Quaderns de filosofia vol. vi núm. 1 (2019): 29-46
eissn: 2341-3042 doi: 10.7203/qfia. 6.1.14823

Javier Anta
LOGOS, Universitat de Barcelona

Indispensability and Effectiveness of Diagrams in Molecular Biology

Received: 6-7-2018 / Accepted: 12-3-2019

Abstract: In this paper I aim to defend a twofold thesis. On one hand, I will support, against Perini [7], the indispensability of diagrams when structurally complex biomolecules are concerned, since it is not possible to satisfactorily use linguistic-sentential representations at that domain. On the other hand, even when diagrams are dispensable I will defend than they will generally be more effective than other representations in encoding biomolecular knowledge, relying on Kulvicki-Shimojima's diagrammatic effectiveness thesis [4][11]. Finally, I will ground many epistemic virtues of biomolecular diagrams (understandability, explanatory power, prediction and hypothesis evaluation) on their cognitive-computational indispensability and their semantic-epistemic effectiveness.

Keywords: Molecular Biology, Diagrammatic Representation, Representational Indispensability.

1. Introduction

The first thing you might notice when opening a biochemistry textbook is the astonishing amount of different visual resources that are employed, for instance schemas, flow charts, structural models, Haworth projections and so on. One could naively assume that the constant use of image-based representations, not just in textbooks but also in important biomolecular practices, only has an insignificant illustrative role as mere visual support of the main linguistically conveyed information. Otherwise and against this common prejudice, I am going to argue in this paper that not all but some visual formats, namely those which are

well-defined, syntactically-behaved and semantically-driven (from now on I will refer to them by the broad term of "diagrams") plays a more than foundational role in the scientific disciple of molecular biology.

Although today is gaining much attention in the literature, the philosophical analysis of representational systems in special sciences and their semantic-epistemic implications is a relatively underdeveloped topic, with the outstanding exception of general and molecular biology [12], [7] or [1]. In this line of inquiry, Sheredos [10] expressed his curiosity on "why do biologists use too many diagrams?". On my lights, the most plausible answer to this wide-scope question would be exactly the same than the one we could give to the more fine-grained "why might scientists prefer diagrammatic representations of information rather than, or in addition to, sentential ones?". A tentative response to both the former and the latter questions (originally formulated by Bechtel and Abrahamsen [1]) will be sketched within the following pages. In the first section, I will argue against Perini [7] that sentential or linguistic formulas are not even possible for representing biomolecules having a high structural complexity, e.g. proteins at their crystallographic or quaternary structure level, and therefore diagrammatic representations would be indispensable within that broad domain. The thesis that diagrams are semantic and epistemically more effective than linguistic representation, in a general context and even when these latter vehicles are available, will be addressed in section 3. I will use Shimojima's thesis of diagrammatic effectiveness and Kulvicki's immediacy thesis (namely, diagrams are representationally effective because their relevant informational content can be highly available) to account for the observational advantages of biomolecular diagrams over formulas and sentences. In the last section, many epistemic virtues of diagrammatic reasoning in molecular biology (e.g. comprehensive, explanatory and evaluative advantages) will be assessed as intrinsically depending on the previously defended indispensability and effectiveness of these representational systems. Now, let's start from the beginning.

2. Indispensability of Diagrammatic Representations in Molecular Biology

First of all, it would be fair to claim that molecular biology is one the scientific area with more variety of representational systems for codifying information about their empirical domain, in a syntax-based and semantically-driven manner. Let us take the illustrative case of the biomolecule D-Glucose, and eight most frequent forms of representing it, as it is depicted in Figure 1. They range from the name "D-glucose", its IUPAC nomenclature (fist on the

left) wherein every piece of information about the molecule remains implicitly referred[1], to one space-filling model of this biomolecule (first on the right), explicitly representing by graphical means a vast amount of physical and chemical properties, like van der Vars forces, which are encoded within the diameter of each ball. As one could learn from this eightfold representation, there no exist a sharp distinction between fully diagrammatic non-diagrammatic, or fully sequential representational systems; the key differences are properly found in the particular mechanisms used for codifying information (for instance in Fisher projection, carbon atoms are represented by chemical symbol "C", while in Haworth projection they are graphically encoded in the vertexes) about the 24 atoms of the D-glucose. It worth mentioning that there also exist fixed semantic codes shared by many representational systems, like the CPK coloring (white for hydrogen, black for carbon…), that allow to systematically interpret certain properties and relations.

