X-ray Diffraction and the Discovery of the Structure of DNA:

A tutorial and historical account of James Watson's and Francis Crick's use of X-ray diffraction in their discovery of the double helix structure of DNA

David T. Crouse
Department of Electrical Engineering
City College of New York, City University of New York, Steinman Hall, Room 614
Convent Ave. at 140th St., New York, NY 10031
Email: dcrouse@ccny.cuny.edu, Phone: (212) 650-5330, Fax: (212) 650-8249

Supplemental Materials for Students

Step 1b: Structure Factor of a Double Uniform Helix

First, the structure factor for one uniform helix should be calculated. To do this, divide up one twist of the helix into N discrete scattering elements, each having an identical form factor F. Figure S1.1 shows a top-down view of the positions of the N scattering elements. Use this information to construct the displacement vectors $\mathbf{d}_j$ in cylindrical coordinates (express $\mathbf{K}$ in cylindrical coordinates as well) and insert this information into Eq. (4).

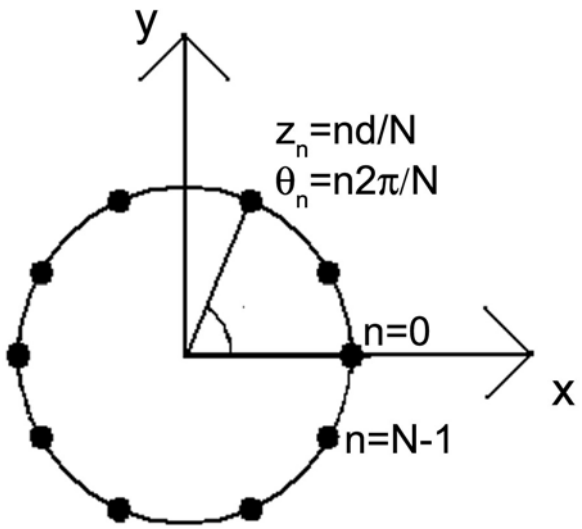

Fig. S1.1 A top view of on complete twist of the single helix composed of N scattering elements.

Now consider splitting up the uniform helix into progressively finer divisions, thereby letting $N \to \infty$ but at the same time with $F \to 0$ such that $N F \to 1$. The summation you just derived can then be expressed as an integral over θ. After using the definition of Bessel functions, you will get the desired expression for $S_m\left(K_\rho, \theta_K, K_z\right)$.

For the addition of the second helix, use Rosalind Franklin's and Raymond Gosling's information that the second helix is displaced in the *z*-direction by $\delta = \frac{3}{8}d$ and include these additional *N* scattering elements in the summation. The z_n values and θ_n values of this second set of scattering elements are related to the first set by:

$$z_n' = z_n + \frac{3}{8}d \qquad \theta_n' = \theta_n \qquad \text{with} \quad n = 0 \to N-1 \qquad \text{(S1.1)}$$

Then do the same as before with $N \to \infty$ and $F \to 0$ such that $N F \to 1$ and the desired expression for the structure factor for the double uniform helix will be obtained.

Step 2: The reciprocal lattice for an array of infinite continuous planes with spacing c

The calculation of the interference function for a lattice of points along the *z*-axis with spacing c is straightforward but a few hints about the calculation of the structure factor for an infinite plane might be useful. It is best to start with Eq. (3) dealing with a continuous distribution of scattering elements but because the shape is an infinite plane oriented in the *x*-*y* plane, the volume integral turns into an area integral with an areal density σ in place of the volume density ρ of scattering elements resulting in the following expression:

$$S(\mathbf{K}) = \int_{plane} F\,\sigma \exp(i\mathbf{K}\cdot\mathbf{d})dA \qquad \text{(S1.2)}$$

where $\mathbf{d} = x\hat{x} + y\hat{y}$. Keeping in mind the definition of Dirac-Delta functions, the above integral should be easy to calculate. If you get any infinite values, just assume because of physical reasons that the integral stays finite; in reality the number of x-ray scattering elements are finite in number and the incident x-ray intensity is finite, hence the diffracted x-ray intensity, which is proportional to $S(\mathbf{K})$, must remain finite.

Step 3: Constructing the Complete Reciprocal Space Structure Using the Convolution Theorem: The Cochran-Crick-Vand Theorem.

