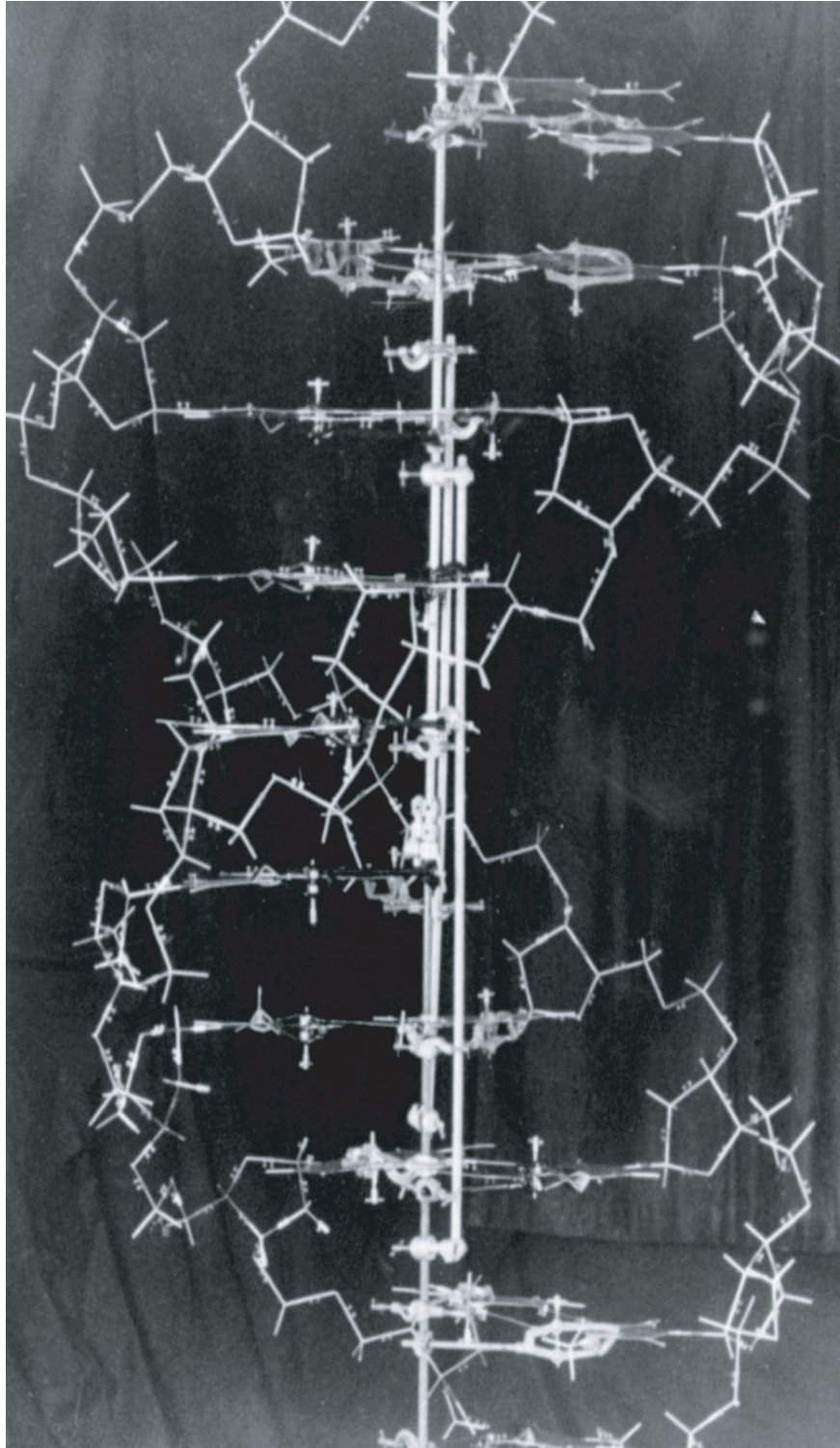
What is a Scientific Model?

Nathan Camillo Sidoli

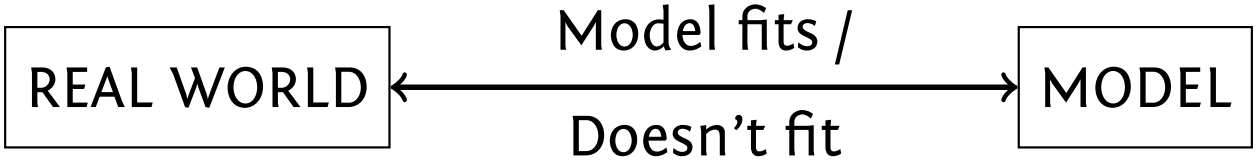
Waseda University, SILS,

Omnibus: Invitation to Natural Sciences



A naive view of scientific models

A simplistic view of the way scientists work with models might be that they observe the physical world, measure it, see what it's like, and then make a model that is an accurate representation of the world.



In this view, the real world is transparent, we have immediate access to it, and we can see exactly what it's like. This is a view of science as analogous to *mapmaking*.

Unfortunately, scientists usually don't have direct access to the things they want to study; they may very far away, or very old, or very small.

Scientists have to interpret the world by producing facts about it, which then take on the form of data.


