

MOSAIC REASONING FOR DISCOVERIES

Boris Stilman

*University of Colorado Denver, Denver, CO, USA &
STILMAN Advanced Strategies, Denver, CO, USA
boris@stilman-strategies.com*

Abstract

We investigate structure of the Primary Language of the human brain as introduced by J. von Neumann in 1957. Two components have been investigated, the algorithm optimizing warfighting, Linguistic Geometry (LG), and the algorithm for inventing new algorithms, the Algorithm of Discovery. The latter is based on multiple thought experiments, which manifest themselves via mental visual streams (“mental movies”). There are Observation, Construction and Validation classes of streams. Several visual streams can run concurrently and exchange information between each other. The streams may initiate additional thought experiments, program them, and execute them in due course. The visual streams are focused employing the algorithm of “a child playing a construction set” that includes a visual model, a construction set, and the Ghost. *Mosaic reasoning* introduced in this paper is one of the major means to focusing visual streams in a desired direction. It uses analogy with an assembly of a picture of various colorful tiles, components of a construction set. In investigating role of mosaic reasoning in the Algorithm of Discovery, in this paper, I replay a series of four thought experiments related to the discovery of the structure of the molecule of DNA. Only the fourth experiment was successful. This series of experiments reveals how a sequence of failures eventually leads the Algorithm to a discovery. This series permits to expose the key components of the mosaic reasoning, tiles and aggregates, local and global matching rules, and unstructured environment. In particular, it reveals the aggregates and the rules that played critical role in the discovery of the structure of DNA. They include the generator and the plug-in aggregates, the transformation and complementarity matching rules, and the type of unstructured environment. For the first time, the Algorithm of Discovery has been applied to replaying discoveries not related to LG and even to mathematics.

1 Introduction

After continuous success in proving the applicability and in taking advantage of the power of Linguistic Geometry (LG) in applications to the modern warfare, over the last 15 years, we decided to investigate if LG would work for the ancient wars. LG was developed by generalizing approach utilized by the most advanced experts in playing chess [1]-[4], [22]. Theoretically, in several papers [34], [35] we demonstrated that Alexander the Great and Hannibal, in effect, “used” LG-like reasoning for their battles, perhaps even consciously,

or, most likely, subconsciously. This means that LG or proto-LG “existed” long before the time when chess was invented, which was about fifteen hundred years ago. Chess served just as a means for rediscovering LG. We concluded that LG appears to be a part of the human intelligence for probably a million years as the major component essential for survival of the fittest during the constant wars [23]. Moreover, LG could have been a component of the Primary Language (of the human brain) introduced by J. von Neumann in 1957 [37]. Even now nobody knows what this language is about. According to von Neumann’s hypothesis, this is the language

used by the human brain for thinking, which is different from all the 6,000 live and dead human languages. It must be much older than any of those, which means that it cannot be symbolic, i.e., cannot not utilize strings of symbols, phonetic or written. People thought and made discoveries, such as mastering fire and making bone tools, for at least a million years, while human languages exist for much shorter time. Moreover, the symbolic part of the human brain, the so-called neo-cortex, did not even exist million years ago. It has been developed later as part of the evolution of the Homo sapiens. Consequently, the Primary Language must be analog. However, we could still use modern algorithmic notation to simulate it if we realize what it is. I believe that the Primary Language is a collection of major algorithms crucial for survival and development of humanity, the underlying "invisible" foundation of all the modern languages and sciences.

It would be interesting to reveal other components of the Primary Language, besides LG. It is likely that they look like LG in some respects. My contention is that that the hypothesized Algorithm of Discovery must be one of such components. In a number of papers, I have been developing a hypothesis that there is a universal Algorithm of Discovery driving all the innovations and, certainly, the advances in all sciences [24]-[29]. I suggested that all the discoverers utilized this algorithm. The Algorithm of Discovery should be a major ancient item "recorded" in the Primary Language [37] due to its key role in the development of humanity. This line of research [24]-[29] involved investigating past discoveries and experiences of construction of various new algorithms, especially, those which I was personally involved in [12]-[35]. Another personal trait I utilized in this research is my vast experience in working with advanced experts, especially, chess experts, including renowned World Chess Champion Professor Botvinnik [1], [3], [4], and [12].

The main hypothesis is that the Algorithm of Discovery is based not on logic but on the so called "visual streams", i.e., mental imaginary movies which run in our brain. (LG is highly visual as well.) This is how it may work. Within the brain, the visual streams run consciously and subconsciously and may switch places from time to time (in relation to conscious/subconscious use). We

may run several visual streams concurrently, morph them, and even use logic for such morphing, although this use is auxiliary. Then we mentally tag some of the objects shown in the movie and create the so-called symbolic shell around the main visual stream. This shell eventually becomes a standard symbolic algorithm that can be communicated to others employing familiar language, logic, mathematics, etc. I named this approach "visual reasoning". While the visual component is, in general, pretty sophisticated, reasoning component is relatively simple. Fortunately, the full scale mental visibility is rarely used in discoveries, and, in my opinion, the limited visibility can be simulated with a reasonable effort. The reasoning component is certainly within the scope of the modern software development.

By tracing and replaying discoveries in actual development of the theory of LG, I revealed the dynamics of visual streams, especially, the means for focusing streams in desired direction. This way, I made several steps to discovering the Algorithm of Discovery [24]-[29]. In this paper, I will make the next step to advance this process by utilizing it theoretically to replay different, non-LG discoveries. Specifically, I will use this Algorithm to replay the discovery of the structure of DNA by Watson and Crick [6], [38], [39], [40]. This structure was revealed in the course research of a number of scientists over almost a hundred years. However, the major breakthrough was achieved by Watson and Crick as a result of several thought experiments conducted in 1951-1953. Those experiments were supported by constructing metal and wire, and even cardboard, physical models of the molecule of DNA. In this paper, I will replay four major experiments. One of the real life experiments to be replayed was conducted by Pauling and Corey [11], [38]. Interestingly, three experiments, out those four, failed to make a discovery. In spite of that, they produced new information and served as the major advances towards the fourth experiment that reached the final goal of revealing the structure of DNA. In this reenactment, I will utilize published research papers and introspections [2], [5], [6], [7] [11], [38], [39], [40] as well as other sources. From those introspections and research papers, I will try to glean more details about the operations of the Algorithm of Discovery based on the previous refinements that utilized discoveries in LG [25]-[29]. Note that traditionally

most of the research papers describe the Validation stream, i.e., validation of a discovery which has already happened. Very few of the papers describe the Observation and the Construction streams, i.e., the actual discovery struggle. The following experiments will be replayed.

- The triple helix model with the backbone and magnesium in the center.
- The triple helix model with the backbone chains in the center linked via hydrogen bonds.
- The double helix model with the backbone chains outside and double like-with-like bases in the center.
- The double helix model with mutually reversed nucleotide chains (the correct model of DNA).

In replaying those experiments I will pay attention to the so-called mosaic reasoning (Section 3) because they provide an excellent opportunity for investigating this major component of the Algorithm of Discovery. The first preliminary results on mosaic reasoning will be presented in Section 3.

This is my sixth journal paper on the subject. In the previous papers [25]-[29], I introduced various features of the Algorithm of Discovery that were revealed by doing case studies. In the next section, I will introduce the reader to a yet another, more comprehensive version of this Algorithm as I see it at the current stage of our research. The Algorithm includes a number of components that work in concert. However, not all the components are utilized for every discovery. For example, thought experiments that will be replayed in this paper will widely use mosaic reasoning while other components will be used on a limited scope.

2 The Algorithm of Discovery

The Algorithm of Discovery operates as a series of thought experiments, which interface with the rest of the brain and with external environment via imaginary animated movies (plays) which I named visual streams. These streams may or may not reflect the reality. This interface is constructive, i.e., visual streams could be morphed in the desired direction.

The input to the Algorithm is also a visual stream, which includes several visual instances of the object whose structure has to be understood or whose algorithm of construction has to be developed. Sometimes, the object is dynamic, i.e., its structure is changing in time. Then the input visual stream includes this visual dynamics. As a rule, neither the structure of the object nor the details of the dynamics are present in the stream. It simply replicates (mimics) the natural or imaginary phenomenon. The task of the Algorithm of Discovery is to understand its structure including dynamics and/or develop an algorithm for reconstructing this object including its changes in time. This understanding happens in several stages. Importantly, it always ends up with the process of actual reconstruction of the object employing the construction set developed by the Algorithm on the previous stages. If the Algorithm investigates a natural object this imaginary reconstruction may be totally unrelated to the construction (replication) utilized by the nature. Usually, this reconstruction process is artificially developed by the Algorithm of Discovery with the only purpose to reveal the structure of the object. However, if the algorithm of natural replication is the goal of discovery than the Algorithm of Discovery will employ a set of different visual streams to reveal the relevant components utilized by the nature.

All the visual streams are divided into classes, Observation, Construction and Validation. They usually follow each other but may be nested hierarchically, with several levels of depth.

The visual streams operate in a very simple fashion similar to a child construction set. The Construction stream utilizes a construction set and a mental visual prototype, a model to be referenced during construction. This is similar to a list of models pictured in a manual (or a visual guide) enclosed to every commercial construction set. It appears that all the thought experiments in LG related to construction investigated so far, [24]-[29], utilized those manuals. Imagine a child playing a construction set. He needs a manual to construct an object by looking constantly at its picture included in this manual. This model comes from the Observation stream as its output. It is not necessarily a real world model. It is not even a model from the problem statement. It is created by the Observation stream

out of various multiple instances of the real world objects by abstraction, specifically, by erasing the particulars. A final version of the object constructed by the Construction stream should be validated by the Validation stream.

The Algorithm of Discovery initiates the Observation stream, which must carefully examine the object. It has to morph the input visual stream and run it several times to observe various instances of the object from several directions. Often, for understanding the object, it has to observe the whole class of objects considered analogous. If the object is dynamic (a process) it has to be observed in action. For this purpose, the Observation stream runs the process under different conditions to observe it in different situations. The purpose of all those observations is erasing the particulars to reveal the general relations behind them. Once those relations appeared a construction set and a visual model have to be constructed by the Observation stream. Both are still visual, i.e., specific, – not abstract. However, they should visually represent an abstract concept, usually, a class of objects or processes, whose structure is being investigated. For construction, the Observation stream utilizes the Construction visual stream with auxiliary purpose (which differs from its prime purpose – see below). Note that the model construction is different from the subsequent reconstruction of the object intended to reveal its structure. This model may differ substantially from the real object or class of objects that are investigated. Its purpose is to serve as a manual to be used for references during reconstruction.

When the model and the construction set are ready, the Algorithm of Discovery initiates the Construction stream with its prime purpose. This purpose is to construct the object (or stage the process) by selecting appropriate construction parts of the set and putting them together. If an object has a sequential nature the construction also takes place sequentially, by repetition of similar steps. An example of such sequential construction is the discovery of the structure of DNA (this paper), where the structure of several helixes coiled about each other has been reconstructed as sequence transformations of a structural component called a generator.

At some point of construction, the parts are tagged symbolically and, in the end, visual reasoning with symbolic tagging turns into a conventional

symbolic algorithm to be verified by the subsequent Validation stream.

Models and construction sets may vary significantly for different problems. Construction of the model begins from creation of the construction set and the relations between its components. Both items should be visually convenient for construction. The Algorithm of Discovery may utilize a different model for the same object if the purpose of development is different. Such a different model is produced by a different visual stream.

In many cases the Algorithm of Discovery employs “a slave” to visually perform simple tasks for all types of visual streams. This slave may be employed by the Construction stream to “see” construction parts and put them together. More precisely, imagine a child playing a simplistic construction set. To avoid offending children, I had named this personality a Ghost. This Ghost has very limited skills, knowledge and, even, limited visibility. The Observation stream may utilize the Ghost to familiarize itself with the optional construction set, to investigate its properties. Next, the Construction stream may use the Ghost to perform the actual construction employing those properties. Eventually, the Validation stream may use the Ghost to verify visually, if properties of the constructed object match those revealed by the Observation stream. In all cases, the Ghost is guided by the Algorithm of Discovery or, more precisely, by the respective visual streams.