The only hint that can be given for this step without giving away the answer is to use the reciprocal space lattice of the lattice of planes (representing the *z*-coordinates of the phosphates) for $G(\widetilde{K})$ and the reciprocal space lattice of the double *uniform* helix (representing the *x* and *y* coordinates of the phosphates) for $F(K-\widetilde{K})$.

Supplemental Materials for Instructors

Interference Function for a 1D Array of Points

The interference function for a lattice of points with spacing d is:

$$I(\mathbf{K}) = \sum_{m=0}^{\infty} \exp(i\mathbf{K}\cdot\mathbf{T}) = \lim_{N\to\infty} \sum_{m=0}^{N-1} \exp(iK_z dm)$$

$$= \lim_{N\to\infty} \exp(iK_z(N-1)d/2) \frac{\sin\left(K_z \frac{d}{2} N\right)}{\sin\left(K_z \frac{d}{2}\right)} = \begin{cases} \phi N & \text{if} \quad K_z = \frac{2\pi}{d}\text{integer} \\ 0 & \text{otherwise} \end{cases} \qquad \text{(S2.1)}$$

where $\mathbf{K} = K_x\hat{x} + K_y\hat{y} + K_z\hat{z}$, $\mathbf{T} = m\,d\,\hat{z}$ are the translation vectors in reciprocal space and coordinate space respectively with $m = \text{integer}$. Also, ϕ is a phase term that cancels when evaluating the intensity of the scattered X-ray beams. It is seen that for the intensity of diffracted X-ray beams to be nonzero, K_z has to be the following:

$$K_z = \frac{2\pi}{d} m \qquad m = \text{integer} \qquad \text{(S2.2)}$$

There are no restrictions on K_x and K_y. Thus, it is seen that the reciprocal lattice of a lattice of points is a series of planes along the z axis with spacing $2\pi/d$.

Step 1b: Structure Factor of a Double Uniform Helix

As was stated in the Supplemental Materials for Students, you first need to calculate the structure factor for a single uniform helix. For the single uniform helix, first consider splitting up one twist of the helix it up into *N* scattering elements, each having an identical form factor *F,*

and later letting $N \to \infty$ and $F \to 0$ such that $N F \to 1$ and changing the summation to an integral. For one twist being composed of *N* scattering elements, it is seen from Fig. S1.1 that *z* and θ vary linearly with *n* according to the following equations:

$$z_n = \frac{d}{N} n \qquad \text{and} \qquad \theta_n = \frac{2\pi}{N} n \qquad \text{with} \quad n = 0 \to N-1 \tag{S2.3}$$

In the summation for the structure factor Eq. (4), it is the dot product $\mathbf{K} \cdot \mathbf{d}_n$ that needs to be evaluated. Using cylindrical coordinates, we have:

$$\mathbf{K} = K_\rho \hat{\rho}(\theta_K) + K_z \hat{z} \tag{S2.4}$$

$$\mathbf{d}_n = r\hat{r}(\theta_n) + z_n \hat{z} \tag{S2.5}$$

$$\mathbf{K} \cdot \mathbf{d}_n = rK_\rho \hat{\rho}(\theta_K) \cdot \hat{r}(\theta_n) + K_z z_n = rK_\rho \cos(\theta_n - \theta_K) + K_z z_n \tag{S2.6}$$

where θ_n is the azimuthal angle of the nth scattering element and θ_K is the azimuthal angle of $\mathbf{K}$. The equation for the structure factor then is:

$$S_m\left(K_\rho, \theta_K, K_z = \frac{2\pi}{d} m\right) = \sum_{n=0}^{N-1} F \exp(i\mathbf{K} \cdot \mathbf{d}_n) = \sum_{n=0}^{N-1} F \exp\left(i\left(rK_\rho \cos(\theta_n - \theta_K) + K_z z_n\right)\right) \tag{S2.7}$$

Using Eq. (6) for K_z, the equation that was derived for z_n, and using the relationship between Δn (which is equal to 1) and $\Delta\theta$ that Eq. (S2.3) provides, Eq. (S2.7) can be expressed as in a different form:

$$S_m\left(K_\rho, \theta_K, K_z = \frac{2\pi}{d} m\right) = \frac{NF}{2\pi} \exp(i\theta_K m) \sum_\theta \exp\left(i\left(rK_\rho \cos(\theta - \theta_K) + (\theta - \theta_K)m\right)\right) \Delta\theta \tag{S2.8}$$