早稲田
出入口1
都電
早稲田駅
新目白通り
1丁目
西早稲田
早稲田
鶴巻町西
区立西早稲田
地域交流館
西早稲田
パークタワー
西早稲田東
ウイング
櫻荘
一番館
りそな銀行
エル
カスティーヨ
スカイコート
ヌーベル早稲田
ビットイン西
早稲田駐車場
リサイクル
ステーション
わせだ
中国料理
花月
アートハビ
テーション
(株)矢田
進幸堂
山浦印刷(株)
(有)松豊
堂製袋
有紀書房
双葉印刷(株)
寺尾有料
駐車場
ドルチェ
目白台
松田製本
ソナーレ
早稲田
アムス西
早稲田
シェモア
コーポ岩立
エバー
グリーン
セブンイレブン
西早稲田1丁目店
グランシティ
早稲田
藤和シティ
コープ早稲田
まいば
すけっと
ParkWell
早稲田弐番館
のがみ
矯正歯科
エスコート
西早稲田
早稲田大学
西早稲田
ビルディング
19号館
早稲田大
図書館
アサヒパレス
早稲田
西早稲田
一同
早稲田
アパート
西早稲田
保育園
早稲田大学
22号館
イルマーレ
西早稲田
西早稲田南
ウイング
フレール
早稲田
サンハウス
WUEOkumaAvenue
プリマス
ナガタ
S・K早稲田
観音寺
ベナント
オギワラ
フラット
ワセダ
早稲田社会
教育センター
リーガロイヤル
ホテル東京
ブラザー
横山
(株)板橋造花
完之荘
寿司源
くるみ荘
ハウゼ井上
(株)デ
文
20号館
大隈庭園
Uni-
Shop&Cafe125
早稲田大学
教職員健康
管理室
24号館
文
シルキー
早稲田
ラフィーネ
早稲田
吉田畳店
生協本部
早稲田大坪内博士
記念演劇博物館
文
17号館
文
文
14号館
銅清本田
板金工業所
ゲーム
インABC
グランド
坂フジタ
カーム
スクエア
サナエプリント
センター
松の湯
早稲田大学29-7
号館・研究室
ファミリーマート
早稲田大学
早稲田
キャンパス
文
1号館
現代政治
経済研究所
シャンブルN
ニュー
ライフ光栄
コーポS・U
ユニハイム
コスモス
早稲田大学29-4
号館研究室
早稲田大学29-6
号館研究室
早大甘泉寮
29-3号館
文
法学部
會津博物館
大隈講堂
文
早大演劇
研究会
(有)森林
印刷所
等和荘
早稲田大学
エクステンション
センター本館
CasadelSol
エスペース
ワセダ
原嶋刺繍店
大隈記念
タワー
小野梓
記念館
ワセダ
エルドラド
(株)成文堂
ドミール
立山
グランベル
西早稲田
プレシャス
ハイネス
早稲田
アサショー
ビル早稲田

Images, maps and models

Images are a type of two-dimensional scale model. A map is more abstract than an image, but less abstract than a theoretical model.

The map contains a bunch of information that *isn't there* in the real world, but is about the real world. It is a representation of the real word, but it is obviously not the same thing as what it represents.

- It contains a *similarity of spatial structure* with part of the real world.
- It contains symbols whose meaning is socially constructed, and it is embedded in practices of *social convention*.
- It contains only certain *selected features* of the real world.
- It also contains some *mistakes*, or misleading signs.

Is there such a thing as a *perfect map*?

Analog models

Scienists often use analogies to try to *understand* the physical world. For example, Descartes argued that an animal's body was like a machine, Rutherford argued that the atom was like the solar system.

Analog models are usually useful in the early stages of research, when scientists are trying to understand new phenomena. At later stages, when they are set to evaluate the model, an analog model is less useful.

Analog model

An analog model is something that a scientist uses as a metaphor or analogy for an object. It is used to help people think about a new object using other objects that they are already familiar with.

Nevertheless, when scientists talk about *models*, they are usually not referring to analog models.

Scale models

Sometimes scientists work with physical models of the subject that they are studying. These are a type of scale model in the ordinary sense of the word, just as we talk about a model airplane or a dollhouse.

Scale models are used in science and engineering as a way to present ideas or theories and as a research tool. It is much easier, or cheaper, to carry out tests on scale models. Also, the structure of some objects can best be understood with scale models, because they are so big, or so small. (solartoscale.com, scaleofuniverse.com)

Scale model

A scale model is a physical representation of the three-dimensional structure of an object.

Again, when scientists use the term *model*, they generally do not mean scale models.

Theoretical models and hypotheses

A theoretical model does not exist anywhere other than in our minds, or as the abstract subject of verbal or mathematical description, or in a computer as the initial conditions and subsequent states of a program, etc.

- Sometimes it's possible to build a scale model of a theoretical model, but not always.

The vast majority of the theoretical models that scientists work with remain as speculative hypotheses. Many of them can be discarded on the basis of things that are *already known* from observations or experiments.

In order for a model to be treated as the representation of a fact, it has to be checked against the real world. But we cannot do this directly. Instead, we have to generate *predictions from the model*, and *data from the real world*, and compare these against each other.

An example: The molecular structure of DNA

In the early 1950s, there was a race between a number of scientists to determine the molecular structure of deoxyribonucleic acid (DNA). At the time, not everyone realized that this was an important race. There had been a few experiments that seemed to indicate that DNA was the carrier of genetic information, but not everyone was convinced – proteins were regarded as another possibility.

This story involves a number of different people with widely different backgrounds, training and ability. Some of them were young students and postdocs, others were established stars, at the top of their field.

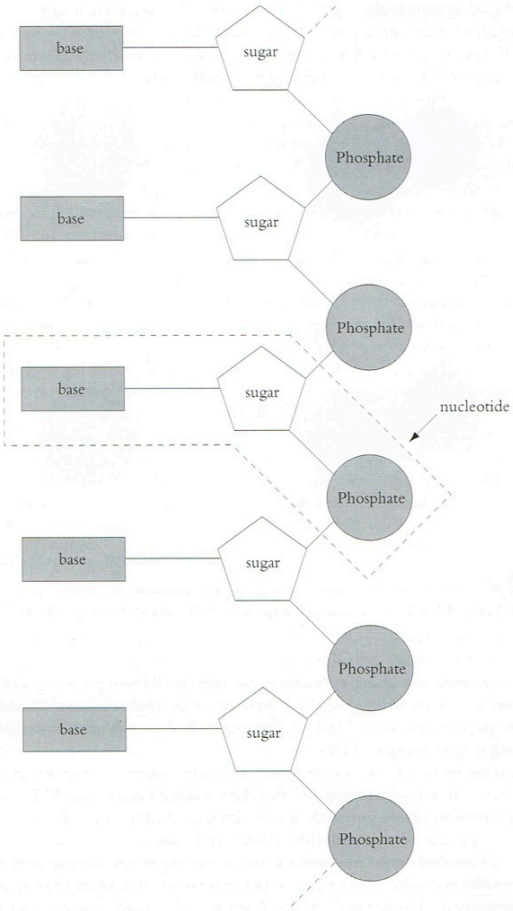
In a surprising twist, the structure was actually determined and published by the most unlikely pair, although it required the work of everyone involved.

