isotropy. We demonstrated our approach on 3 synthetic and 3 real datasets and compared our results to the ones from classical deconvolution [9, 12] and state-of-the-art super resolution methods. We further showed that a standard 3D segmentation pipeline performed on outputs of IsoNet-2 are essentially as good as on full isotropic data. It seems apparent that approaches like the ones we suggest bear a huge potential to make microscopic data acquisition significantly more efficient. For the liver data, for example, we show (Fig. 3) that only 12.5% of the data yields isotropic reconstructions that appear on par with the full isotropic volumes. This would potentially reduce memory and time requirements as well as laser induced fluorophore and sample damage by the same factor. Still, this method can, of

course, not fill in missing information: If the axial sample rate would drop below the Shannon limit (with respect to the smallest structures we are interested in resolving), the proposed networks will not be able to recover the data. Source code will be released at <https://github.com/maweigert/isonet>.

Acknowledgments. We thank V. Stamataki, C. Schmied (Tomancak lab), S. Merret and S. Janosch (Sarov Group), H.A. Morales-Navarrete (Zerial lab) for providing the datasets, and U. Schmidt (all MPI-CBG) for helpful feedback. Datasets were recorded by the Light Microscopy Facility (LMF) of MPI-CBG.

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). doi:[10.1007/978-3-319-10593-2_13](https://doi.org/10.1007/978-3-319-10593-2_13)
2. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)
3. Economo, M.N., Clack, N.G., Lavis, L.D., Gerfen, C.R., Svoboda, K., Myers, E.W., Chandrashekar, J.: A platform for brain-wide imaging and reconstruction of individual neurons. *Elife* **5**, e10566 (2016)
4. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., Stelzer, E.H.K.: Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science (New York, N.Y.)* **305**, 1007–1009 (2004)
5. Jog, A., Carass, A., Prince, J.L.: Self super-resolution for magnetic resonance images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 553–560. Springer, Cham (2016). doi:[10.1007/978-3-319-46726-9_64](https://doi.org/10.1007/978-3-319-46726-9_64)
6. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
7. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
9. Lucy, L.B.: An iterative technique for the rectification of observed distributions. *The Astron. J.* **79**, 745 (1974)
10. Morales-Navarrete, H., Segovia-Miranda, F., Klukowski, P., Meyer, K., Nonaka, H., Marsico, G., Chernykh, M., Kalaidzidis, A., Zerial, M., Kalaidzidis, Y.: A versatile pipeline for the multi-scale digital reconstruction and quantitative analysis of 3D tissue architecture. *Elife* **4**, e11214 (2015)
11. Preibisch, S., Amat, F., Stamataki, E., Sarov, M., Singer, R.H., Myers, E., Tomancak, P.: Efficient bayesian-based multiview deconvolution. *Nat. Methods* **11**, 645–648 (2014)
12. Richardson, W.H.: Bayesian-based iterative method of image restoration. *JOSA* **62**(1), 55–59 (1972)

13. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
14. Sezgin, M., Sankur, B.I.: Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **13**(1), 146–168 (2004)
15. Tsien, R.Y.: The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998)