As was already discussed, the initial visual model is usually guided by a very specific prototype, where the Observation stream has actually erased the particulars. However, this specificity does not reduce generality in any way. This sounds like a paradox. My point is that every component of this model carries an abstract class of components behind it. This way visual reasoning about the model drives reasoning about abstract classes, which is turned eventually into the standard formal reasoning. This happens as follows. A visual model drives construction of the formal symbolic model so that the key items in a visual model have tags representing the respective formal model. At first, the formal model is incomplete. At some stage, a running visual stream is accompanied by a comprehensive formal symbolic shell. Running a shell means doing formal derivation, proof, etc. synchronized

with a respective visual stream. While the shell and the stream are synchronized, the visual stream drives execution of the shell, not the other way around. For example, a formal proof is driven by animated events within the respective visual stream. The visual streams, usually, run the creation of the visual model, the construction set and the final construction of the object several times. During those runs as a result of persistent tagging the symbolic shell appears. Multiple runs utilize the same visual components but during initial runs the synchronization of the stream and the shell is not tight. Further on, synchronization is tightened by morphing the visual model and/or adjusting symbolic derivation if they initially mismatch. Eventually, the stream and the shell switch their roles. In the end, it appears that the stream becomes the animated set of illustrations, a movie, driven by the running symbolic shell. For example, during the final runs (and only then), the visual streams, presented in [24]-[29], are driven by the constraints of the ABG, the abstract set theory and/or the productions of the controlled grammars. At this point the visual stream and the symbolic shell can be completely separated, and the visual stream can be dropped and even forgotten.

A stream may schedule other streams by creating almost a program with “procedure calls”. Essentially, it may schedule a sequence of thought experiments to be executed in the future. These experiments will in their turn initiate new visual streams. In this case, the purpose, the nature, and the general outcome of those experiments should be known to the stream created this sequence. However, this sequence is different from the list of procedure calls in conventional procedural (or imperative) programming. The algorithms of those “procedures”, i.e., the algorithms to be produced by the respective thought experiments are generally unknown. The experiments are not programmed – they are staged. The actual algorithm should be developed as a result of execution of such experiment. In a sense, this is similar to the notion of declarative programming when a function is invoked by a problem statement while its body does not include an algorithm for solving this problem.

The ability of a visual stream to schedule a sequence of thought experiments permits to create a nested top-down structure of visual streams with several levels of depth. Though, I do not think that

the actual depth of nested programmed experiments ever exceeds two or three.

It is likely that all the technological inventions and discoveries of the laws of nature include “optimal construction” or, at least, have optimization components [9]. Thus, various construction steps performed by the Algorithm of Discovery require optimization, which, certainly, makes construction more difficult. As the appearance of this algorithm is lost in millennia it could not certainly utilize any differential calculus even for the problems where it would be most convenient. For the same reason, it could not utilize any approximations based on the notion of a limit of function. In that sense, in order to reveal its optimization components, the most interesting problems to be investigated should lack continuity compelling the Algorithm of Discovery to employ explicitly those components. Based on several case studies [29], I suggested that this optimization is performed by the imaginary movement via approaching a location (or area) in the appropriate imaginary space. Having such space and means, the Algorithm employs an agent to catch sight of this location, pave the way, and approach it. Contrary to the function based approach, which is static by its nature, the Algorithm operates with dynamic processes, the visual streams. Some of those streams approach optimum (in a small number of steps); other streams show dynamically wrong directions that do not lead to the optimum and prevent the Algorithm from pursuing those directions. Both types of streams represent proximity reasoning. I suggested that proximity reasoning plays a special role for the Algorithm of Discovery as the main, if not the only, means for optimization. Important steps of various discoveries include construction of the spaces for proximity reasoning as well as initiating motion within those spaces [29]. Proximity reasoning is a type of visual reasoning [28]. This implies that the Algorithm should reason about the space where distances are “analogous” to the 3D Euclidian distances. Roughly, when we approach something, the distance must be visually reduced, and this should happen gradually. The space for proximity reasoning provides means to evaluate visually if the animated images representing various abstract objects approach each other or specific locations.

In the next section I will consider yet another type of visual reasoning that permits focusing the Construction stream of the Algorithm of Discovery, the main topic of this paper. Section 3 should be considered as a guide followed by the Algorithm of Discovery in the four experiments (Sections 6 through 9).

3 Mosaic Reasoning

The name of mosaic reasoning was introduced due to the analogy of the Construction stream operation with assembling a mosaic picture of small colorful tiles. Another, maybe, even more transparent analogy is known as a jigsaw puzzle when a picture is drawn on a sheet of paper and then this paper is cut into small pieces, mixed up, to be assembled later into the original picture. As Sir G. Thompson [36] pointed “. . . the progress of science is a little like making a jig-saw puzzle. One makes collections of pieces which certainly fit together, though at first it is not clear where each group should come in the picture as a whole, and if at first one makes a mistake in placing it, this can be corrected later without dismantling the whole group”. Both analogies, the pictorial mosaic and the jigsaw puzzle, represent well the key feature of the Algorithm of Discovery construction set. However, I prefer the former because the jigsaw puzzle looks more like an assignment in reassembling a construct, a picture, which has already been created and, certainly, well known. In that sense, a tile mosaic is created from scratch, including choosing or even creating necessary tiles. In addition, a jigsaw puzzle is re-assembled out of pieces based on random cuts. On the contrary, in pictorial mosaic, in many cases, every tile should have unique properties; it should be shaped and colored to match its neighbors precisely. A similar specificity is related to a group of adjacent tiles, the aggregate.

Returning to the Algorithm of Discovery, for many discoveries, the components of the construction set should be developed with absolute precision, in the way that every part should be placed to its unique position matching its neighbors. I will use the same name, the tiles, for those construction parts. If precision is violated the final mosaic will be ruined and the discovery will not happen. Though a group of tiles, an aggregate, may be configured properly, its correct placement in the mosaic

may be unclear and requires further investigation. Moreover, a tile itself may have complex structure which may require tailoring after placement in the mosaic. In some cases, a tile is a network of rigid nodes with soft, stretchable links, which is the case for the DNA mosaic considered in this paper.

Mosaic reasoning may lead to the observation, construction, and validation steps of the Algorithm of Discovery operating with tiles and aggregates of tiles. Overall, mosaic reasoning requires tedious analysis of the proper tiles and their matching rules. Investigation of the matching rules is the essential task of the Observation stream. Multiplicity of those rules and their specificity with respect to the classes of construction tiles make the actual construction very complex. Selecting a wrong tile, wrong tailoring, choosing a wrong place, or incompatible neighbors may ruin the whole mosaic. The matching rules are the necessary constraints that control the right placement of the tiles. Missing one of them, usually, leads to the wrong outcome because the Algorithm of Discovery is pointed in the wrong direction.

Some of the matching rules impact mosaic locally while other rules provide global constraints. The *global matching rules* include the requirement of the top-down analysis and construction, the global complementarity rule, certain statistical rules, the transformation rules, etc. For many if not all natural objects and processes, their structure is not reducible to a combination of the components. Large groups of tiles, i.e., large aggregates, may obey the rules which are hardly reducible to the rules guiding placement of singular tiles. Such matching rules must be understood globally first, implemented in the mosaic skeleton construction, and, only then, reduced to the placement of the specific tiles. An example considered in this paper is the choice of the helical structure of the DNA molecule including the number of strands (Sections 6 through 9). The rule of the global complementarity means that placement of one aggregate may determine precisely the adjacent aggregate. In case of DNA, one strand of the helix with the sequence of the base tiles attached to it determines the unique complementary second strand with the corresponding sequence of the base tiles (Sections 8 and 9). The global statistical rules related to the whole mosaic may reflect the relationship between major

structural components, the large aggregates. If understood and taken into account by the Observation stream, they may focus the Construction stream and lead to a quick discovery. In the case of DNA, the so-called Chargaff rules [5] reflect the structural relationship between the base tiles of the complementary strands of the double helix (Sections 8 and 9). Yet another class of global matching rules is called transformation rules. This is an algorithm for reconstructing an aggregate out of another aggregate and placing this aggregate in the proper location. Applied sequentially, such a rule permits to turn an aggregate, the so-called *generator*, into the set of adjacent aggregates. In some cases this set of aggregates may cover the whole mosaic. For example, the helical mosaic of the DNA molecule could be constructed if the generator and the translation-rotation transformation are defined (Sections 6-9). Interestingly, this type of construction may be utilized by the Algorithm of Discovery as a convenient procedure to reveal the structure of an object, e.g., the DNA molecule, while the nature may use a totally different algorithm for generating and replicating the same object.

The *local matching rules* include the local complementarity rule, the interchangeability rule, etc. The local complementarity means, roughly, that a protrusion on one tile corresponds to the cavity on the complementary adjacent tile. For the DNA molecule this is usually a hydrogen bond of a base tile (a protrusion) that corresponds to a negatively charged atom of the adjacent tile (a cavity), Sections 7 through 9. The local complementarity often expresses itself in the requirement of various kinds of symmetry within the pairs of matching construction tiles. The whole class of the local matching rules is based on interchangeability. In simple terms, if two aggregates that include several tiles are not identical but interchangeable, their internal structure may be unimportant. There are several levels of interchangeability. Two aggregates could be essentially the same, i.e., their skeletons coincide. Importantly, those skeletons must include nodes which serve as the attaching points of the aggregates to the rest of the mosaic. The notion of an internal skeleton depends on the problem domain and is specific for different types of mosaic. For example, two different aggregates for the DNA mosaic may have identical ring structures but the atoms and respective bonds that do not belong to those structures may

be different (Section 9). Another lower level of interchangeability of the aggregates does not require their skeletons to coincide. The only requirement is that the attaching points of those aggregates are identical (Section 8). In all cases interchangeability means that the stream can take one aggregate off the mosaic and replace it with another. This will certainly change the picture but the whole structure will stand. I named those aggregates *plug-ins*. It appears that plug-ins played crucial role in the discovery of the structure of DNA because such a plug-in was a key component of the helical generator (Sections 6 through 9).

Besides mosaic structural components that include tiles, aggregates, global and local matching rules, there is an unstructured component that I named a mosaic environment. Such environment may impact the structure of tiles, aggregates, application of matching rules, and the whole mosaic while being relatively unstructured itself. In case of DNA, this was the water content whose lack or abundance could seriously impact the structure of the whole mosaic.

As I already mentioned, the Algorithm of Discovery does not search for a solution in the search space. Instead, it constructs the solution out of the construction set employing various tools and guides. The right choices of the construction tiles and the matching rules by the Observation stream permitted focusing the Construction stream to produce the desired mosaic, i.e., to make a discovery.

In this paper, as I did in [24]-[29], I will follow visual streams, including Observation, Construction and Validation streams, in detail, in order to reveal the general rules (constraints) of the thought experiments staged by the Algorithm of Discovery. I will continue pursuing my hypothesis that those constraints are the same for all those experiments and reflect the core of the Algorithm of Discovery. For the first time, I will focus on the experiments related to the discoveries that reveal the major role of the mosaic reasoning. This is also my first paper on the Algorithm of Discovery with analysis of the non-LG related discoveries.

Figures included in this paper are mostly variants of the figures from [11], [38], [39], and [40]. However, here, they are not just illustrations but snapshots of the visual streams initiated by the Algorithm of Discovery.

4 Observation Stream: A Preliminary Analysis

I will attempt to recreate the knowledge of the structure of DNA available to Watson and Crick in the Fall of 1951 [38], at the time when they started their joint pursuit. I assumed that the same knowledge was available to Pauling and Corey working on the same problem approximately at the same time [11]. I would like to emphasize that my goal is not to praise or criticize those great scientists but to understand how they utilized the Algorithm of Discovery in order to reveal its major components.