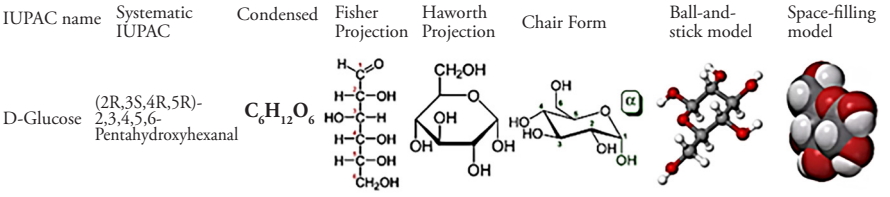


Figure 1. Representational manifold for the molecule D-glucose.

Laura Perini, one of the main philosophers devoted to assessing the representational and epistemic role of diagrams in biology, argues that the defining feature of diagrams (which properly demarcate them from linguistic representations) is the meaningfulness or significance of spatial properties and relations among the syntactically articulated graphical elements of the representing structure [5]. The syntactically-based and semantically-driven graphical behavior of diagrams is what differentiate diagrams from other kind visual representations, like pictorial ones[2] [11]. For instance, one cannot graphically alter the relative position of the

[1] It would be highly controversial to assume proper names like "D-glucose" as representations if we consider representations as a sort of "homeomorphic relations" between the symbol and the represented phenomena. Here we are going to assume that representation relations are referred to different procedures of codifying information.

[2] Pictorial means of representation are those usually characterized as exploiting graphical resources but lacking of compositionality-systematicity. For instance, it cannot be possible to systematically articulate a new electron microscope photography "C" just from other EM images "A" and "B" (even when they depict the same protein, having the same content) precisely because they have neither well-defined syntactic rules nor compositional behavior.

bottom hydroxyl group "OH" attached to the anomeric or first carbon in the chair form (third on the right in Figure 1, green colored) without altering its semantic content: the lower "OH" opposed in the ring to the CH_2OH group (indexed on carbon 5 and 6), which is known as a "trans" arrangement, graphically represents the specific alpha-anomeric structure of this molecules. In this sense, if this OH were just 1mm lower it would constitute a *meaningless* change, since this kind of structural diagrams are semantically sensitive not to the absolute location but to relative position (whether the OH is positioned "below" or "on the left", as depicted in Figure 3) of their graphical elements[3]. This particular graphically-codified anomeric structure of the D-glucose, indexed in linguistic representations by an "a-" or "alpha-", is only explicitly represented in the three diagrams on the right (Figure 1). Thus, diagrams are those representational systems wherein you can systematically change their meaning by syntactically manipulating graphical elements.

The fact is that diagrammatical alternatives in molecular biology are incredibly rich. Projective mechanisms of representation are particularly well-suited for codifying three-dimensional information in a schematic two-dimensional format: Fisher projective system make graphically explicit the organic or carbon-centered branching of biomolecules, grasping its chiral properties; the cyclic structure of carbohydrate become represented by means of Haworth representational system, which do not depict the actual but an idealized three-dimensional configuration of biomolecules (for that representational aim is effectively used its "chair form" projection). These diagrammatic mechanisms translate symbolic conventions of chemical notation, as used in condensed formulas, into sophisticated means of graphical descriptions of extensional structures.

In biochemistry, even the simplest object (for instance, the hydroxyl group "OH") possess many structural subtleties. This plurality of biomolecular structures posit an important question for the purpose of this paper: could every piece of structural information about a molecule be linguistically codified or not? Perini assumes the idea that "analysis of diagrams shows that their content can be expressed with linguistic representations" [7, p. 257] or in other words:

> **(1) Diagrammatic Dispensability**: The informational content of a certain diagram or set of diagrams can be equivalently represented on a linguistic-sentential format.

Based on the notion of "computational equivalence" of Larkin and Simon, Perini took for granted that, although cognitively essential for

[3] I should thank an anonymous reviewer for suggesting me this point.

understanding certain complex phenomena, the content of biomolecular diagrams could be fully translated into serial or linguistic representation. She defends that sentential representations are always available, either as long conjunctive formulas or as a list of linguistic descriptions of each atom as the one we could find on a computer render software, and in this sense, any diagrammatic codification of the same data would be semantically dispensable satisfying (1). As it has just mentioned, Perini remarks that biomolecular diagrams are those kinds of representations which must be understood as "cognitively indispensable" (or "essential", in her terms) for epistemic agents, not just to grasp complex information about those phenomena, but also to explain them:

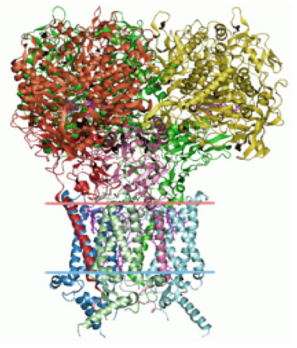
> The list of individual atomic coordinates would do little for a human in terms of understanding how these locations add up to the functional capacities of the complex. A serial representation of the positions of amino acids is readily available; it can be printed from the same electronically stored file of atomic coordinates which was used to make the diagram of the structure [5, p. 267]

I will support, in the forthcoming sections of this paper, Perini's idea of biomolecular diagrams being cognitively indispensable (namely, epistemic agents needs diagrams for obtaining biomolecular knowledge) and explanatorily powerful; but, up to this point, I argue that (1) do not holds for the cases of codifying information about macromolecules with a high structural complexity; which is a foundational claim, since molecular biology and biochemistry are empirical domains wherein complex structures can be found everywhere. Linguistic-sentential representations are not always available in this domain. Then, one might have robust reasons to support the following thesis:

> **(2) Diagrammatic Indispensability (at High Structural Complexity)**: The informational content of a determinate diagram or set of diagrams cannot be either computed nor equivalently represented on a sentential-linguistic format when this informational content possess a high level of structural complexity.

The motivation underlying (2) is not just that the information contained in a sentential representation of a complex macromolecule cannot be cognitively processed, but moreover, that this information cannot be (computationally) processed at all. Then, this would become a problem about the general computational impossibility (being human cognition a particular kind of computation) of processing such amount of information contained in

a single formula describing complex macromolecules at the atomic level. As Perini suggested in the above quote, the only actual serial formula of this kind is just a string of software code (a code sample is depicted on the right of image 2) whose unique semantic role is rendering the macromolecular diagram, which cannot be directly used for any epistemic activity.



Informational Inequivalence

Semantic Indispensability

(Untranslatability)

Epistemic Indispensability

(Unpredictability)

```
ATOM    886  CZ   PHE A 117      19.914   4.171   2.606  1.00 14.00
ATOM    887  N    THR A 118      23.856   7.434   6.252  1.00 11.56
ATOM    888  CA   THR A 118      23.660   7.280   7.655  1.00 10.93
ATOM    889  C    THR A 118      22.259   6.982   7.820  1.00 15.34
ATOM    890  O    THR A 118      21.447   7.435   7.051  1.00 14.33
ATOM    891  CB   THR A 118      23.982   8.621   8.138  1.00 22.78
ATOM    892  OG1  THR A 118      23.375   9.780   7.827  1.00 25.00
ATOM    893  CG2  THR A 118      25.586   8.842   8.226  1.00 25.00
```

Figure 2. Representing the high structural complexity of human deoxyhemoglobin, 4hhb. (Left) Crystal structural three-dimensional model from the Protein Data Bank, RCSB. (Right) Software code specifying a list of atomic values required exclusively for rendering the structural model.

In which sense should be understood structural complexity in molecular biology? Let us assume as a paradigmatic case of complex macromolecule the human deoxyhemoglobin, which is a protein containing a polypeptide chain of 154 amino acid (at the monomeric level) or 1579 atoms (at the chemical level). (i) Every amino acid encompasses atomic properties and relations among atoms; (ii) the primary structure or polypeptide chain of the deoxyhaemoglobin encompasses molecular properties, intramolecular and intermolecular relations and higher-order relations among atoms; (iii) its secondary structure (a-helix, for instance) entails higher-order intramolecular and intermolecular relations and exponential-order relations among atoms. We reach the critical structural complexity threshold at the level of (iv) deoxyhemoglobin's tertiary (as well as quaternary) structure, wherein the so-called computational complexity of a possible symbolic description became logarithmic: which means that exist an asymptotic approach to the use of infinite computational resources to solve the task.

For achieving this purpose in a sentential manner, it would not suffice with the chemical formula containing every atom within deoxyhemoglobin; notice that this kind of representation would have no structural information at all. Otherwise, it would be required one hypothetical formal-mathematical language able to express every single relational value, (e.g. complete set of atomic coordinates, relative positions, angular separation, etc.) required

for an exhaustive description of the protein or any other similarly complex macromolecule. Rejecting Perini's argument, I have just argued above that that no software code (see image 2) would ever satisfy those requirements. My claim is that such hypothetical symbolic apparatus would also suffer from the previously mentioned computational impossibility. At this point of structural complexity, only diagrammatic devices can computationally perform the representational work, and Perini is aware of it, in spite of her supporting (1):