Letting $N \to \infty$ and $F \to 0$ with $N F \to 1$, we can change the summation given by Eq. (S2.8) into the following integral:

$$S_m\left(K_\rho, \theta_K, K_z = \frac{2\pi}{d} m\right) = \frac{\exp(i\theta_K m)}{2\pi} \int_0^{2\pi} \exp\left(i\left(rK_\rho \cos(\tilde{\theta}) + m\tilde{\theta}\right)\right) d\tilde{\theta} \tag{S2.9}$$

where $\widetilde{\theta} = \theta - \theta_K$. Using the fact that $i = \exp(i\pi/2)$ and the definition of Bessel functions, the desired expression for the structure factor of a single uniform helix can be obtained:

$$S_m\left(K_\rho, \theta_K, K_z = \frac{2\pi}{d}m\right) = J_m(rK_\rho)\exp\left(im\left(\theta_K + \frac{\pi}{2}\right)\right) \tag{S2.10}$$

To add in the second helix oriented in the way specified by Franklin and Gosling, start with the summation Eq. (S2.7) describing the structure factor for one helix of *N* scattering elements and add in another set of *N* scattering elements with the following z'_n and θ'_n dependencies:

$$z'_n = \frac{d}{N}n + \frac{3}{8}d \qquad \text{and} \qquad \theta'_n = \theta_n = \frac{2\pi}{N}n \qquad \text{with} \qquad n = 0 \to N-1 \tag{S2.11}$$

The structure factor for the double helix of 2*N* scattering elements oriented is then:

$$\widetilde{S}_m = \sum_{n=0}^{N-1} F\exp\left(i\left(rK_\rho\cos(\theta_n - \theta_K) + K_z z_n\right)\right) + \sum_{n=0}^{N-1} F\exp\left(i\left(rK_\rho\cos(\theta_n - \theta_K) + K_z\left(z_n + \frac{3}{8}d\right)\right)\right)$$

$$= \left(1 + \exp\left(i\frac{3\pi}{4}m\right)\right)\sum_{n=0}^{N-1} F\exp\left(i\left(rK_\rho\cos(\theta_n - \theta_K) + \theta_n m\right)\right) \tag{S2.12}$$

As before, changing the summation index from *n* to θ, letting $N \to \infty$ and $F \to 0$ with $NF \to 1$, and using the definition of Bessel functions results in the equation:

$$\widetilde{S}_m\left(K_\rho, \theta_K, K_z = \frac{2\pi}{d}m\right) = \left(1 + \exp\left(i\frac{3\pi}{4}m\right)\right)J_m(rK_\rho)\exp\left(im\left(\theta_K + \frac{\pi}{2}\right)\right) \tag{S2.13}$$

Step 2: The reciprocal lattice for an array of infinite continuous planes with spacing c

For the interference function for a periodic array of points with spacing *c*, the exact same calculation as was done for Step 1a is performed but with a spacing c instead of d. It is seen that for the X-ray diffracted intensity to be nonzero, K_z has to be the following:

$$K_z = \frac{2\pi}{c} m \qquad m = \text{integer} \tag{S2.14}$$

The calculation of the interference function places no restrictions on K_x and K_y. Restrictions on K_x and K_y will be a result of the structure factor for the infinite plane.

For the structure factor of an infinite plane oriented in the *x*-*y* plane, start with Eq. (3) but use the real density σ in place of the volume density ρ of scattering elements resulting in the following:

$$S(\mathbf{K}) = \int_{plane} F\,\sigma \exp(i\mathbf{K}\cdot\mathbf{d})dA = F\,\sigma\left(\int_{plane}\exp(iK_x x)dx\right)\left(\int_{plane}\exp(iK_y y)dy\right) \tag{S2.15}$$

where $\mathbf{d} = x\hat{x} + y\hat{y}$. The integrals within the brackets are recognized as delta functions:

$$\int_{-\infty}^{\infty} e^{iK_x x}dx = 2\pi\delta(K_x) \qquad \text{and} \qquad \int_{-\infty}^{\infty} e^{iK_y y}dy = 2\pi\delta(K_y) \tag{S2.16}$$

In reality, these integrals will remain finite but will be nonzero only when K_x and K_y are both equal to zero.