The state of knowledge in 1951

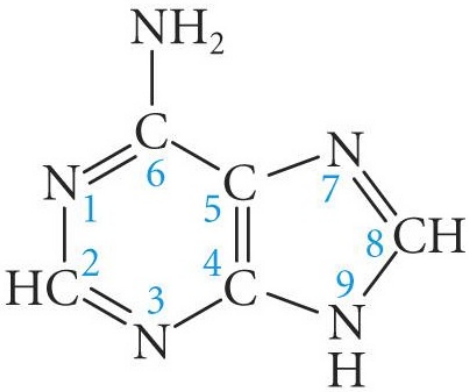
In 1951, a DNA molecule was thought to consist of one or more chains of nucleotides, called a polynucleotyde.

The nucleotides were known to be linked by a sugar and phosphate backbone. Hanging off of the backbone was a series of bases. In many ways, the structure was assumed to be similar to the structure of proteins, which was also largely unknown at this time.

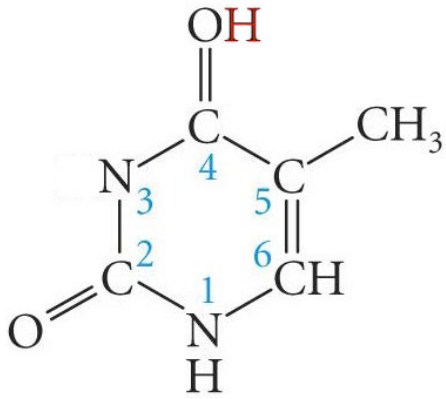
The chemical structure of the bases was also known. The bases were known to be two pyrimidines, cytosine (C) and thymine (T), and two purines, adenine (A) and guanine (G). (But the configuration of T and G came in two forms.)



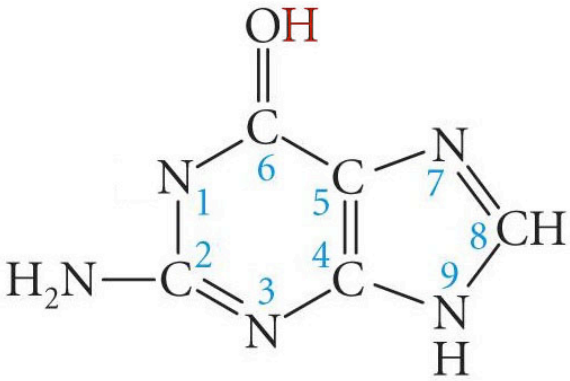
Knowledge of the nucleic bases in 1951



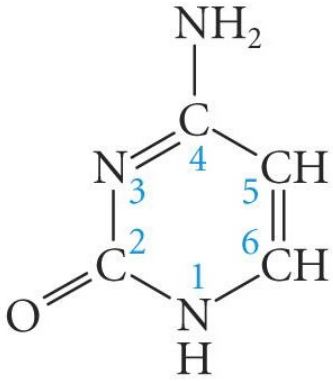
Adenine (A)



Thymine (T)



Guanine (G)

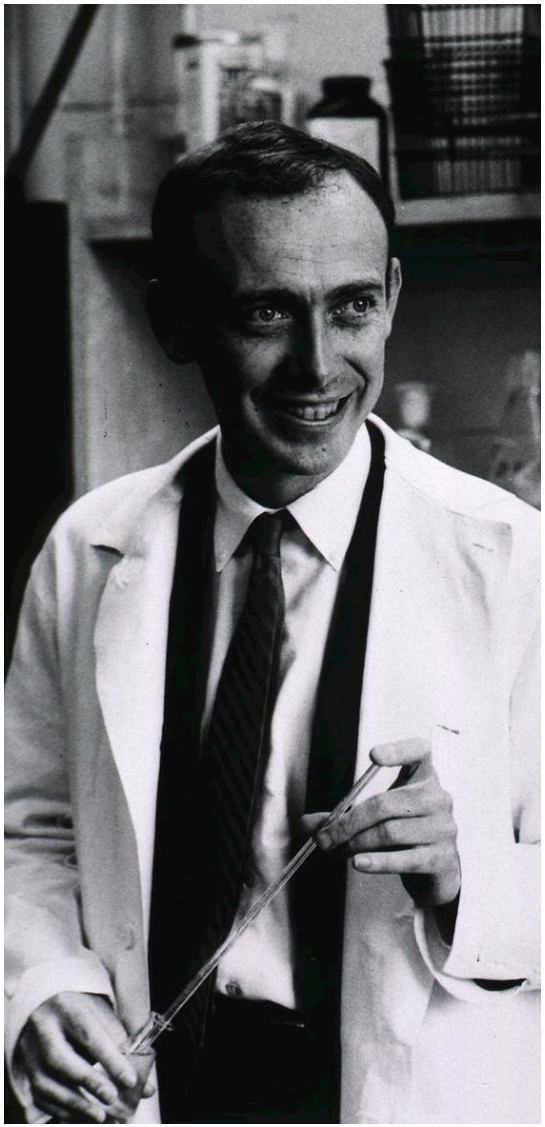


Cytosine (C)

Purines

Pyrimidines

Watson arrives at the Cavendish, 1951



James Watson (1928–) finished his PhD in 1950, working under Salvidor Luria on phage genetics.