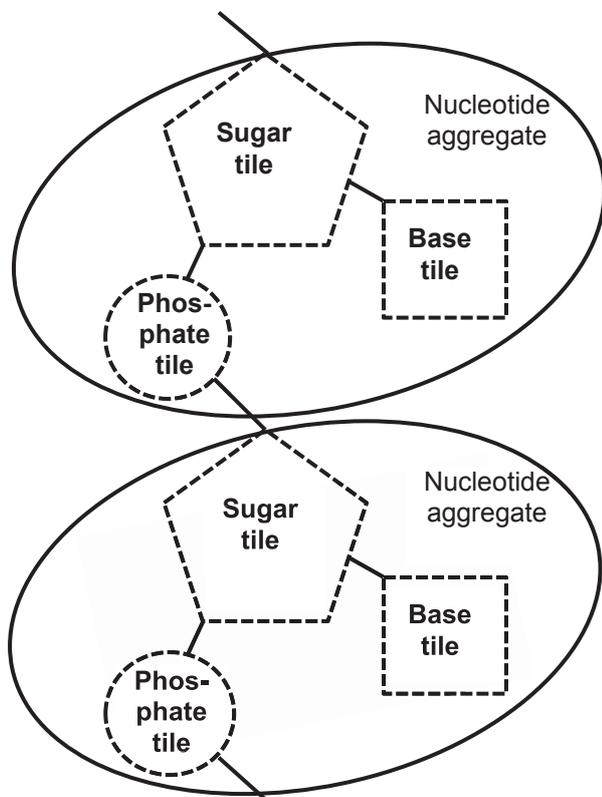


Figure 1. A short section of DNA: two nucleotide aggregates

The Algorithm of Discovery initiated the Observation visual stream. This stream utilized data about the structure of DNA that were obtained in the past experiments [6], [11], [38], [39] over more than a hundred years. The Observation stream had also drawn information for preliminary analysis from [2] and [7].

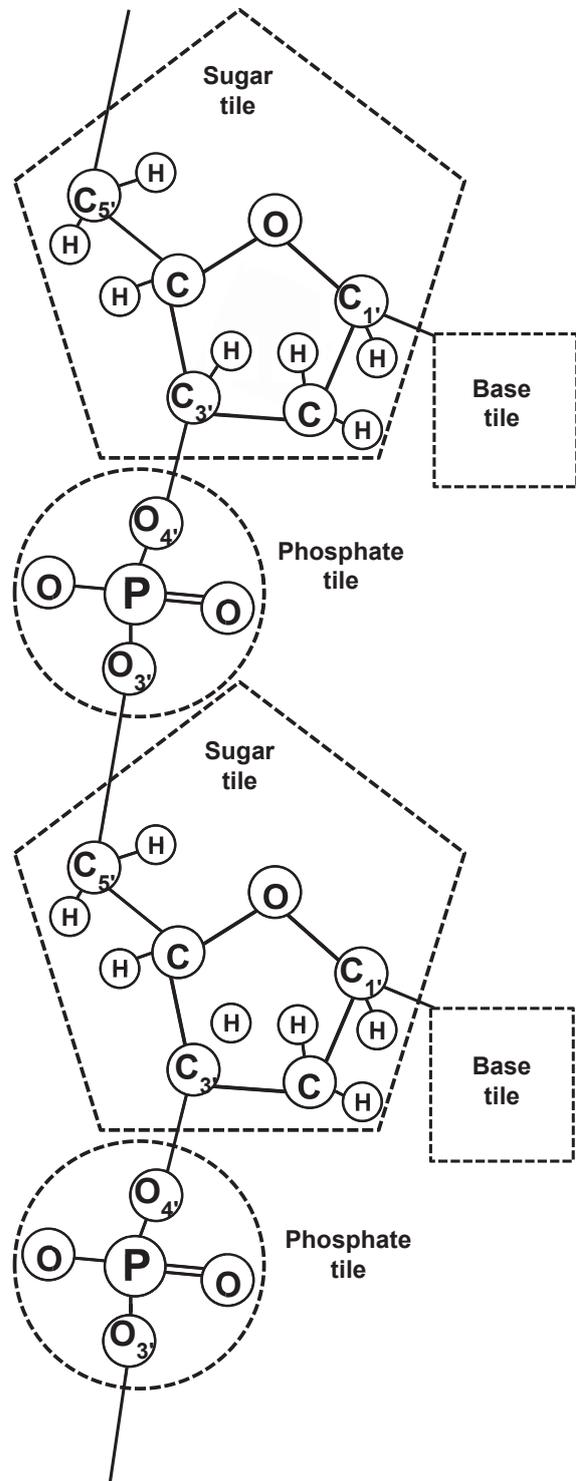


Figure 2. A detailed structure of the sugar-phosphate backbone

The Observation stream utilized the following preliminary knowledge to develop the construction set and the model to pass them to the Construction stream. It was known that DNA consists of very large molecules, even the largest of all known at the time. It is built of smaller blocks, the nucleotides. It was assumed that it carries genetic information though there was no a firm opinion about that. From the fact that DNA could crystallize the Observation stream concluded that it has regular structure, which can be reconstructed by the Construction stream employing the construction set.

The 3D α -helix model of proteins developed by Linus Pauling [10] served as a prototype model for this construction. Even, development of the construction set and the construction itself were modeled employing the Pauling's approach (Sections 6 and 7). Indeed, Watson noted [38] that "Pauling's accomplishment was a product of common sense, not the result of complicated mathematical reasoning. Equations occasionally crept into his argument, but in most cases words would sufficed. The key to Linus' success was reliance on the simple laws of structural chemistry. The α -helix had not been found by only staring at X-ray pictures; the essential trick, instead, was to ask which atoms like to sit next to each other. In place of pencil and paper, the main working tools were a set of molecular models superficially resembling the toys of preschool children." In my opinion, Pauling applied the Algorithm of Discovery in its purest form. Despite of the Pauling failure in discovering the structure of DNA (Section 7), his approach was a clear example of mosaic reasoning with "molecular tiles". Watson and Crick succeeded by applying consistently mosaic reasoning of the Algorithm of Discovery, using Pauling's approach as a pattern.

Developing a construction set required constructing mosaic tiles and matching rules. The tiles were the components of a nucleotide. In that sense, the nucleotide can be considered an aggregate. As shown in Figure 1, it consisted of the two smaller aggregates, the Sugar-Phosphate backbone and the Base tile. The backbone included two tiles, the Sugar tile and the Phosphate tile whose matching rules were well defined. These same rules linked one nucleotide aggregate to another (Figure 1, Figure 2).

They were based on the phosphodiester bonds of the carbon atom number 5, $C_{5'}$, and the carbon atom number 3, $C_{3'}$, of the Sugar tile [38], [6]. With respect to the Phosphate tile those matching rules involved $O_{4'}$ and $O_{3'}$ atoms of oxygen, [11] and [6]. The sugar tile was attached to the Base tile through the carbon atom $C_{1'}$ [11], [38], [6]. The question remained if those rules were universal. In other words, the stream should have figured out if the intertile and internucleotide bonds were identical. This hypothesis had been around for a long time, however, it was not proved chemically.

The Observation stream reasoned as follows. If those bonds were not identical the DNA molecules packed together could not form the crystalline aggregates studied by X-rays by many researchers including experiments presented in [2] and [7]. The stream noted that the DNA molecule contained a very large number of nucleotides, about 30×10^4 , linearly linked together in a regular way [40]. More precisely, it assumed that a nucleotide was a 3D structural "period" repeated many times. While the regularity of the sugar-phosphate backbone was not questioned, the base tiles varied widely from one nucleotide to another.

Four types of nucleotides were found in DNA. They differed from each other by their Base tiles which were divided into two groups, a purine (Adenine and Guanine) and a pyrimidine (Cytosine and Thymine), Figure 3. Those tiles were attached to the carbon atom $C_{1'}$ of the Sugar tile (Figure 2), through the ring nitrogen atom 3, N_3 , in the case of Pyrimidine tiles, and the ring nitrogen atom 9, N_9 , in the case of the Purine tiles [11], [6], Figure 3. It was also known that only one of those Base tiles was attached to any given Sugar tile. The Base tiles had planar shape with a thickness of 3.4 (angstroms) [7]. Fortunately, the internucleotide matching rules did not involve the Base tiles, i.e., only the Sugar and the Phosphate tiles, Figure 2. The Observation stream concluded that the variability of the Base tiles did not violate the regularity of those matching rules and the periodical nature of the entire molecule of DNA. However, some kind of irregularity of nucleotides should have been present. However, this irregularity should have been "erased" or hidden by "erasing the particulars". Indeed, if the Base tiles or the Base sequences (within the chain of nucleotides) would be identical, all the DNA molecules would have been identical and all the

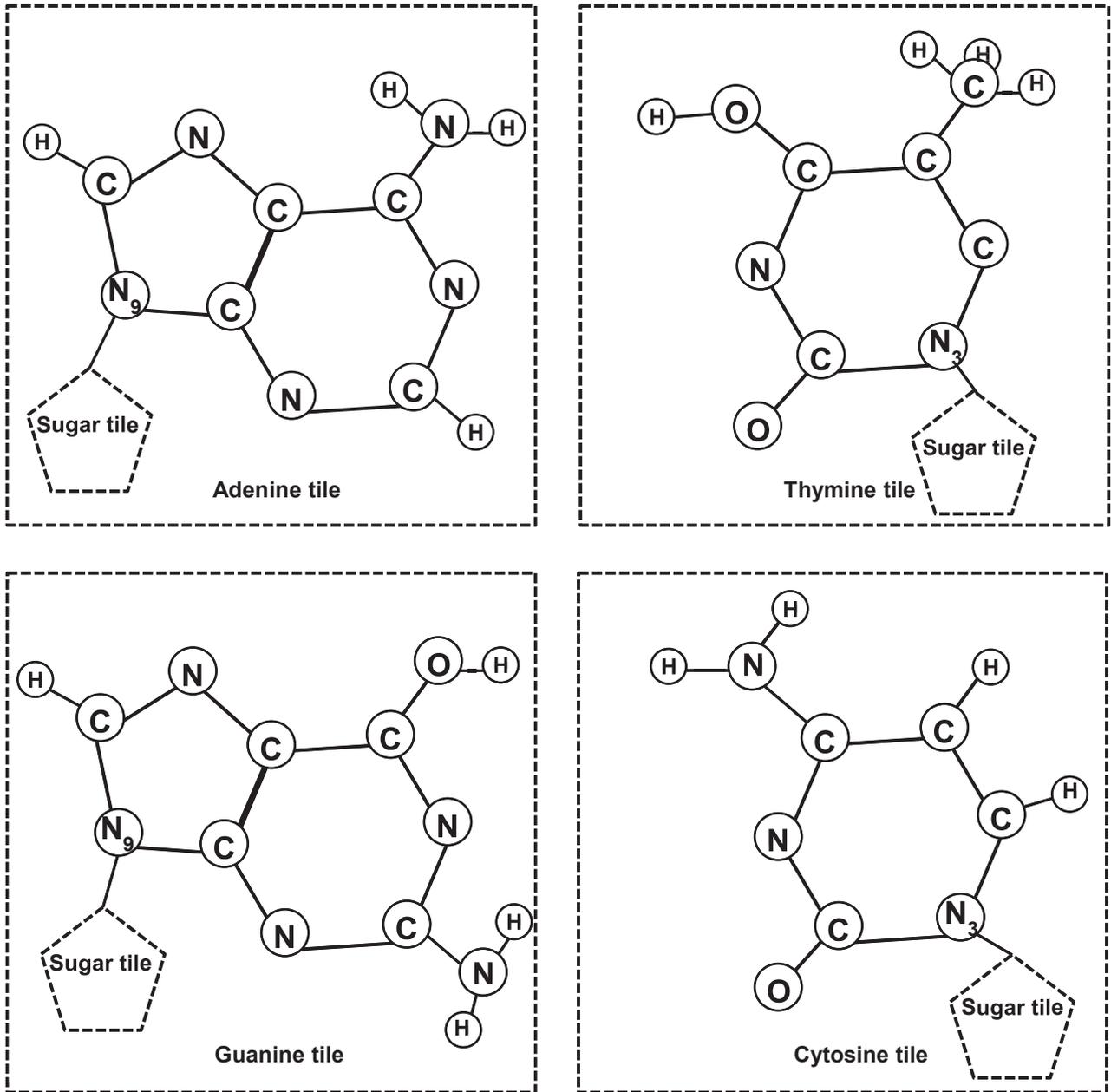


Figure 3. Base tiles, purines (left) and pyrimidines (right)

genes would become indistinguishable from each other. Thus, they would have carried the same genetic information and all the species would have been copies of each other.