Finally combining these two results above, the reciprocal lattice of a lattice of infinite planes is a set of points with $\mathbf{K}$ vectors:

$$\mathbf{K} = \frac{2\pi}{c} m\,\hat{z} \qquad m = \text{integer} \tag{S2.17}$$

Step 3: The Complete Reciprocal Space Lattice Using the Convolution Theorem

In the convolution integral (Eq. (11)), use the reciprocal space lattice of the lattice of planes (representing the *z*-coordinates of the phosphates) for $G(\widetilde{K})$ and the reciprocal space lattice of

the double uniform helix (representing the *x* and *y* coordinates of the phosphates) for $F\left(K-\widetilde{K}\right)$. Assuming that there are P phosphates per twist (i.e., $c=d/P$), $G\left(\widetilde{K}\right)$ is:

$$G\left(\widetilde{K}\right)=\begin{cases}\phi N \quad if \quad K_z=\dfrac{2\pi}{c}n=\dfrac{2\pi}{(d/P)}n=\dfrac{2\pi}{d}Pn & n=\text{integer}\\ 0 \quad otherwise\end{cases} \tag{S2.18}$$

Because the set of K_z values in Eq. (S2.14) is a subset of the K_z values given by Eq. (S2.2), no new points in reciprocal space are produced with K_z values other than those given by Eq. (S2.2). Equation (11) now becomes a summation:

$$\Omega\left(rK_\rho,\theta_K,K_z=\frac{2\pi}{d}m\right)=\int_{-\infty}^{\infty}G\left(\widetilde{K}\right)F\left(K-\widetilde{K}\right)d\widetilde{K}$$

$$=\sum_{n=-\infty}^{\infty}\left(1+\exp\left(i\frac{3\pi}{4}(m-Pn)\right)\right)J_{(m-Pn)}\left(rK_\rho\right)\exp\left(i(m-Pn)\left(\theta_K+\frac{\pi}{2}\right)\right) \tag{S2.19}$$

As stated in Section 2, the result you get for $\Omega(K)$ will be oscillatory about the K_z axis (i.e., as a function of θ_K) but these oscillations will not be observed because in the experiment, the DNA strands may be rotating or you have many DNA strands with random orientations. What is desired is the absolute magnitude of $|\Omega(K)|^2$ averaged over $0\le\theta_K\le 2\pi$. This quantity is easy to calculate:

$$\left|\Omega\left(rK_\rho,K_z=\frac{2\pi}{d}m\right)\right|^2=\frac{1}{2\pi}\int_0^{2\pi}\Omega\left(rK_\rho,\theta_K,K_z=\frac{2\pi}{d}m\right)\Omega^*\left(rK_\rho,\theta_K,K_z=\frac{2\pi}{d}m\right)d\theta_K$$

$$=4\sum_{n=-\infty}^{\infty}\cos^2\left(\frac{3\pi}{8}(m-Pn)\right)J^2_{(m-Pn)}\left(rK_\rho\right) \tag{S2.20}$$

Now to show the important relations given by Eq. (14) and (15) we have:

$$\left|\Omega_{avg}\left(rK_\rho,K_z=2\pi(-m)/d\right)\right|^2=4\sum_{n=-\infty}^{\infty}\cos^2\left(\frac{3\pi}{8}(-m-Pn)\right)J^2_{(-m-Pn)}\left(rK_\rho\right)$$

$$= 4\sum_{n=-\infty}^{\infty} \cos^2\left(\frac{3\pi}{8}(m+Pn)\right)\left((-1)^{-m-Pn}\right)^2 J^2_{(m+Pn)}(rK_\rho)$$

$$= 4\sum_{n=-\infty}^{\infty} \cos^2\left(\frac{3\pi}{8}(m-Pn)\right) J^2_{(m-Pn)}(rK_\rho)$$

$$= \left|\Omega_{avg}\left(rK_\rho, K_z = 2\pi m/d\right)\right|^2 \qquad \text{(S2.21)}$$

where in the second to the last step, the standard technique of redefining the summation index was done. Now for Eq. (15) we have:

$$\left|\Omega_{avg}\left(rK_\rho, K_z = 2\pi(m+P)/d\right)\right|^2 = 4\sum_{n=-\infty}^{\infty} \cos^2\left(\frac{3\pi}{8}((m+P)-Pn)\right) J^2_{((m+P)-Pn)}(rK_\rho)$$