He spent some time at Copenhagen, studying biochemistry, and some time in Naples, reading genetics papers. He heard a talk by Maurice Wilkins (1916–2004), King's College, London, on using x-rays to study the structure of DNA.

He decided to go the Cavendish, Cambridge, where they were studying big molecules with x-rays. He asked Luria to help him get a postdoctoral fellowship there. At the Cavendish, he met Francis Crick (1916–2004), who was an older PhD student, working on applying physics to organic molecules.

Wilkins' x-ray image of DNA

When Watson heard Wilkins' talk in 1950, he was especially struck by an image that Wilkins showed, which had been produced by x-rays that had been passed through a DNA molecule interacting with a photographic plate.

This image impressed on Watson the need to learn *x-ray crystallography* in order to try to determine the structure of the molecule. This was a technique that Watson had previously known little about.

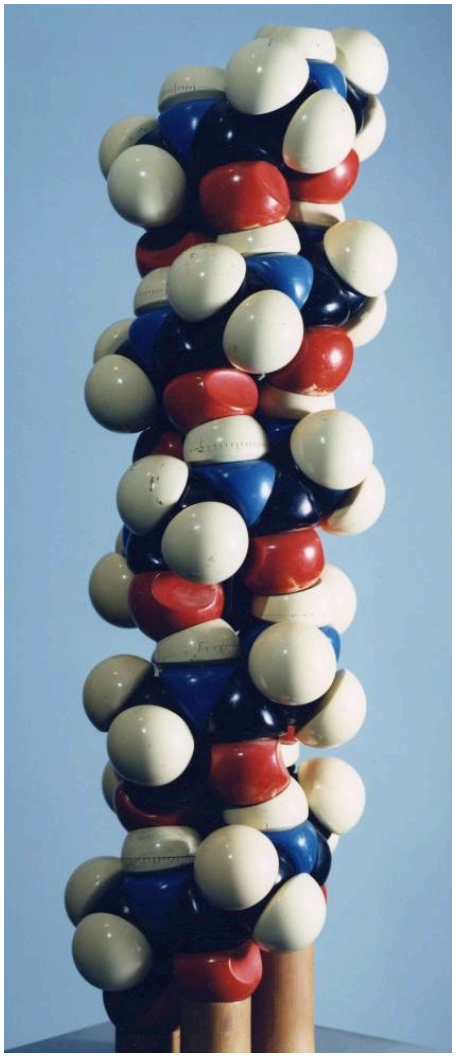
Pauling cracks the α-helix

Linus Pauling (1901–1994), a professor at CalTech, was one of the top structural chemists in the world. He wrote the 1939 book *The Nature of the Chemical Bond*.

His work helped lay the groundwork for figuring out the structure of large molecules, such as proteins.

In 1948, while he was laid up with a cold, he drew a polypeptide chain of roughly correct dimensions on a strip of paper and folded it into a helix maintaining the correct chemical bonds.

In 1951, he published a paper (with Robert Corey) giving the complete chemical structure of the α-helix, a key component of protein structure, using x-ray evidence and exhibited with a physical model.



CalTech model, 1951

DNA at the Cavendish

With the publication of Pauling's paper, Watson and Crick realized both (1) that the Pauling's model-building approach might work for DNA and (2) that its structure might be helical. That is, they used the α-helix as an analog model.

The head of their laboratory, Sir Laurence Bragg, an x-ray crystallographer, realized that the work on DNA might lead to something important and gave Watson and Crick permission to work on the topic, as long as the people at King's College, London, didn't mind.

Wilkins and his group at King's College had been working on x-raying DNA for some time, but Wilkins was getting fed up with the problem, due to interpersonal difficulties that he was having with one of his coworkers.

Franklin and the King's group

In January 1951, Rosalind Franklin (1920–1958) began working on making x-ray images of DNA at King's College. Franklin was one of the world's experts on x-ray imaging, but she soon ran into difficulties with Wilkins, with whom she did not get along.

Franklin had been led to believe she would have DNA all to herself and she disagreed with Wilkins about technical issues.

Soon there were two, competing, groups working on x-ray crystallography at King's College. Franklin's group had exclusive access to the better samples of DNA.

The 3-chain model

Crick made his first major contribution with a mathematical theory of how x-rays are diffracted by helically shaped molecules.

Watson went down to London to listen to a talk by Franklin about her work. Watson didn't take notes at talks, so based on what he could remember, Crick worked out that there would be between two and four chains in the model.

They decided to go with three chains, and put the bases hanging out on the outside – like a protein. Using wire, and specially made metal plates they constructed a scale model in under a month.

They invited the King's group up to Cambridge to see the model, but Franklin took one look at it and pointed out that there were not enough places for water to bind with it. Watson had misremembered what she has said about how much water was involved! It was a theoretical model that could be discarded on the basis of known facts.

Taking a break

It was such an embarrassment that Bragg forbid Watson and Crick to do any more work on DNA. Watson started working on x-raying the tobacco mosaic virus; Crick went back to work on his PhD.