Unfortunately, all the known details of the construction tiles and matching rules were seriously incomplete. The known details provided only partial information on the chemical structure of each of the six tiles, their aggregates, and intertile matching rules. However, the key question was about the global matching rules, specifically, about the 3D structure of the whole molecule of DNA, which was far from being apparent. Following the top-down approach the Observation stream focused on the global structure, because, if established, it could impact the local matching rules and even the structure of some of the tiles which were not built in stone. The major impact was expected to be made on the 3D configuration of intertile connections, interatomic distances and angles. The Observation stream considered the construction tiles (and their connections) as literally soft stretchable bodies.

With the hazy picture of the global 3D structure of the DNA molecule, important information on the local 3D configuration of a nucleotide was available. The Observation stream utilized the “standard” visual model of a pyrimidine nucleotide (Figure 4) published by Furberg [7] and reproduced by Watson in [38]. Furberg concluded that other nucleotides should be similar in this respect. Moreover, in [7] he described his “Construction stream”, with options of the intertile matching rules. (I am not going to speculate if he was aware of the algorithmic nature of the process he was describing.) Furberg determined that the plane of the Base tile was almost perpendicular to the plane, in which most of the Sugar atoms lie (which is the plane of the paper, Figure 4). The base plane is shown in gray behind the sugar tile (shown in blue). The Phosphate tile shown in brown is located in front of the Sugar tile. The $N_3-C_{1'}$ bond lies in the Base plane (Figure 4) and the bond $C_{2'}-C_{3'}$ is approximately parallel to the plane of the Base. The bond $N_3-C_{1'}$ is a single bond, and so rotation about this bond is possible. However, Furberg remarked, not all relative orientations of the Sugar and Base tiles are equally feasible, as in many positions the hydrogen atoms (small blue spheres) at $C_{2'}$ and $C_{3'}$ of the sugar come unfavorably near the atoms of the Base.

He determined that the most favorable position to be is the one shown in Figure 4, which is found in the crystal structure of Cytidine, one of pyrimidines. This is where he suggested options for the intertile matching rules. Furberg noted that there is a certain range of favorable positions; the Sugar ring may be rotated about 45° , clockwise and counterclockwise, on either side of the position shown in Figure 4 without bringing atoms of the Sugar and the Base too close together. This optional matching rule was certainly utilized in all the subsequent thought experiments and was passed to their Observation streams as an input (Sections 7-9). Remarkably, Furberg also noted that rotating the Base tile 180° about the bond $N_3-C_{1'}$ gives a structure which energetically would not appear differ much from the one shown in Figure 4. Historically, the value of this matching rule was neither understood nor utilized until the real life prototypes of Experiments Three and Four (Sections 8 and 9) when the double helix of two complementary strands was constructed. Its value could have been understood only in combination of a plug-in with several other matching rules such as the replication rule and the rule of complementary strands, which were missing at the time. In absence of the additional rules the Observation stream observed the base rotation rule and did not pass it further.

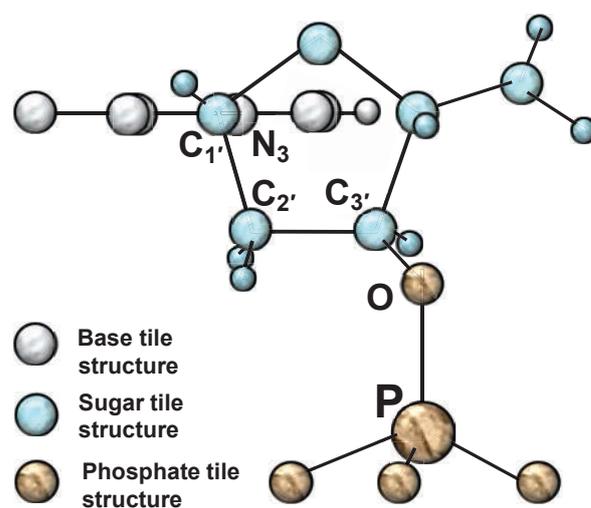


Figure 4. A standard 3D configuration of a nucleotide

There was no experimental evidence on the exact orientation of the Phosphate tile [7]. However, from a consideration of the van der Waals forces the position with the bond P-O roughly perpendicular

ular the Base plane was considered as preferred [7], with other less favorable positions located within the 45° clockwise and counterclockwise rotation of the Phosphate tile around the C_{3'}-O bond. All the matching rules considered above, except for the Base rotation rule, were passed to the subsequent experiments.

5 Four Experiments

The following Sections describe four thought experiments that utilized the preliminary analysis of the Observation stream (presented in Section 4). All of them were performed by the Algorithm of Discovery employing the standard visual components, Observation, Construction and Validation stream. The experiments are named One, Two, Three and Four according to timing of the respective historical events which those experiments replay. Experiment One is a replay the first attempt of discovering the structure of DNA by Watson and Crick [6], [38], [40]; Experiment Two is a replay of construction of the model of DNA suggested by Pauling and Corey [11]. Though Experiments One and Two utilized similar preliminary analysis of the Observation streams, the subsequent reasoning was different. However, both experiments produced essentially similar models generated by the Construction streams. Experiments Three and Four are the replays of the last two experiments by Watson and Crick, the double helix with like-with-like base tiles and the final successful experiment of the correct double helix construction. In all cases, the Observation stream of the subsequent experiment utilized the outcome of the Validation stream of the previous experiment. The only exception is related to the Experiments One (Watson and Crick) and Two (Pauling and Corey). The latter could not utilize the outcome of the Validation stream of the former because nothing was published, and two groups did not communicate on the issue.

6 Experiment One

According to Watson [38], the real life prototype of this thought experiment was actually completed in 1951 as a combination of the thought experiment with metal and wire physical model construction.

The Observation stream evaluated data about various types of DNA. According to various sources [6], [38], the pattern shown in the X-ray pictures was the same for all kinds of DNA ranging from viruses to mammals. This contradicted the fact that the ratios of various nucleotides varied from one source to another. Thus it should have been expected that the size and the shape of the molecule would vary correspondingly. In addition, it was known that the sequence of nucleotides within a DNA chain is irregular; hence, DNA should not form a repetitive structure and much less a crystalline structure. But it did! The likely conclusion was that the structure was based upon features common to all nucleotides. Such common features were located in the phosphate-sugar backbone (Figure 2), with an “average” Base tile attached to each Sugar tile. It could be considered as an idealized polynucleotide. This was a typical application of the erasing the particulars approach by the Observation stream to create a model.

For such a model it was plausible to assume that all the Sugar and Phosphate tiles were located in the respective “equivalent” positions and had identical environments for every nucleotide. The stream concluded that one nucleotide was related to another by one matching rule of transformation (Section 3). In case of one chain this would be a composition of a rotation about a fiber axes and a translation along this axes. Multiple repetitions of this transformation led to a helical mosaic. This also meant that an “average” nucleotide was a mosaic generator for such a helix (Section 3). The Observation stream was concerned about discovering a precise structure of the generator. The stream planned to construct “a couple of nucleotides”, the generator, and then turn them into the whole mosaic. Indeed, Watson wrote: “As long as we could be sure it was a helix the assignment of the positions for only a couple of nucleotides automatically generated the arrangement for all the other components” [38].

Out of various specific models for the global 3D structure of DNA [2], [7], [38], the Observation stream focused on the helix model. In particular, it considered the Pauling’s α -helix model for proteins which had just one strand. However, from the X-ray pictures available at the time it was known that the diameter of the DNA “cylinder” is about 20, which is wider than a single polynucleotide strand. The

number of strands could have been determined if the Observation stream would know the angles at which the helix appeared to zig-zag. The greater the number of strands within the given cylinder (of the fixed diameter) the steeper each of the strands, i.e., it would go nearly parallel to the axes of the fiber. If there is only two or even one strand the angle would be sharper. Good X-ray pictures could show this angle but such pictures available at the time were of dismal quality. The X-ray experimental data were compatible with 2, 3 or 4 strands [38].

In making a decision about the number of strands, the Observation stream had to deal with the issue of rather high density of DNA [2], [6] based on the helical structure with 10 nucleotides per period. This high density indicated the presence of the two or even three times more nucleotides. This led the Observation stream to conclude that the DNA molecule consists of several polynucleotide chains and they are coiled around common axes.

The Observation stream had to decide what part of the nucleotide to place in the center of the generator, i.e., close to the fiber axes. The stream concluded that this decision should be based on the packing consideration at the center of the molecule and that the core should contain the tiles common to all the nucleotides. This decision effectively ruled out the Base tiles because they had very little in common (Figure 3). The stream further concluded that placing the Sugar tiles in the center would not work either due to the irregular shape of those tiles (Figure 2). This left the Phosphate tiles in the center of the generator.

Putting those Phosphate tiles in the center made them responsible for the forces that should held the three chains together. Each Phosphate tile carried a negative charge pushing away other Phosphate tiles. This negative charge could have been neutralized by the salt bridges in which divalent cations like Mg^{++} could hold together three Phosphate tiles. The Observation stream introduced those cations as an additional tile for the DNA mosaic.

One more item of the input data was required for the Observation stream – the mosaic environment. As discussed in Section 3, such environment may impact the structure of tiles, aggregates and the whole mosaic while being relatively unstructured itself. In case of the DNA molecule this was water content. Watson provided the Observation stream

with data about water content of the DNA samples upon which the X-ray experiments were conducted. Inadvertently, he reduced this amount ten times! This erroneous data was received by the Observation stream.

The Observation stream scheduled the 3-STRAND(Mg) construction experiment. For this experiment the Algorithm of Discovery initiated the Construction stream and employed the Ghost. The Ghost constructed a rough model of the generator based on the three nucleotides as was suggested by the Observation stream. He utilized the mosaic construction set of seven types of tiles introduced at the preliminary analysis and at the Observation stages of this experiment. The Ghost applied repetitive transformation of translation-rotation to the newly constructed generator to create triple helix. A number of parameters could have been adjusted to finalize this model. For the visual stream replay, I assumed that the Ghost constructed the DNA model, which was very close to the one developed a year later by Pauling and Corey. This is the model that resulted from Experiment Two (Figure 5 and Section 7). The major difference was that in this model the three strands were linked by the cations of magnesium Mg^{++} instead of the hydrogen bonds. Those cations were located in the space near the center of the generator, between the three Phosphate tetrahedrons (Figure 5). This was the DNA mosaic submitted by the Construction stream to the Validation stream.

The first conclusion made by the Validation stream was the “discovery” of the realistic water content in the DNA molecule and its impact on the whole model. This input was provided by Franklin [38] who indicated that the hydrophilic Phosphate tiles should be located so as to interact with water molecules, the mosaic environment, on the periphery of the molecule while the hydrophobic Base tiles should be packed into the core of the generator. This fact alone certainly invalidated the whole DNA mosaic.

The second nail into the coffin of the DNA model constructed in Experiment One was also related to profound impact of the water environment. The Validation stream concluded that Mg^{++} cations would be surrounded by tight shells of water molecules and would not be able to hold three helical strands together.