$$= 4\sum_{n=-\infty}^{\infty} \cos^2\left(\frac{3\pi}{8}(m-P(n-1))\right) J^2_{(m-P(n-1))}(rK_\rho)$$

$$= 4\sum_{\tilde{n}=-\infty}^{\infty} \cos^2\left(\frac{3\pi}{8}(m-P\tilde{n})\right) J^2_{(m-P\tilde{n})}(rK_\rho)$$

$$= \left|\Omega_{avg}\left(rK_\rho, K_z = 2\pi m/d\right)\right|^2 \qquad \text{(S2.22)}$$

It is easy to see why the fifth layer line will be slightly stronger than other layer lines because two terms with J_5^2 will occur in the summation and the cosine terms in front of these two fifth order Bessel functions are close to 1. Using the same analysis, it is easy to see why the forth and sixth layer lines are weak.

Figure S2.1 shows $\left|\Omega_{avg}\left(rK_\rho, K_z = 2\pi m/d\right)\right|^2$ for $m = 0 \to 5$. Table S2.1 lists the values $\left(rK_\rho\right)_m^{1st\ \max}$ and $\left(rK_\rho\right)_m^{2nd\ \max}$ that produce the first and second local maximum for $\left|\Omega_{avg}\left(rK_\rho, K_z = 2\pi m/d\right)\right|^2$ respectively for each m. Table S2.1 also lists the theoretical comparison ratios $\left(rK_\rho\right)_m^{1st\ \max} \Big/ \left(rK_\rho\right)_{m=1}^{1st\ \max}$ that will be compared with the experimentally obtained values.

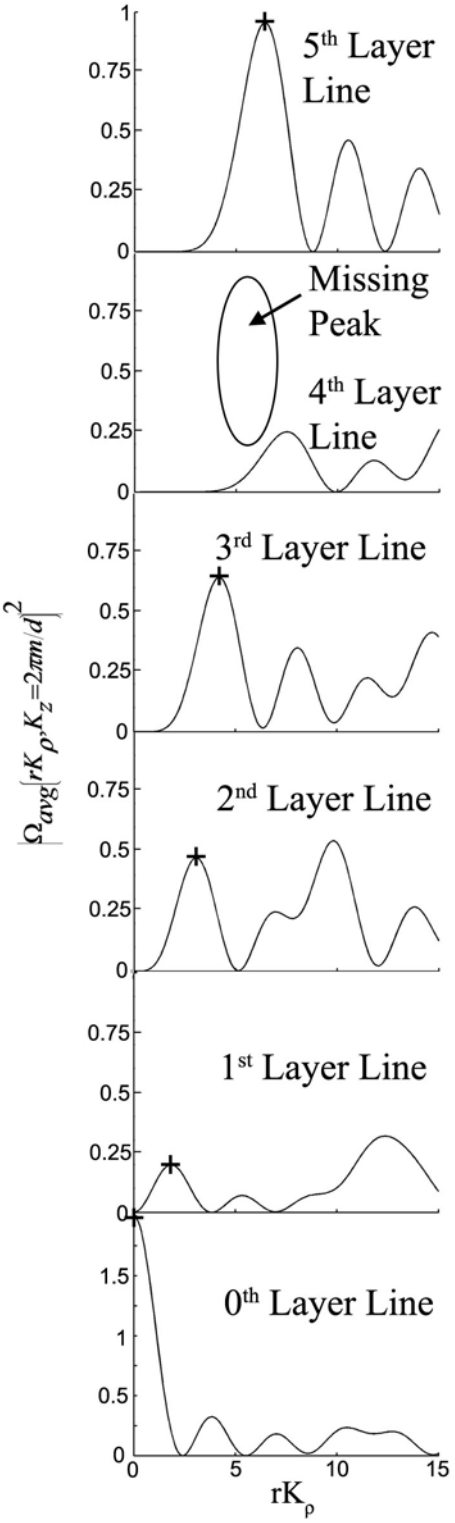

Fig. S2.1 The reciprocal space lattice layers averaged over $\theta_K = 0 \to 2\pi$ as a function of rK_ρ. It is seen that the fourth layer line is weak and missing the first peak and that the fifth layer line is strong.