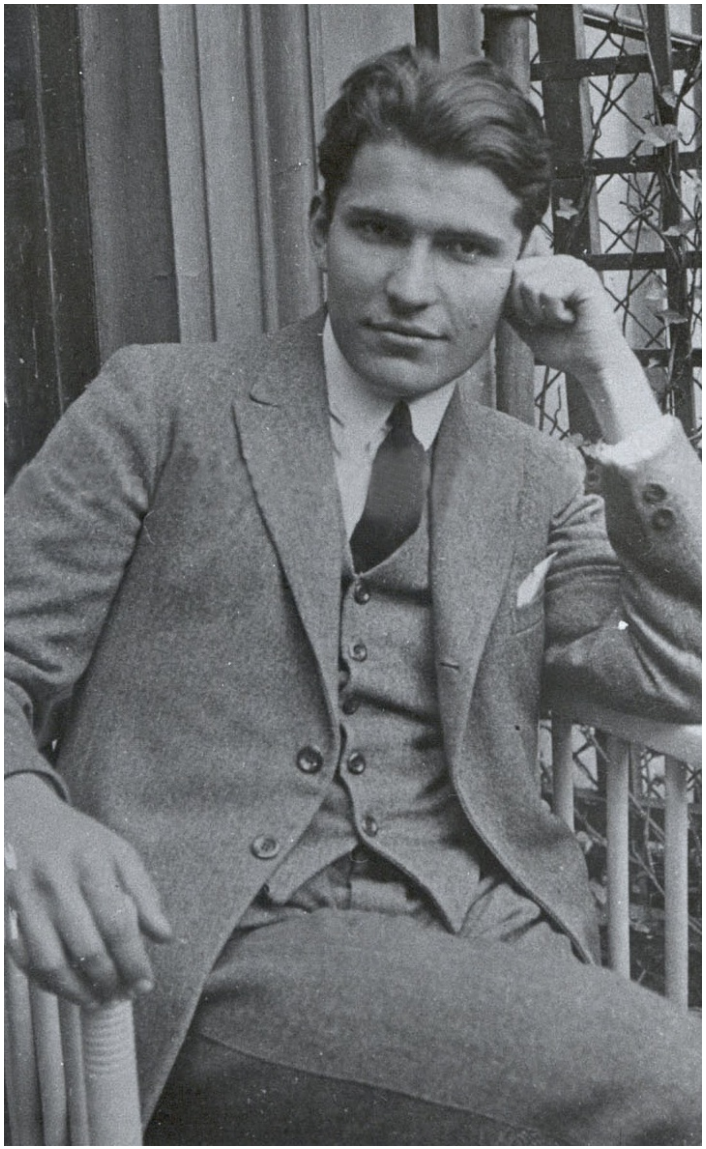
But other people were still working on DNA:

- Work that was going on at Cold Spring Harbor, NY, pointed even more securely to the importance of DNA in genetics.
- Erwin Chargaff published some results about the relative quantities of the base pairs.
- Pauling published a structure of DNA.

As soon as Watson and Crick read Pauling's paper, they realized that the proposed structure was wrong, but they also knew that Pauling would figure out his own mistake soon and would then turn all his energy towards the structure of DNA.

This was enough for Bragg to let them get back to work on DNA.

Chargaff's rule



Erwin Chargaff (1905–2002) was a brilliant Jewish-Austrian biochemist who had emigrated to the US as a result of Nazi policies.

He had come to the conclusion early on that DNA was the genetic material and began to study it using traditional biochemical techniques, such as breaking it down into its constituent parts.

He showed that, in all samples of DNA, $\%A = \%T$ and $\%G = \%C$. He explained this to Watson and Crick in 1952, but they seemed not to understand the significance of these findings.

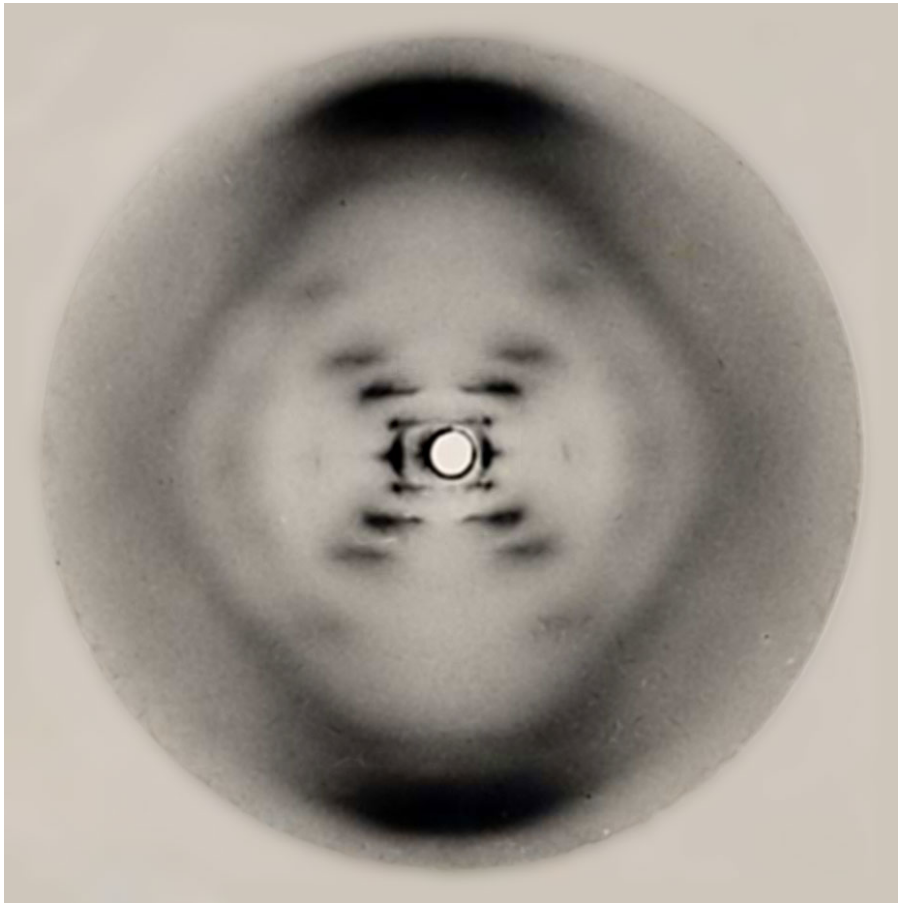
The 2-chain model

Watson went down to London to enlist the help of Wilkins and Franklin. Franklin, unimpressed with Watson's abilities, turned him down. Wilkins did not feel he could get involved with DNA, but he did show Watson one of Frankin's new photographs, *without her permission*. This alerted Watson to the existence of an unpublished research report – which he then obtained by round-about means.