> What makes structural models so important is the fact that amino acid chains do not simply stretch out in a line. They wind around in complicated formations. This means that side groups on amino acids that are very far from one another in terms of position on the amino acid chain might be located right next to one another in the protein [6, p. 267]

The main reason of why this computational impossibility is not the case for crystallographic models or structural diagrams is because of their idiosyncratic information-encoding mechanisms: they unload complex structural information extensionally codified across the representation, while in the case of sentential representations the intensional codification remain fully implicit in the symbols, involving an overload of computational resources to perform the same descriptive task. Up to this point, we should remark that even a sophisticated structural-diagrammatic model won't be able to represent all possible spatial relations within a determinate macromolecule, since it would rapidly exceed any conceivable set of computational resources. Traditional biomolecular diagrammatic systems (see Figure 1) are representationally constrained limited to a single level of organization, i.e. ball-and-still models to atomic structures, ribbon diagrams to polymeric units, and so on. Interestingly, new software-enhanced structural diagrams, satisfactorily encoding multilevel (atomic, molecular and polymeric) information, have been developed recently.

The most important databases of macromolecules, for instance the Protein Data Bank (PDB) or Proteopedia, employ complex 3D or stereographic (so they are not "static" or "printed") diagrammatic-structural representation[4] of proteins and other complexly structured biomolecules to organize relevant and novel information about the field. PDB diagrammatic-models, like the one depicted in Figure 2, encode information about several levels of biomolecular information in an interactive way: you can select whether to visualize (i)

[4] This 3D representation (2D projected onto the screen or the paper) could be stereographically rotated, so you could visualize the back of the protein in Figure 2.

its tertiary-secondary structure (protein strand, helix structures), (ii) its amino acid chain (primary structure) or even (iii) their underlying atomic architecture. These recent software-based diagrammatic representations also allow to overlap its protein-level interface (e.g. Figure 2) with its underlying atomic constitution, so you can visualize the exact location of oxygen atom within some helix structure.

Of course, there also exist databases of protein and DNA sequences, being NCBI one of most used, but one important fact within the field of bioinformatics is that structural-diagrammatic model already contains (and make explicit) its sequential or primary structure and not the other way around. This asymmetry is foundational for a second biomolecular practice that has recently gained much interdisciplinary attention, which is known as PSP[5] (Protein Structure Prediction) and consists on inferring highly complex (tertiary-quaternary) structure from its sequence or sentential representation. For carrying PSP, it is necessary continuous work and sharing computational resources in a worldwide scientific community (named CAMEO3D) exclusively devoted to the performance of these structural predictions. Even with both huge international cooperative effort and incredibly sophisticated new techniques, the current maximal accuracy for predicting *just the secondary structure* of a protein (notice that its biological functions rely on its tertiary-quaternary structures) from its primary structure is about 80%. Therefore, it would be required an asymptotically infinite quantity of computational resources for predicting higher protein structures from sequences of polypeptides.

…

Up to this point, we can firstly answer the Bechtel-Abrahamsen question "why might scientists prefer diagrammatic representations of information rather than, or in addition to, sentential ones?" in the following term: when information about highly complex biomolecular structures is concerned, molecular biologist can only have structural models or diagrammatic representations (even when they are restricted to a single level of organization) since *there are no* equally informative sentential-formulaic encodings of these macromolecules available. Furthermore, as I have defended: *there cannot be* linguistic representations at those levels of structural complexity. In the following section, I will try to give an answer

[5] I should thank Nuño de la Rosa for reporting me the fact that it cannot be predicted the tertiary or quaternary structure of any protein from its primary or sequential structure by any current software.

to the above question for general biomolecular contexts of lower structural complexity.

3. The Reasonably Effectiveness of Biomolecular Diagrams

It has been argued during the previous section that there are no (and moreover, there cannot be) properly linguistic or sentential representations of complexly structured biomolecules; therefore diagrammatic ones or structural models stand as indispensable means of codifying biomolecular information in the domain of high structural complexity. In this section I will argue that, additionally to their indispensability, diagrammatic systems are much more (i) representational and (ii) inferentially effective than formula-based means, not just in the not narrow domain of complex biomolecules, but in the general disciplinary context of molecular biology. Secondly, I will also try to offer one suitable naturalistic explanation of this reasonable effectiveness of biomolecular diagrams.