Table S2.1. K_z values, and local maxima $\left(rK_\rho\right)_m^{\max}$ of $\left|\Omega_{avg}\left(K_\rho, K_z = 2\pi m/d\right)\right|$ for $m = 0 \to 5$ are given.

m	Peak 1 $\left(rK_\rho\right)_m^{\max}$	Peak 2 $\left(rK_\rho\right)_m^{\max}$	Comparison Ratios $\left(rK_\rho\right)_m^{\max} / \left(rK_\rho\right)_1^{\max}$
0	0	3.8	0
1	1.8	5.3	1
2	3.0	6.8	1.67
3	4.2	8.0	2.33
4	Missing	7.5	Missing
5	6.4	10.5	3.56

Step 5: Comparison of Theory and Experiment

Even though the diffraction pattern is a bit unclear in Fig. 3, we can distinguish four sets of diffraction peaks besides the undiffracted beam at different x_m and y_m values. We have the following data for the diffraction peaks labeled $0 \to 5$. Expanding the figure to a full page and measuring the x_m and y_m values yields the values listed in Table S2.2. The camera constant that is used depends on the degree that the X-ray diffraction photograph is expanded; if you zoom in on the photograph by a factor of two, the camera constant has to be increased by a factor of two as well. Let us assume we have a camera constant $s = 5\ cm$ for the x_m and y_m values yields the values listed in Table S2.2. Also, copper Kα x-rays are used that have a wavelength 1.54 Å. The values of x_m and y_m and Eqs. (16) - (17) allow the values $\left|K_z\right|_m$, $\left|K_\rho\right|_m$ and $\left|K_\rho\right|_m / \left|K_\rho\right|_{m=1}$ to be calculated and are listed in Table S2.2. The normalized values $\left|K_\rho\right|_m / \left|K_\rho\right|_1$ in Table S2.1 should match up with the normalized values in Table S2.2 in the following way:

1. Peak 0 is the undiffracted beam and corresponds to $\left|K_z\right|_{m=0}$.

2. Peak 1 corresponds to $\left(K_z\right)_{m=1} = \frac{2\pi}{d}1$ and all following $\left|K_\rho\right|_m$ values are normalized relative to $\left|K_\rho\right|_{m=1}$.

3. Peak 2 corresponds to $\left(K_z\right)_{m=2} = \frac{2\pi}{d}2$ and the value $\left|K_\rho\right|_2 \Big/ \left|K_\rho\right|_1$ should be equal to 1.67.

4. Peak 3 corresponds to $\left(K_z\right)_{m=3} = \frac{2\pi}{d}3$ and the value $\left|K_\rho\right|_3 \Big/ \left|K_\rho\right|_1$ should be equal to 2.33.

5. There appears to be a gap where Peak 4 should be and this agrees with the theoretical results that state that the fourth layer line is weak.

6. Peak 5 corresponds to $\left(K_z\right)_{m=1} = \frac{2\pi}{d}5$ and the value $\left|K_\rho\right|_5 \Big/ \left|K_\rho\right|_1$ should be equal to 3.56.

It is seen that these measured and theoretical values largely match, therefore Properties 1-3 of the DNA structure are verified.

Now using Eq. (16) and the fact that $\left|K_z\right|_{m=1} = 2\pi/d$ to calculate d yields $d = 33.5$ Å.

Finally, using Eqs. (16) - (17) to calculate $\left|K_\rho\right|_{m=1}$ and the value of $\left(rK_\rho\right)_{m=1}^{1st\ \max}$ listed in Table S2.1 to calculate r yields $r = 11.3$ Å. These values are fairly close to the values given by Watson, Crick, Wilkins and Franklin and considering that we do not have the original X-ray diffraction photograph, the level of agreement is acceptable.

Table S2.2. Diffraction pattern values x_m and y_m and other important values. We are assuming $s = 5\ cm$

Peak	x_m	y_m	$\left\|K_z\right\|_m$	$\left\|K_\rho\right\|_m$	$\left\|K_\rho\right\|_m \Big/ \left\|K_\rho\right\|_1$
0	0	0	0	0	0
1	0.195	0.230	0.187	0.159	1
2	0.335	0.455	0.369	0.273	1.72
3	0.520	0.715	0.575	0.423	2.66
4	Missing				
5	0.800	1.190	0.933	0.647	4.07