Franklin's group had taken clear photos of B-form DNA, which contains more water than A-form DNA. The photos showed a clear x-shaped pattern, which Crick's theory predicted must result from a helix.

Watson returned to Cambridge and began model-building. He decided to try a 2-chain model, on the vague analogy that many things in nature come in pairs. They now decided to try a model in which the bases faced inward – as opposed to proteins, in which the amino acids hang off the α-helix. He choose to link like-with-like bases (ignoring Chargaff's results). These were all *guesses*.

Wilkins' 1950 x-ray image,
A-form DNA



Franklin's 1952 x-ray image,
B-form DNA

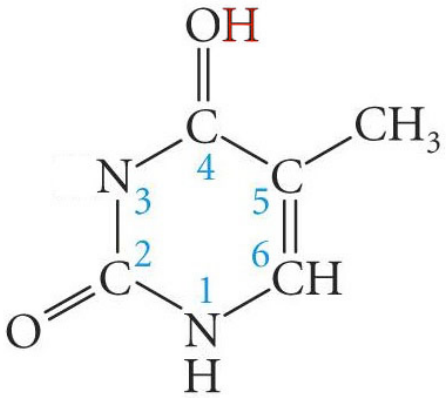
Cardboard nucleic acids

Watson built a model with like-to-like paring of the bases in the inside of two helixes. The bases aren't all the same size, so this produced bulges in the backbone.

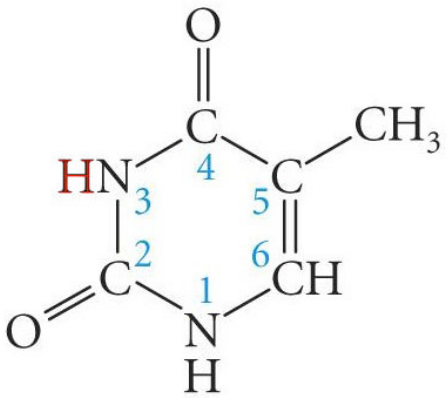
He showed his idea to Jerry Donohue (1920–1985) an American theoretical chemist, who was visiting the Cavendish and sharing an office with Watson and Crick. Donohue immediately pointed out that the diagram that Watson had found for guanine (G) and thymine (T) in the reference books was, from a quantum-mechanical perspective, highly unlikely. He suggested using the keto, as opposed to enol, forms.

Since there were no metal models of these forms available, Waston made his own models with stiff cardboard. By playing around with these scale models of the correct forms of the nucleic acids, he could see that the adenine and thymine (A-T) and guanine and cytosine (G-C) paired together in the same shape. This *explained* Chargaff's results.

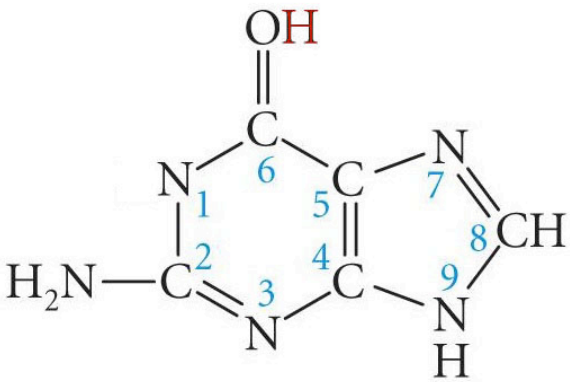
Enol and keto nucleic acids



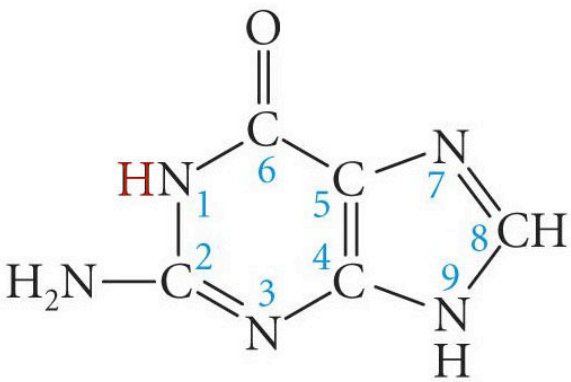
Thymine (T)



Thymine (T)



Guanine (G)



Guanine (G)