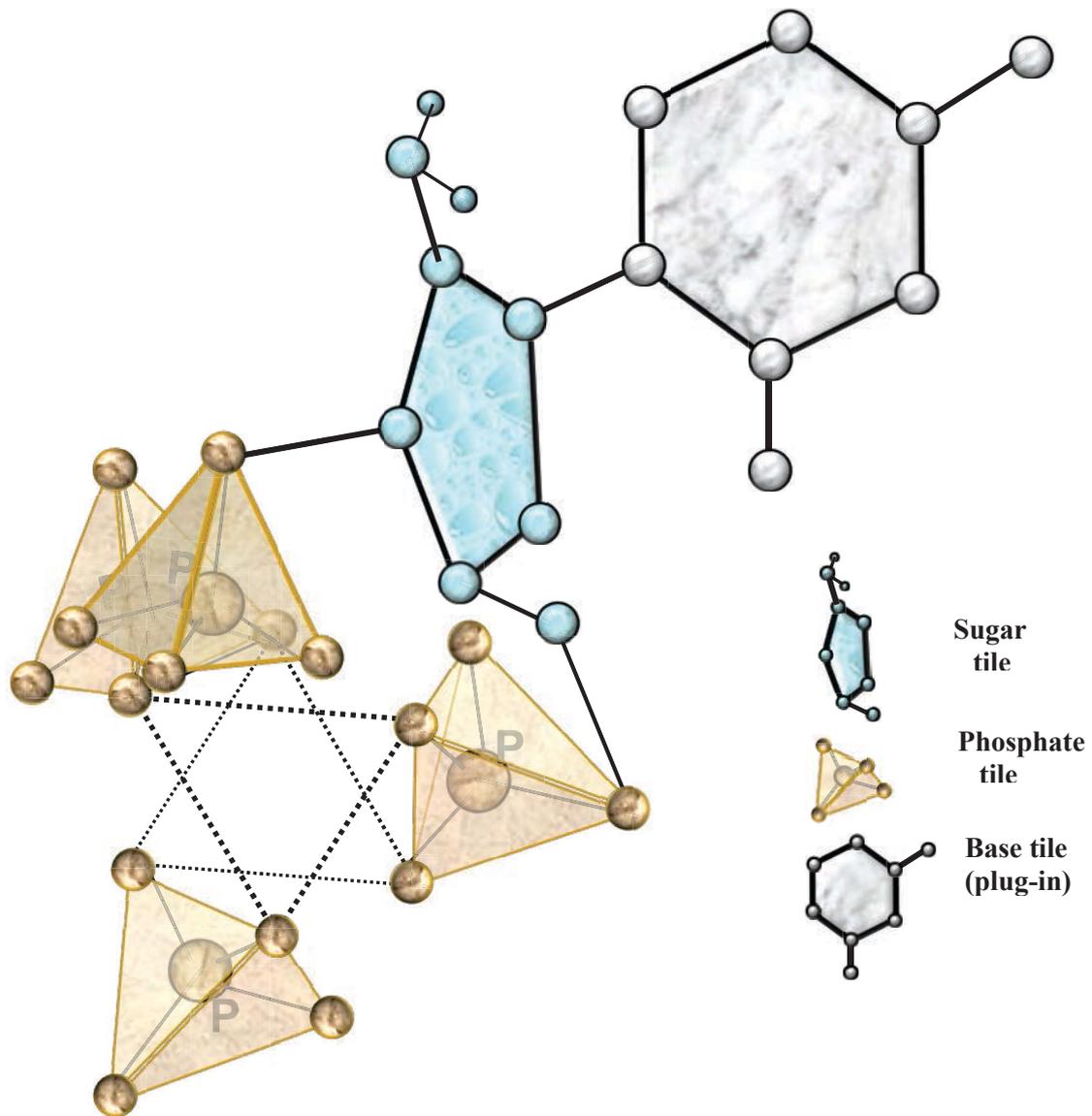


Figure 5. A cross-section of the DNA mosaic generated in Experiment Two

At the time, the Validation stream did not consider both observations as ground truth but it passed them as a highly probable to the Observation streams of Experiments Three and Four. It appears that the Observation stream of the Experiment Two (Section 7) was not aware of the water impact either.

7 Experiment Two

This experiment was not a descendant of Experiment One and did not utilize its output, though, its Observation stream used the same preliminary analysis (Sections 5, 6). To derive the global rule about the structure of the DNA mosaic, the Observation stream investigated global matching rules borrowed from the other models. Specifically, the stream investigated the transformation rules. For this purpose the Observation stream reviewed configuration of the variety of polypeptide chains in many proteins. This voluminous input was available due to substantial experience of the developers [10], [11]. In all cases this was the α -helix. The stream concluded that those mosaics consist of identical amino-acid aggregates (except for differences in the side chains). In proteins, there was only one matching rule that permitted to turn a generator aggregate into the neighboring aggregates, and, eventually, to construct the whole mosaic. Applying one type of transformations, the rotation-translation, led to creating the α -helix. The Observation stream further suggested that a similar matching rule could be applicable to the construction of the molecule of DNA out of multiple nucleotides. The matching rule based on the 3D translation-rotation would lead to a helix. In such case, the whole helix should be within a cylinder with approximately circular cross section.

The above conclusion was the first step to constructing a DNA mosaic generator. Reiterating a standard transformation to this aggregate the whole mosaic could be generated. This generator may include plug-ins, if necessary. As introduced in Section 3, two aggregates are called plug-ins if they are not necessarily identical but interchangeable. In such case their internal structure may be unimportant for the mosaic as a whole. In a simple case, those interchangeable aggregates could have just a single attaching point to the rest of the mosaic.

The stream received additional input supporting the suggested transformation rule [11]. In particular these were electronic micrographs that showed DNA to be fibrous in nature. The small fibrils, the molecules of DNA, looked circular of constant diameter in their cross-section. The ultracentrifuge experiments permitted to estimate this diameter to be about 20. The X-ray pictures showed a series of reflections compatible with a hexagonal lattice. Yet another input to the Observation stream included data about the length of 1.12 per residue along the fiber axes while the X-ray pictures demonstrated strong meridian reflection, spacing 3.4. The stream concluded that this reflection corresponds to a distance along the DNA axes equal to three times the distance per residue, i.e., $1.12 \times 3 \approx 3.4$. The stream further concluded that this reflection represents a unit consisting of three residues.

If the molecule were a single helix, the reflection at 3.4 would mean regularity in the sequence of the Base tiles or existence of a structural unit, an aggregate, composed of the three different nucleotides. The Observation stream concluded that there was low probability of existence of such structures which meant, essentially, that they do not exist. While the triple aggregates of nucleotides, indeed, are still unknown to exist, now, we know that such conclusion was too hasty. In reality, the DNA structural unit does include a pair of different nucleotides, which combined give a regular structure, the generator in Experiments Three and Four (Sections 8 and 9). This mistake was not an occasional oversight but a conclusive result based on the information available to the Observation stream.

In the meantime, the stream provided the alternative explanation of the X-ray data that the cylindrical molecule consists of *three* chains, which are coiled about one another. Each chain is a helix with translation equal 3.4, and those chains are related to each other except for differences in the Base tiles. This way the Observation stream established the global matching rule for the DNA mosaic: the universal generator should consist of the three nucleotides located in the adjacent sections of the three separate chains.

Several questions left to be answered. The first one was about the structure of the core of this generator, the part closest to the fiber axes. The stability of the model depends on how well atoms are

packed together, i.e., if all the inter-tile and intra-tile matching rules hold. The most “difficult area” of the generator was in the neighborhood of the helical axis than farther away where the distances between the “equivalent” respective atoms in the adjacent nucleotides are greater.

The stream considered three options for the composition of the core: the Base tiles, the Sugar tiles, or the Phosphate tiles. The Observation stream employed the Ghost for exploration of those options.

Due to their plane shape perpendicular to the DNA axes (Figure 4) and varied shapes of the four Base tiles (Figure 3) it was impossible to pack them along the axes of the helix in such a way that suitable bonds could be formed between the Sugar and Phosphate tile so that the established matching rules could hold. This choice was eliminated. It should be noted that it was certainly difficult to create a core of the generator out of triples of the Base tiles. However, certain pairs of the Base tiles do match to form a plug-in (employing the hydrogen bonds, Section 9).

The second option with the Sugar tiles in the core was also eliminated because the shape of this tile is such that close packing along the helix axes (they are almost parallel) is difficult: the matching rules would not hold.

The stream concluded that the core of the molecule is formed of the Phosphate tiles. Strangely, the stream utilized the erroneous input, i.e., the erroneous structure of the Phosphate tile HPO_4^- , which was different from the one provided by the preliminary analysis (Section 4) and prior research. The mistake was related to the presence of the hydrogen atom (Figure 2 and Figure 4). The stream visualized the Phosphate tile as a tetrahedron without explicitly showing the atom of hydrogen (Figure 5).

The Observation stream scheduled a 3-STRAND(H) experiment to actually construct the DNA mosaic and passed all the required information to the Construction stream. The Construction stream employed the Ghost, the construction set of six different tiles and the matching rules discovered by the Observation stream. The Ghost began construction of the generator from the closed-packed core near the DNA axes (Figure 5). Thus, he

was constructing all three helical strands simultaneously. The Ghost put three Phosphate tiles, HPO_4^- , in the mosaic; they are shown as three brownish tetrahedrons, Figure 5. The Phosphorus atom P is shown as a larger sphere inside of each tetrahedron. Six oxygen atoms, shown as brownish smaller spheres, two from each Phosphate tetrahedron, are located closer to the center that coincides with the axes of the molecule, which is perpendicular to the paper, Figure 5. A similar group of three tetrahedrons, the next layer of the three helix chains, could be superimposed on the first one, with translation (shift) by 3.4 along the DNA axes and azimuthal counter-clockwise rotation of 105° . The Ghost utilized the first layer of Phosphate tetrahedrons, the component of generator, to construct the second instance of generator by composition of the upward translation and counter-clockwise rotation. Only one tetrahedron of the second layer (dark brownish color) linked to the Sugar tile (blue pentagon ring) is shown in Figure 5. As a result of subsequent transformation the neighborhood of the DNA axes was filled with oxygen atoms which change their azimuthal orientation by 60° from layer to layer.

One of the outer oxygen atoms of each HPO_4^- tetrahedron attached to the Sugar tile is shown in Figure 5 for one of the chains. The hydrogen atom of each Phosphate tile attached to one of the two inner oxygen atoms provided hydrogen-bond connection with the inner oxygen atoms of the other Phosphate tiles in the same layer. Hydrogen bond and its role of a facilitator of the complementarity matching rule is discussed in Section 3. Electronegative atoms in the Phosphate tile are all the atoms of oxygen. Appearance of a hydrogen atom in one tile should be complemented by the electronegative atom in the other tile at the bonding point. In Figure 5 those hydrogen bonds are shown with dotted lines. The Observation stream provided information that those bonds could be formed within the same layer only. (As it was discovered later by the Validation stream, existence of those hydrogen bonds was a totally non-realistic assumption.) Two layers of the Phosphate tiles were nearly superimposed but not connected to each other in one chain. According to the intertile matching rules the Sugar tile was constructed to link the Phosphate tile in one layer to the Phosphate tile in the next layer. One Sugar tile with connections to the lower and upper Phosphate tetrahedrons is shown in Figure 5. The plane of the

Base tile (metallic color) was constructed perpendicularly to the DNA axes – it is parallel to the plane of the paper.

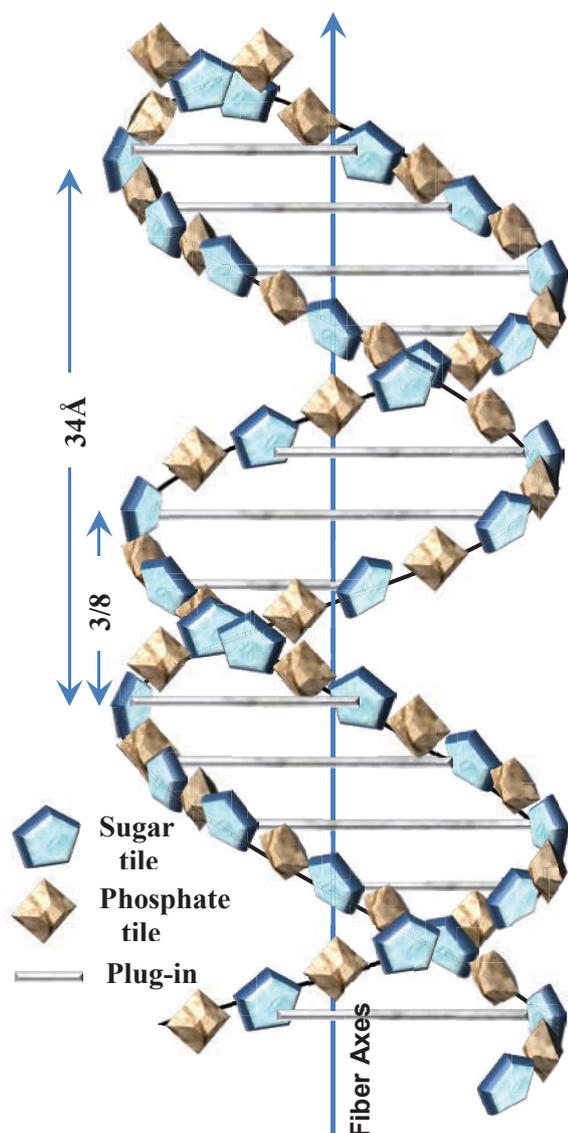


Figure 6. Global structure of the DNA mosaic for the 2-strand helix

The generator included three of each, the Phosphate tiles, the Sugar tiles and the Base tiles, i.e., three nucleotides. Each nucleotide aggregate of Phosphate, Sugar and Base tiles was included in one chain. Each instance of the generator could include different triple of Base tiles, three out of four. For this generator, a Base tile served as a plug-in (Section 3) which means that those tiles

could interchange each other in different instances of the generator (during construction). As a result of successive transformations of the generator, the Base tiles were superimposed in parallel layers with minor counter-clockwise rotation around the DNA axes and the distance between the successive layers was 3.4. This put all the Base tiles into a column arranged almost directly above one another. The only Base tile shown in Figure 5 is the Cytosine tile. The Ghost completed the 3-STRAND(H) construction of the DNA mosaic and the Construction stream initiated the Validation stream.

There were at least two different Validation streams for this experiment. One of those streams was initiated as usual, immediately after the Construction stream, at the place of “discovery”, which was Caltech in Los Angeles. It confirmed the construction based on the information available to the Algorithm of Discovery at the time. Interestingly, the Validation stream noted that the Base tiles on the periphery of the mosaic, i.e., on the periphery of the generator, have special features [11]. Those tiles were located in the way that their hydrogen-bond forming groups were directed radially, in the same plane as Base tiles, perpendicular to the fiber axes. This should have permitted the DNA molecule to interact with other molecules. This conclusion certainly underscores the great value of the hydrogen-bond components of the Base tiles, though in the wrong context. As we would know, those bonds are the keys to forming the Base tile plug-ins (Figure 7 and Figure 9).

The second Validation stream was initiated a little later at Cambridge University, UK, where a preliminary version of the article [11] was obtained. The major objection of the stream against the model constructed by the 3-STRAND(H) was violation of the standard matching rules of Chemistry. The whole DNA generator structure was based on the hydrogen bonds that were holding together the three helical chains. Without the hydrogen atoms those chains would break away. However, the Phosphate tiles had never contained bound hydrogen atoms. Due to DNA being an acid, under the real life conditions, even if present, those atoms should have been dissociated and the Phosphate tiles should have been negatively charged. This would also mean presence of positively charged ions like sodium or magnesium nearby to neutralize negative charges

[38]. The erroneous input to the Observation stream of the Experiment Two that permitted to construct the generator, based on the non-existent hydrogen bonds linking the Phosphate tiles, invalidated the whole construction.

8 Experiment Three

Historically, the real life prototype of this experiment was caused by the two factors, the results of Experiment Two and the appearance of the new X-ray picture of the DNA molecule, the so-called B-form of DNA [38], [40]. Just information about Experiment Two, in spite of its incorrect result, caused Watson and Crick to push the organizational levers for restarting their thought experiments suspended for 15 months.

The X-ray picture available to Experiment One, a crystalline A-form, occurred at 75% relative humidity and contained 30% of water, the DNA mosaic unstructured component. It appeared that at higher humidity the DNA fibers take up more water, increase in length by 30% and assume the B-form. It was suggested that the B-form was less ordered form of the molecule than the A-form. It meant that the B-form was para-crystalline, i.e., individual molecules were packed parallel to one another but not regularly arranged in space. Remarkably, transition from A-form to B-form was reversible, which meant that these were forms of the same structure.

The B-form picture clearly showed the black cross reflection in the center of the structure, which could be easily attributed to the helix shape of the molecule even without standard calculations. However, the specific data brought to the Observation stream by this picture was not really new. This data included the strong indication of the helical structure, the cylinder shape of the molecule with the diameter of 20 and the period of 34, with which the helical strands were repeating themselves when they coiled around the molecule axes. The most important data was the zig-zag angle (Section 6) that could implicitly permit computing the number of strands that formed helix. Assuming that the number of strands was two, the X-ray picture of the B-form suggested that those strands were not spaced equally along the molecule axes, but were displaced from each other by $3/8$ of the period [40], Figure 6. Unfortunately, those calculations were not conclusive.

In reality, the results of Experiment Two and the new X-ray picture simply caused the start of Experiment Three and reinforced input data available from Experiments One and Two including negative conclusions.

The Observation stream scheduled a 2-STRAND experiment for constructing a 2-strand model with Sugar-Phosphate backbone tiles outside. This should have been done by temporarily ignoring the Base tiles. This means that 2-STRAND was intended to construct the DNA mosaic via transformation of the new generator. This generator should have included a pair of Sugar tiles (blue diamonds), a pair of Phosphate tiles (brown diamonds), and a pair of Base tiles (“metallic plates”, perpendicular to the paper’s plane, linking blue diamonds), Figure 6.

A triple of tiles, a nucleotide aggregate, consisting of one tile from each pair belonged to each strand. Membership in the pair of Base tiles (the strand bridges) could vary from one pair to another, with all of them taken out of the set of the four Base tiles (Figure 3). It was not clear if those Base tiles included in this pair should be linked together.

According to Watson [38], due to his “lack of 3D vision”, the 2-STRAND experiment was actually completed in 1953 not just as a thought experiment but as a combination of the thought experiment with the physical model construction. Watson actually took a part their “particularly repulsive backbone-centered” physical model (the 3-STRAND(Mg) constructed for Experiment One, Section 6) and twisted it into a shape compatible with the X-ray evidence (Figure 6). This new model included the bridges of Base tiles (of unknown structure) holding the two strands together (Figure 6).

Assuming success of the 2-STRAND experiment the Observation stream generated a requirement to produce a new type of nucleotide(s) aggregate. This requirement was demanded by the 2-STRAND’s newly constructed model with identical 3D configurations of the corresponding backbone tiles in each strand on the outside of the helix (Figure 6). This was the first hard interchangeability requirement in a series of experiments considered in this paper. As discussed in Section 2,

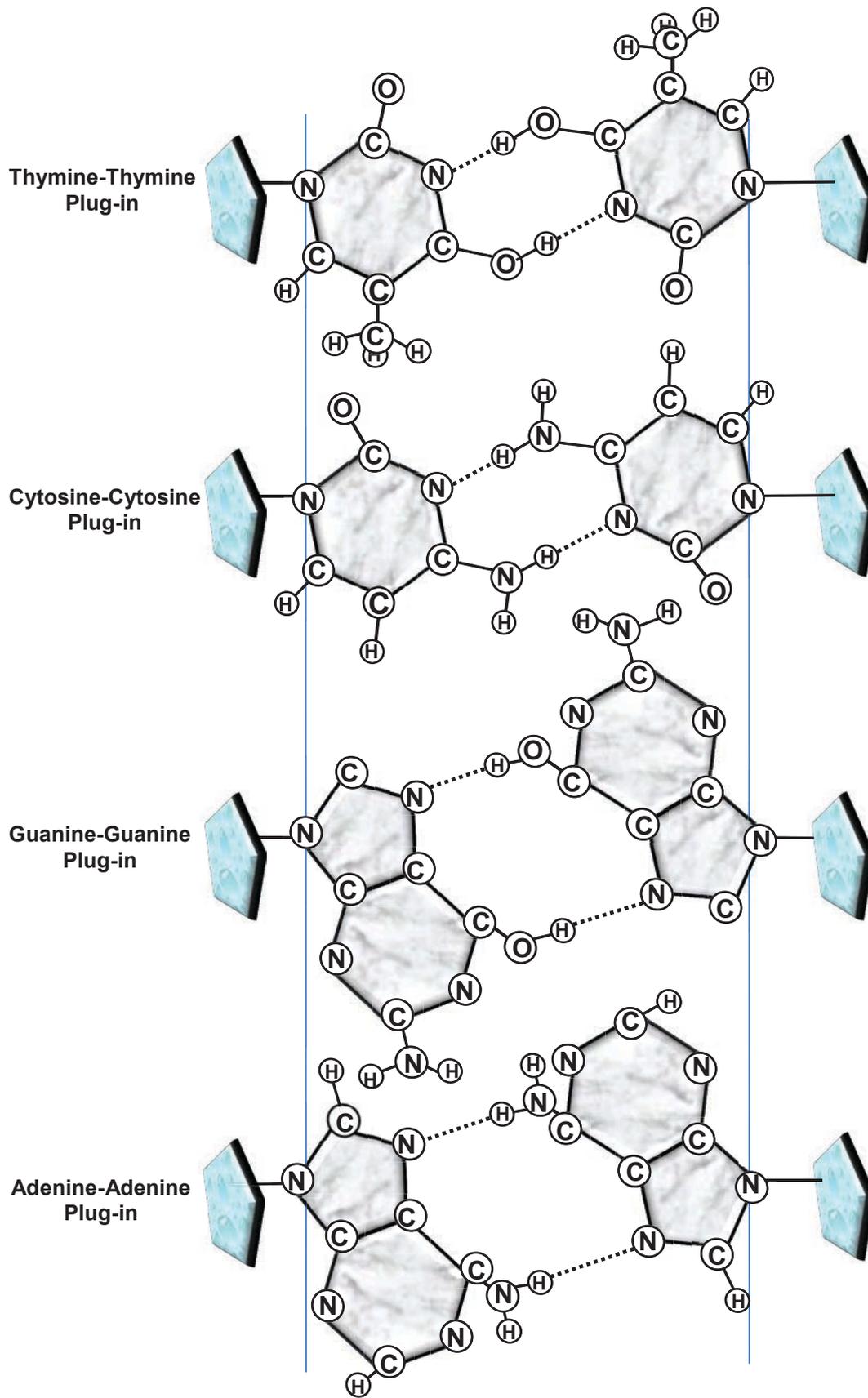


Figure 7. Plug-ins constructed in Experiment Three

if two aggregates are not identical but interchangeable, their internal structure may be unimportant for the mosaic as a whole. Such aggregates are called plug-ins. They can interchange each other. For the Experiments One and Two, the plug-ins were the singular Base tiles located on the periphery of the model with one attaching point each, and, thus, easily interchangeable. For the current experiment, it was necessary to arrange a regular pattern consisting of the Base tiles between the identical backbone tiles that belonged to different strands of the double helix. Regularity meant that all such patterns (plug-ins) should have followed the same matching rules, i.e., all the plug-ins should have been completely interchangeable. Specifically, those rules required that each plug-in should have the same planar shape and the same two attaching points determined by the corresponding Sugar tiles of the two strands (Figure 2, Figure 6). The distance between those attaching points should have been a constant. Importantly, those new plug-ins should have accommodated internally all the four different Base tiles (one, two, maybe more, per plug-in) in such a way that permitted an arbitrary sequence of base tiles along the polynucleotide chain (Figure 3 and Figure 6). Essentially, it was a combinatorial requirement for the Observation stream.

The stream scheduled a construction experiment PLUGIN-1 to try patterns consisting of a singular Base tile only. The Construction stream employed the Ghost who attached randomly the Base tiles to the specific nitrogen atoms of sugar in both strands coiled around the helix axes. PLUGIN-1 ended up in total failure because all the Base tiles had really different shape.

Watson vividly described his attempts to perform this experiment with his physical model [38]. My attempts to “randomly twisting two polynucleotide chains” (with base tiles between them) “around each other should result in a mess. In some places the bigger bases must touch each other, while in other regions, where the smaller bases would lie opposite each other, there must exist a gap or else backbone regions must buckle in.” The same troubles happened to the Ghost.

This failure motivated the Observation stream to consider an option of creating interchangeable patterns consisting of the two base tiles, possibly, attached to each other. The stream requested ad-

ditional input with information on the Base tiles matching rules. The response was that this should be a complementarity matching rule of the mosaic picture (Section 2) based on the hydrogen bonds. Electronegative atoms in the Base tiles include nitrogen and oxygen. Appearance of a hydrogen atom in one tile should be complemented by the electronegative atom in the other tile at the bonding point. This complementarity was visualized as a bump or protrusion in one tile and the cavity in the other. The Observation stream scheduled the PLUGIN-2 experiment.

The Algorithm of Discovery initiated the Construction stream which employed the Ghost perform PLUGIN-2 by combining pairs of Base tiles as adjacent plates (Figure 7). The Ghost used the complementarity matching rule as described above. It did not take long to realize that each Adenine tile would form two hydrogen bonds to another Adenine tile related to it by a 180° rotation. It means that the complementary plate was the same one turned around. Similar bonding could hold for other pairs of the Base tiles, Thymine, Cytosine, and Guanine. The potential existence of such rotated Base tiles was established by Furberg long before this experiment [7], Section 4. The knowledge of rotated tiles was available for the Observation streams of Experiments One and Two (Sections 6 and 7) but was not utilized because there was no purpose for doing so.