Enol forms

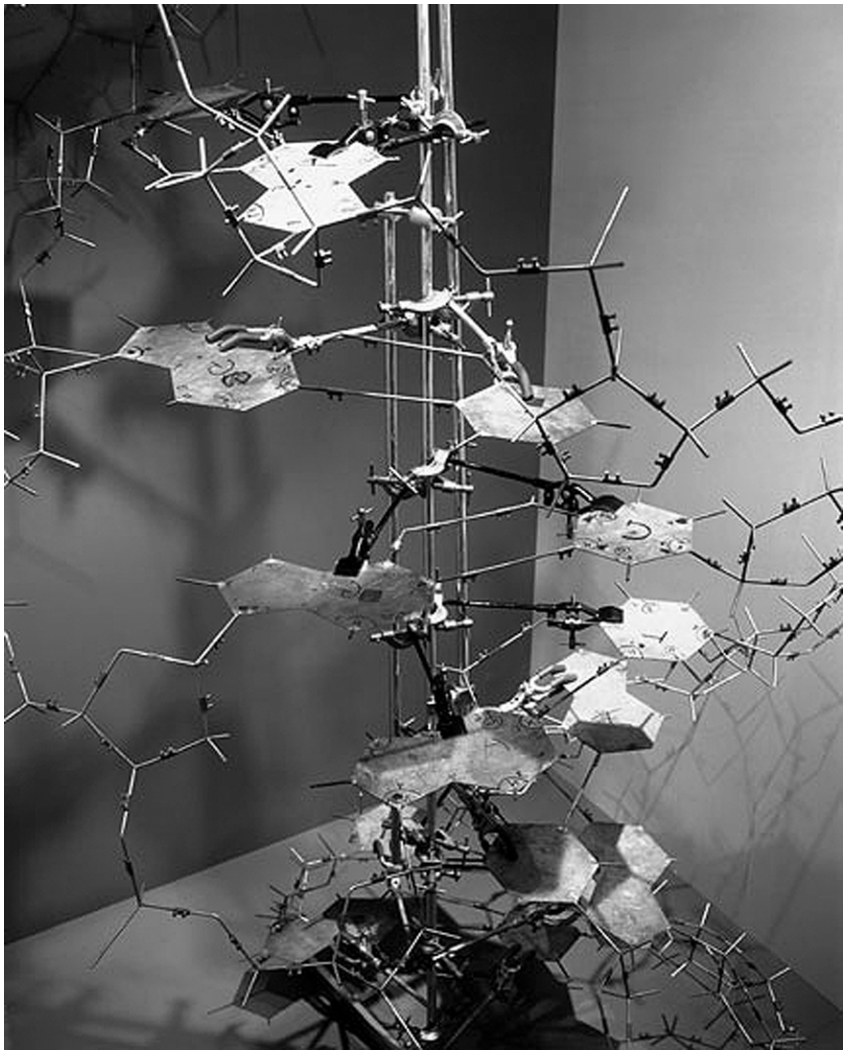
Keto forms

A scale model

When the machine shop delivered the correct bases, Watson and Crick began to assemble a scale model. They used a *plumb line* and a *measuring stick* to make sure the backbone was in the shape required by Crick's theory.

They used Pauling's *The Nature of the Chemical Bond* to confirm that all of the bonds represented in the model were theoretically correct.

Again, they invited up the King's group, who agreed that the model fit the data. They published a one-page paper explaining the model.



equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

WHILE the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the *n*th layer line being proportional to the square of J_n, the *n*th order Bessel function. A straight line may be drawn approximately through

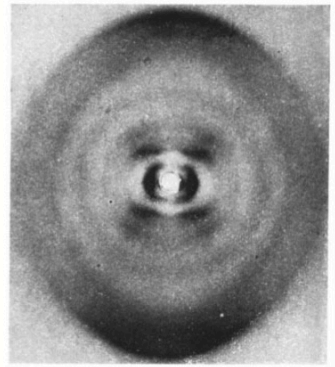


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats *n* times along the helix there will be a meridional reflexion (J_0^2) on the *n*th layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[5] being to reproduce the intensity distribution about the origin around the new origin, on the *n*th layer line, corresponding to *C* in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

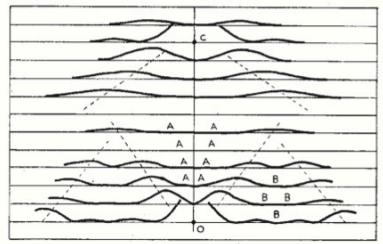
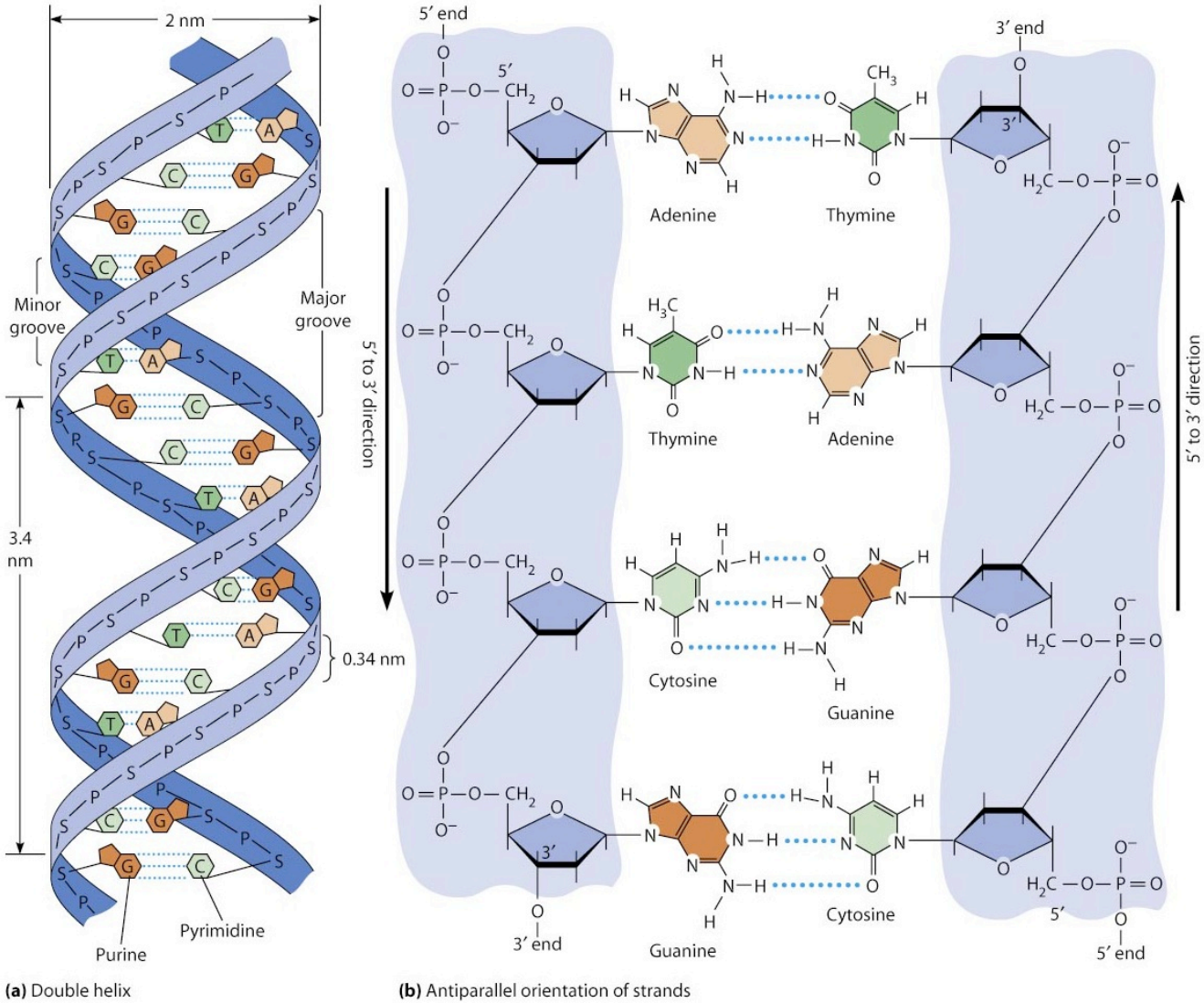