Returning to the PLUGIN-2 experiment the like-with-like pairing seemed satisfactory, Figure 7. Since the purines, the Adenine and Guanine tiles, and the pyrimidines, the Thymine and Guanine tiles, had different shape, the backbone strands had shown minor in-and-out buckles depending which like-with-like pairs were in the center. However, the Construction stream assumed that those minor helix inaccuracies should not violate the global structure and, probably, would not contradict the B-form of the X-ray pictures. Those minor inaccuracies are not shown in Figure 7.

The Algorithm of Discovery initiated the Validation stream whose first conclusion was that the existence of the two identical base tiles in each plug-in and two backbone sequences showed that the whole DNA molecule could be broken into two identical halves, each one associated with a helical strand. Both strands should have the same sequence

of Base tiles but their orders should be opposite, i.e., they should go in opposite directions. The Validation stream concluded that for this model the whole mosaic picture of DNA could be created by replicating a half of this mosaic simply by using the first half as a template. This conclusion suggested an algorithm for duplication of the whole mosaic by separating its halves (two strands with divided plug-ins), and using each of them separately as templates for completion to the whole mosaic. In a sense, two strands of DNA could have served as parents to the two identical children molecules of DNA.

Unfortunately, further analysis by the Validation stream was not so positive. The stream concluded that the 3D structure of the two out of four Base tiles was wrong. It was not a mistake of the Observation stream; the stream just used incorrect input information available at the time to the discoverers. Appearance of the new sources of data about the so-called tautomeric forms for the Guanine and Thymine tiles invalidated the first model generated by the PLUGIN-2 experiment (Figure 8). The correct Base tiles for the DNA molecule should have hydrogen atoms attached to different locations. They are called the “keto” forms.

The Validation stream scheduled the new PLUGIN-2 experiment with the modified Base tiles. For this purpose it initiated the Construction stream which employed the Ghost who attempted the like-with-like combinations of those tiles. Unfortunately, shifting the hydrogen atoms to the new locations made the relative sizes of the Thymine-Thymine and Guanine-Guanine plug-ins so different from the Adenine-Adenine and Cytosine-Cytosine plug-ins that they could not be considered as plug-ins any more (compare with Figure 7). They would have certainly required substantial bending of the backbone chains which would violate the cylindrical shape of the molecule and the helical symmetry demonstrated by the B-form in the X-ray pictures.

The Validation stream raised two more objections against the like-with-like model. The first one was related to the rotation angle of the helical structure. In order to keep the 34 period, this model would have required a small rotation angle, just 18° , which was not supported by the X-ray pictures. The second objection was related to the so-called Chargaff rules [5]. A series of past biochemical experi-

ments demonstrated that the ratios of the total numbers of certain types of the Base tiles were constant for all the DNA molecules extracted from different living organisms. Specifically, the ratios of Adenine to Thymine tiles as well as Guanine to Cytosine tiles were both close to one [38]. This was a statistical fact, whose significance was neither understood nor explained in the previous thought experiments on construction of the DNA model. Moreover, the Chargaff rules had never been included on the list of requirements for the Observation stream. Note that direct inclusion of those rules could have focused both the Observation and the Construction streams in the right direction (see below). This time it was demanded (by Crick) that the model should have provided a structural support to the Chargaff rules, which the like-with-like model did not provide [38]. Unfortunately, even this direct demand meant the inclusion of the Chargaff rules in the Validation stream only.

9 Experiment Four

The real life prototype of this experiment took place immediately after the prototype of Experiment Three. The Algorithm of Discovery initiated this experiment with the same input to the Observation stream as in the Experiment Three (Section 8). However, this input also included the results of the construction experiments 2-STRAND and PLUGIN-2. This meant that the Observation stream could restart the PLUGIN-2 experiment for constructing the generator but exclude the like-with-like plug-ins. On request of the Observation stream the Construction stream employed the Ghost to consider various combinations of the base tiles (excluding like-with-like). The real life prototype of Experiment Four was performed by the same team of Watson and Crick as the prototypes of Experiments One and Three.

The Ghost quickly found complementarity of the hydrogen bonds in the pairs of the Adenine-Thymine and the Guanine-Cytosine tiles (Figure 9). He realized that Adenine can only pair with Thymine, and Guanine only with Cytosine. If Adenine would have tried to pair with Cytosine it would not be able to form hydrogen bonds, since there would be two hydrogen atoms near one of the bonding positions, and none at the other, instead of one

in each [40]. This would mean violation of the local complementarity rule.

The ring structures of the new plug-ins were identical. This matching rule was never a requirement for constructing a plug-in but it was discovered visually by overlaying two plug-in flat shapes shown in Figure 9. These overlays and necessary rotations were performed by the Ghost. It could be stated as follows. The plug-ins were essentially “the same” (Section 3). For the like-with-like plug-ins from the Experiment Two such essential similarity existed only within the pyrimidine and purine like-with-like pairs, separately. That is, the ring structures of the pyrimidine and purine plug-ins were totally different (Figure 7). Simple overlay shows this difference. In case of the new plug-ins the overlay shows the identical ring structure. As we saw in Experiment Three for aggregates to become plug-ins, they had to have the identical attaching points (Sections 3 and 8). The attaching points of the new plug-ins were the two nodes of their ring structures, Figure 9. These ring structures were identical; hence, the attaching points were also identical. The plug-ins constructed in Experiment Three had identical attaching points but were essentially different contrary to the new plug-ins.

Taking this to the level of mosaic reasoning of the Algorithm of Discovery the new plug-ins are identical if we erase the particulars which in this case are several atoms outside of the ring structures. Future research on mosaic reasoning will show what these particulars are for different discoveries.

The Construction stream passed the complete DNA model with the new plug-ins to the Validation stream. The stream concluded that the Chargaff rules for the new model received simple structural support. Indeed, with totally irregular sequence of plug-ins the number of pyrimidine tiles of Adenine and Guanine would always be equal to the number of purine tiles, Thymine and Cytosine. Specifically, the number of Adenine tiles will always be equal to the number of Thymine tiles; the same is true for the Guanine and Cytosine tiles. The Chargaff rules simply reflected the major structural components of the mosaic, the pairing of the base tiles within the plug-ins and the double-helical structure of the DNA molecule. If available to the Observation stream, those rules could certainly drive the

discovery. It is likely that the mosaic, including the double helix and the special pairing of the plug-ins, could be discovered faster.

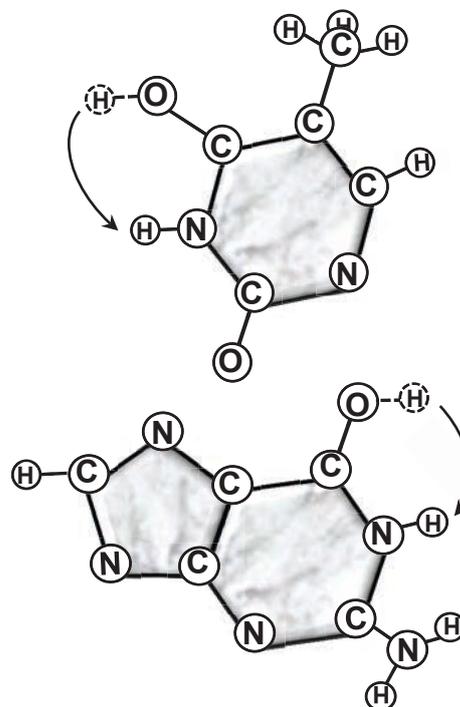


Figure 8. The tautomeric keto forms of the Thymine (top) and Guanine (bottom) tiles which are likely to occur in DNA. The hydrogen atoms are displaced.

The next conclusion of the Validation stream was related to the replication scheme of the mosaic. The structure of the DNA model demonstrated clear division of the mosaic into two halves, each associated with one of the two helical strands. This division goes through the generator and through every plug-in, and divides those plug-ins into two Base tiles which belong to pyrimidines and purines, respectively. Moreover, this division reveals the structure of the mosaic halves as two intertwined chains complementary to each other.

There are two levels of complementarity in the DNA mosaic. The lower level is the complementarity within the generator and every plug-in (Figure 9). The higher level is the complementarity of the entire sequences of Base tiles so that one chain automatically determines the Base sequence of another chain. Based on the known replication patterns, the Validation stream suggested that each of the chains could serve as a template for the synthesis of a chain with the complementary sequence.

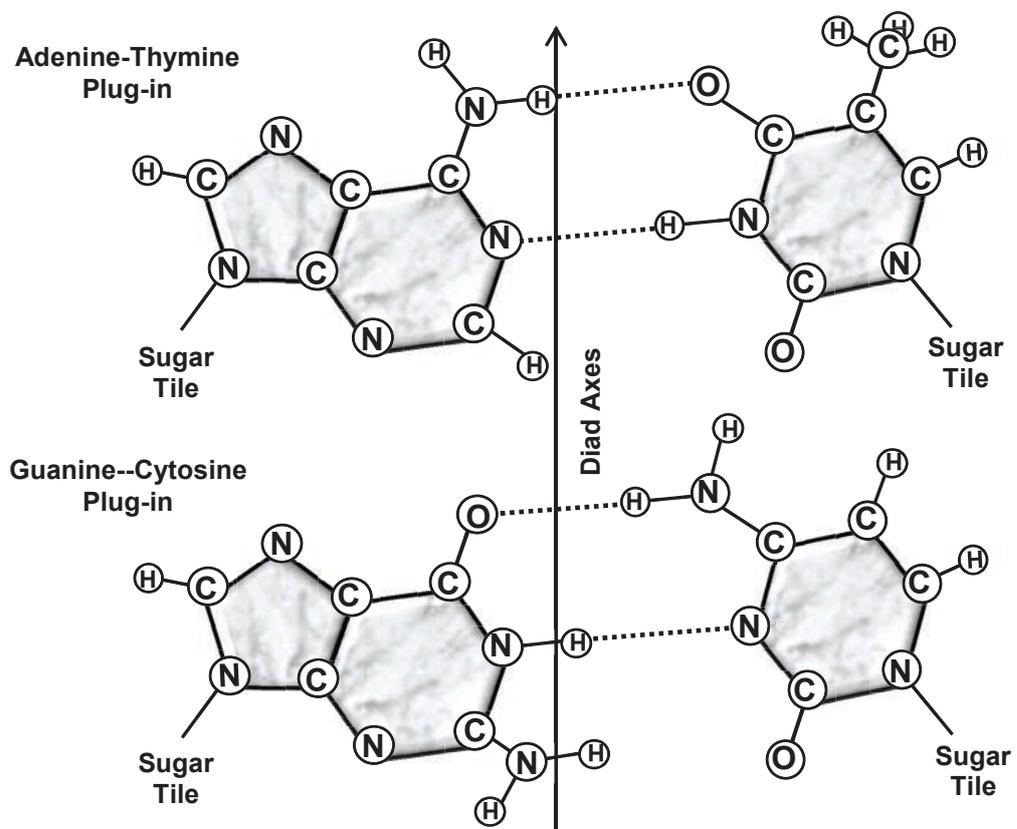


Figure 9. Plug-ins constructed in Experiment Four

Further investigation of the newly constructed DNA model and, specifically, the structure of new plug-ins revealed the symmetry of their attaching points, bonds that join Base tiles and Sugar tiles, Figure 9. They were symmetrical with respect to the straight line, the so-called diad axes that lies in the plane of the plug-in (which is perpendicular to the helical axes). The Validation stream concluded that the new plug-ins could be turned over but the attaching points as well as respective bonds would be preserved: they would match the original points. This means that the same plug-in may have two directions, from the first strand to the second and the other way around. Consequently, the backbone strands could run in the opposite directions.

10 Conclusion

With a bit of humor this paper could be titled “The tale of generators and plug-ins”. When contemplating this paper, before writing and rewriting it, I had no idea of existence of those generators and plug-ins. Moreover, the whole Section 3 “Mosaic Reasoning” included initially just several lines about the analogy of a product of the Construction stream to a pictorial mosaic. All the components of mosaic reasoning included in Section 3 and analyzed in the four experiments have been discovered while replaying those experiments and, concurrently, writing this paper. When contemplating the paper on mosaic reasoning, initially, I was planning to include the analysis of development of several algorithms in LG as I did in all the previous papers on the Algorithm of Discovery. The discovery of the structure of DNA as well as several other discoveries was certainly on the back of my mind. After rereading the Watson’s book [38], I realized that this discovery should be an ideal ground for investigating role of mosaic reasoning in the Algorithm of Discovery. Even then, until the midst of writing this paper, I could not imagine the power and the pervasive role of mosaic reasoning for this one and, likely, for many other discoveries.

In this paper, I made further advancements towards discovering the nature of the Algorithm of Discovery. All of them were demonstrated on four examples. For the first time since my first paper on the subject of discoveries I considered a non-

LG and, most importantly, non-mathematical discovery. Also, for the first time, I considered failed thought experiments. Every discoverer is familiar with experiments that failed to produce a discovery. The majority of them fail. Very few of those failures are published and the reason for publishing is that, at the moment, the failure has yet to reveal itself. It is pitiful that the details of those failures are quickly forgotten, especially, if success has followed. Even in the domain of my long-term research on LG, numerous failures are vaguely remembered. In that sense, it was extremely instructive to read papers that documented failures of the great scientists on their way to a discovery. Still, the vast majority of their papers describe the successful discovery of the structure of DNA where they also, sometimes, analyzed past failures. It appears that this post-success analysis is quite different from the documented descriptions of a failure, especially, when this failure still looks like success [11], or with specific intent to recall past struggle [38]. In their post-success papers scientists usually discuss past failures as optional solutions arguing why those options led nowhere. They simply forget that at the time of considering those options as viable their knowledge as well as information available to the Algorithm of Discovery was different from what they knew after the discovery. I realized that for the purpose of discovering the Algorithm of Discovery, the rare published descriptions of failures are a lot more beneficial than numerous publications on successful discoveries. Success causes persistent polishing the results (with extensive publishing) while failures cause persistent digging for errors (usually, without publishing).

The four experiments discussed in this paper for revealing the nature of mosaic reasoning demonstrate that the same algorithm was applied four times. Only on its fourth application it had led to a discovery. The only difference between those attempts was information available as input to the Observation stream of the Algorithm of Discovery, and this information was gained from one experiment to the next.

In Experiment One, the Algorithm discovered the existence of a generator with a simple plug-in in the DNA mosaic. The generator was constructed employing three nucleotide aggregates with generic Base tiles which served as simple plug-ins. This

construction utilized the construction set of six types of tiles all known before the experiment. The additional tile of the magnesium cation was introduced by the Observation stream. The Construction stream in this experiment managed to apply the translation-rotation transformation to the generator to construct a complete triple helix DNA mosaic. After so many “discoveries” the Validation stream invalidated the whole mosaic. In spite of the failure, Experiment One provided new data for future experiments with the DNA mosaic. They included the high likelihood of the multi-helix generator with the backbone tiles on periphery and the Base tiles inside the structure, the fatuousness of the seventh tile of magnesium, and potential importance of the new type of plug-in (using Base tiles).

In Experiment Two, the Algorithm made discoveries similar to those of Experiment One. The main difference was that this mosaic did not utilize the magnesium cations for linking the three strands of a helix but used the hydrogen bonds which were meaningless as well. Though this experiment was yet another failure, its total mosaic design was significantly more elaborate than the design in the Experiment One. In spite of the real life prototype of this experiment being performed a year later than the prototype of Experiment One, this experiment did not use the outcome of the One because nothing was published. What was probably even more frustrating for the discoverers was that they could not use the outcome of the second Validation stream of Experiment Two as well because it was too late – the real life prototype of Experiment Four performed by a different team had already succeeded.

Experiment Three inherited all the information produced by the previous experiments. In addition, it utilized new X-ray pictures of the DNA molecule which simply reinforced past conclusions. This was the first experiment where the Construction stream produced the double-helix generator with the Base tiles in the core of the structure. It also developed the first like-with-like plug-ins linking two helical strands. Importantly, for the first time, those Base tile plug-ins used hydrogen bonds to link the generator components. Though the like-with-like structure of the plug-ins was incorrect due to the mistaken input data about the structure of Base tiles, the global structure of the DNA mosaic was correct. The real life prototype of Experiment Three was the shortest experiment – it continued just several days.

Experiment Four was the direct descendant of Experiment Three. It created the classic model of DNA based on the double helix generator with complementary (purine-pyrimidine) plug-ins. The major discovery that was made by the Ghost during this experiment was that two Base tiles of totally different structures could be linked into a universal aggregate, the plug-in, to link together the two strands of the double helix. For the first time, the new plug-in provided structural support to the global statistical matching rules, the Chargaff rules. The complementary structure of the two chains of the double helix going in the opposite directions established foundation for the hypothesis of the template base replication of the DNA mosaic.

The role of mosaic reasoning in the Algorithm of Discovery described in this paper as well as its other components will be hypothetical until they will be verified by software implementations. The very first implementations have been initiated at the University of Colorado Denver.

References

- [1] M. Botvinnik, *Chess, Computers, and Long-Range Planning*, Springer-Verlag, New York, 1970.
- [2] W.T. Astbury, *X-Ray Studies of Nucleic Acids*, Symposia of the Society of Experimental Biology, 1, pp. 66–76, 1947.
- [3] M. Botvinnik, *Blok-skema algoritma igry v shahmaty* (in Russian: A Flow-Chart of the Algorithm for Playing Chess), Sovetskoe Radio, 1972.
- [4] M. Botvinnik, *Computers in Chess: Solving Inexact Search Problems*, Springer-Verlag, 1984.
- [5] E. Chargaff, E. 1951. Structure and function of nucleic acids as cell constituents, *Fed. Proc.* 10, pp. 654-659, 1951.
- [6] F. H. C. Crick & J.D. Watson, The complementary structure of deoxyribonucleic acid, *Proc. Roy. Soc., London, A*, 223, pp. 80-96, 1954.
- [7] S. Furberg, On the Structure of Nucleic Acids, *Acta Chemica Scandinavica*, 6, pp. 634-640, 1952
- [8] A. Kott and W. McEneaney (editors), *Adversarial Reasoning: Computational Approaches to Reading the Opponent's Mind*, Chapman & Hall/CRC, 2007.

- [9] A. Miller, *Insights of Genius: Imagery and Creativity in Science and Art*, Copernicus, an imprint of Springer-Verlag, 1996.
- [10] L. Pauling and R.B. Corey, Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets, *Proc. US Nat. Acad. Sci.*, 37(11), pp. 729–40, 1951.
- [11] L. Pauling and R.B. Corey, A proposed structure for the nucleic acids, *Nature*, 171, pp. 346–359, 1953; *Proc. US Nat. Acad. Sci.*, 39, pp. 84–97, 1953.
- [12] B. Stilman, Formation of the Set of Trajectory Bundles, Appendix 1 to the book: *On the Cybernetic Goal of Games*, by Botvinnik, M. M., Soviet Radio, Moscow (in Russian), pp. 70–77, 1975.
- [13] B. Stilman, *Ierarhia formalnikh grammatik dla reshenia prebornikh zadach* (Hierarchy of Formal Grammars for Solving Search Problems), Tech. Report, 105 p., VNIIE, Moscow (in Russian), 1976.
- [14] B. Stilman, A Formal Language for Hierarchical Systems Control, *Int. J. Languages of Design*, Vol. 1, No.4, pp. 333–356, 1993.
- [15] B. Stilman, A Linguistic Approach to Geometric Reasoning, *Int. J. of Computers & Math. with Appl.*, Vol. 26, No. 7, pp. 29–58, 1993,
- [16] B. Stilman, Network Languages for Complex Systems, *Int. J. of Computers & Math. with Appl.*, Vol. 26, No. 8, pp. 51–80, 1993.
- [17] B. Stilman, Linguistic Geometry for Control Systems Design, *Int. J. of Computers and Their Applications*, 1(2): 89–110, 1994.
- [18] B. Stilman, Translations of Network Languages, *Int. J. of Computers & Math. with Appl.*, Vol. 27, No. 2, pp. 65–98, 1994.
- [19] B. Stilman, Linguistic Geometry Tools Generate Optimal Solutions, Proc. of the 4th Int. Conference on Conceptual Structures - ICCS'96, pp. 75–99, Aug. 19–22, 1996, Sydney, Australia.
- [20] B. Stilman, Managing Search Complexity in Linguistic Geometry, *IEEE Trans. on Syst., Man, and Cybernetics*, 27(6): 978–998, 1997.
- [21] B. Stilman, Network Languages for Concurrent Multi-agent Systems, *Int. J. of Computers & Math. with Appl.*, 34(1): 103–136, 1997.
- [22] B. Stilman, *Linguistic Geometry: From Search to Construction*. Kluwer Academic Publishers (now Springer), 416 pp., 2000.
- [23] B. Stilman, Linguistic Geometry and Evolution of Intelligence, *ISAST Trans. on Computers and Intelligent Systems*, Vol. 3, No. 2, pp. 23–37, 2011.
- [24] B. Stilman, Thought Experiments in Linguistic Geometry, *Proc. of the 3d Int. Conf on Advanced Cognitive Technologies and Applications – COGNITIVE'2011*, Sep. 25–30, 2011, pp. 77–83, Rome, Italy.
- [25] B. Stilman, Discovering the Discovery of Linguistic Geometry, *Int. J. of Machine Learning and Cybernetics*, (DOI) 10.1007/s13042-012-0114-8, 20 p., 2012. Printed in 2013, Vol. 4, No. 6, pp. 575–594.
- [26] B. Stilman, Discovering the Discovery of the No-Search Approach, *Int. J. of Machine Learning and Cybernetics*, (DOI) 10.1007/s13042-012-0127-3, 27 p., 2012.
- [27] B. Stilman, Discovering the Discovery of the Hierarchy of Formal Languages, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI) 10.1007/s13042-012-0146-0, 25 p., 2012.
- [28] B. Stilman, Visual Reasoning for Discoveries, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI) 10.1007/s13042-013-0189-x, 23 p., 2013.
- [29] B. Stilman, Proximity Reasoning for Discoveries, *Int. J. of Machine Learning and Cybernetics*, Springer, (DOI): 10.1007/s13042-014-0249-x, 31 p., 2014.
- [30] B. Stilman, V. Yakhnis, O. Umanskiy, Winning Strategies for Robotic Wars: Defense Applications of Linguistic Geometry, *Artificial Life and Robotics*, Vol. 4, No. 3, 2000.
- [31] B. Stilman, V. Yakhnis, O. Umanskiy, Knowledge Acquisition and Strategy Generation with LG Wargaming Tools, *Int. J. of Comp. Intelligence and Applications*, Vol 2, No.4, Dec. 2002, pp. 385–409.
- [32] B. Stilman, V. Yakhnis, O. Umanskiy, Chapter 3.3. Strategies in Large Scale Problems, in [8], pp. 251–285, 2007.
- [33] B. Stilman, V. Yakhnis, O. Umanskiy, Linguistic Geometry: The Age of Maturity, *J. of Advanced Computational Intelligence and Intelligent Informatics*, Vol 14, No. 6, pp. 684–699, Sep. 2010.
- [34] B. Stilman, V. Yakhnis, O. Umanskiy, Revisiting History with Linguistic Geometry, *ISAST Trans. on Computers and Intelligent Systems*, Vol. 2, No. 2, pp. 22–38, Oct. 2010.

- [35] B. Stilman, V. Yakhnis, O. Umanskiy, The Primary Language of Ancient Battles, *Int. J. of Machine Learning and Cybernetics*, Vol. 2, No. 3, pp. 157-176, 2011.
- [36] G. Thomson, *The Inspiration of Science*, Oxford U. Press, London, 1961.
- [37] J. Von Neumann, *The Computer and the Brain*, Yale U. Press, 1958.
- [38] J.D. Watson, *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, Atheneum, New York, 1968. [Scribner Classics Edition, New York, 1996.]
- [39] J.D. Watson & F. H. C. Crick, A structure for deoxyribose nucleic acid, *Nature*, 171, pp. 737-738, 1953.
- [40] J.D. Watson & F. H. C. Crick, The structure of DNA, *Cold Spring Harbor Symposia on Quantitative Biology* 18, pp. 123-131, 1953.