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About *C* on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

A theoretical model



(a) Double helix

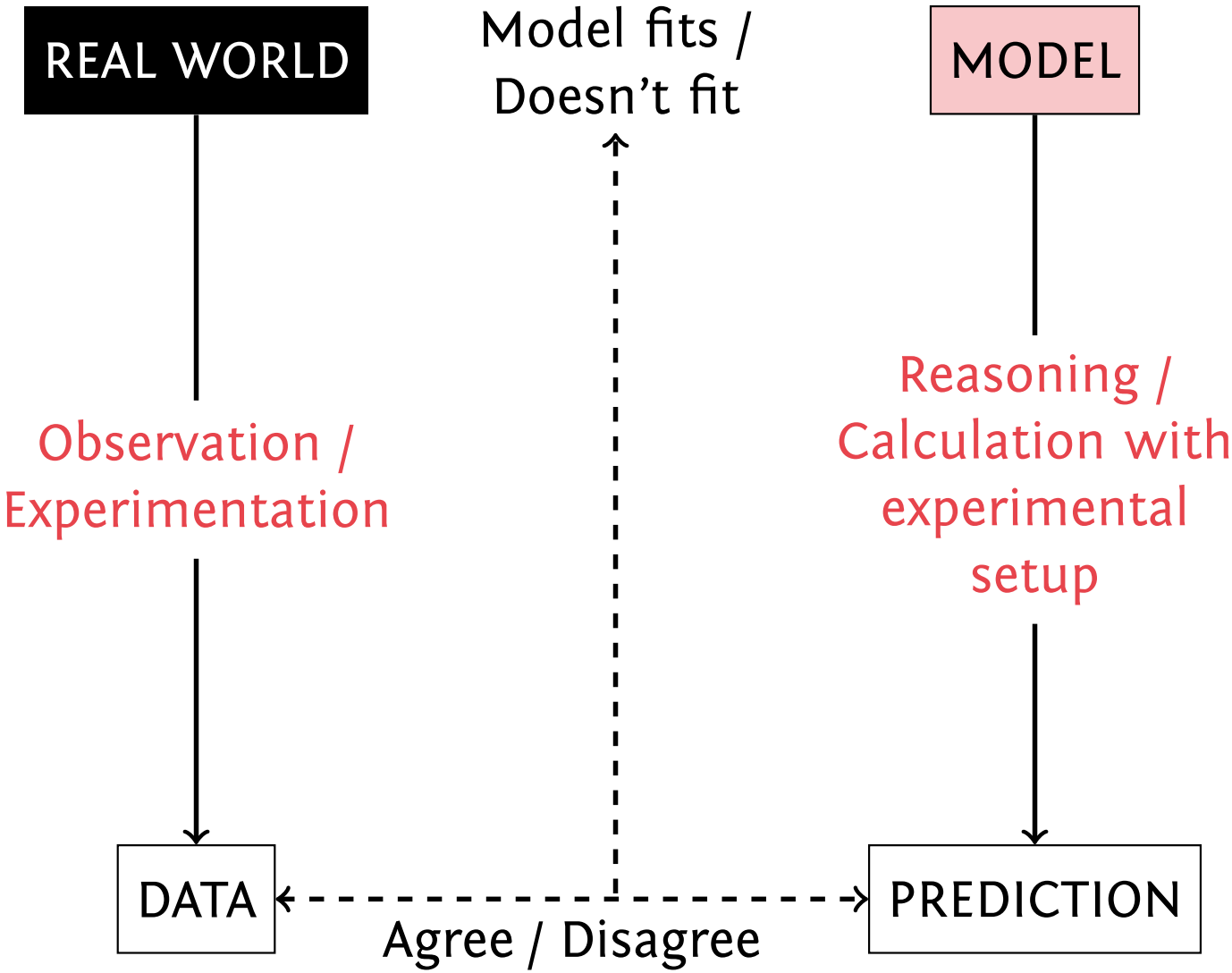
(b) Antiparallel orientation of strands

Working with models

We see a number of aspects of scientific model-building:

- A lot of model-building is guesswork.
- Lots of different types of models are involved: analog models (stairs, the α-helix, symmetry in biology, etc.), scale models (paper, cardboard, metal, plastic, etc.), theoretical models (the quantum mechanics of chemical structure, mathematical theory of x-ray diffraction), etc.
- Gathering data is a difficult process that involves expert knowledge (x-ray photographs, Chargaff's rules, etc.).
- Determining the prediction of the model also involves expert knowledge (the quantum mechanics of chemical structure, mathematical theory of x-ray diffraction, etc).
  - Often one needs to know the details of the experimental set-up to know what the model will predict (different x-ray photographs, etc.).

Components of a scientific episode



Overview

- We have looked at a number of different types of models that are used by scientists.
- We have looked at the determination of the molecular structure of DNA as an example of fitting the predictions of models with data gathered about the real world.
- We have shown how model-building, data collection, prediction and data-fitting can work together as a way of trying to overcome the difficulty presented by the fact that we do not have direct access to the real world.