Shimojima formulated the question of why some diagrams are much more representationally effective than sentential in certain scientific practice context. His answer to that question is that one specific diagram would be representationally effective when its codification of information entails certain "observational advantages", like the possibility of reading-off its content without performing inferences. This is equivalent to what Kulvicki [4] calls the "immediacy of information" due to the extractability, syntactic salience and semantic salience. For instance, a chair form projection of a glucose molecule would be highly effective in encoding enantiomeric and anomeric structure of this molecule if one could tell a-D-glucose from b-D-glucose by simply reading-off the biomolecular (semantically salient, in Kulvicki's terms [4]) content of diagrams, as shows Figure 3. At least, this projective format would much more effective when concerning anomeric structures than other diagrammatic or sequential representational systems, like a fisher projection or a condensed formula respectively, simply because these two cannot representationally distinguish between a-D-glucose and b-D-glucose while a chair form projection can.

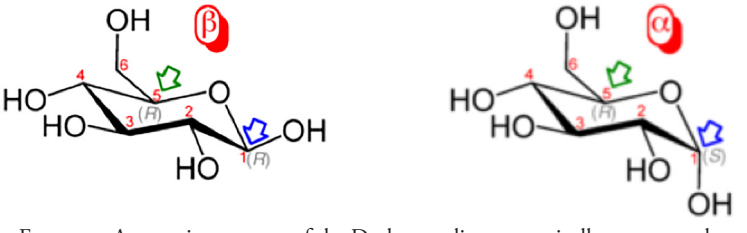


Figure 3. Anomeric structure of the D-glucose diagrammatically represented.

In this sense, Shimojima would claim that a chair form projection of a a-D-glucose can be used to read off its anomeric properties, which constitute a "observational advantage" against formulaic representations like the IUPAC name "a-D-glucose". In words of Kulvicki [4], the anomeric information would be immediately available from the diagrammatic representation precisely because of its syntactic-semantic salience. Even a 6-years old child could tell two anomerically different D glucoses apart from their chair form projections (by simply reading their informational content off from the below/right relative position of OH), but obviously she cannot do the same from their IUPAC names. Then, it can be said that biomolecular diagrams are representationally effective in many levels of biomolecular abstraction (i.e. atomic, monomeric or polymeric) because of the semantic-syntactic immediacy of their graphically-encoded informational content: this is "Kulvicki-Shimojima's effectiveness thesis", henceforth KSET.

Up to this point, it would be important to make the distinction, firstly stated by Larkin and Simons [5], between two representations being "informationally equivalent" (namely, the two implicitly or explicitly contain exactly the same information) and two representations being "computational" or "inferentially equivalent". This latter concept refers to two representations with the same inferential power, wherein the same kind of inferences can be performed on both. So, in the previous case of the "a-D-glucose" and its chair form projection, one could notice that the two representations are informationally equivalent but inferentially inequivalent, since a molecular biologist could effectively infer much more information from the projective diagram (e.g. where are located the binding atoms within the D-glucose molecule, how would it behave within an acid environment, etc.) than from its IUPAC name.

In the line of Suárez's [6] inferentialism, Bueno [2] developed the view that diagrams could be treated no just as syntactically-structured representations but moreover as epistemically significant "inferential devices". Then, the key difference between diagram-based inferences and formula-based inferences is that the former ones are (perceptually) carried immediately on the representational symbols, while the latter ones are (cognitively) performed mediately based on the meaning of the representation. As far as these two clearly differentiated inferential behaviors would determine different ways of obtaining biomolecular knowledge, we should also distinguish between the semantic and the epistemic dimension of KSET:

(3) Semantic Effectiveness of Diagrams: The semantic or representational effectiveness of a diagrammatic system is given by the (computational) exploitation of geometric-topological resources for explicitly codifying and extracting relevant information.

(4) **Epistemic Effectiveness of Diagram**: The epistemic effectiveness or inferential power of a diagrammatic system is given by the (algorithmic/cognitive) exploitation of geometric-topological resources for processing relevant information.

Departing from this two-fold thesis, we can offer an answer to the question of why diagrams are so (semantically) effective and have so many observationally advantages within molecular biology. The essential idea is that by graphically exploiting geometric-topological resources (e.g. projective spaces, invariant connections and shapes, geometrical configurations, etc.) in diagrammatic means it could be possible to encode relevant structural information about biomolecules without inflating syntax exponential or logarithmically, as in the case of sentential representations, or losing operational rigor, as in the case of pictorial means. Or in other terms: the semantic effectiveness of one particular representational system within a certain domain (e.g. structural models and diagrams in molecular biology) could be understood by means of representational mechanisms making relevant information more available. By relevant information in the biomolecular domain I specifically mean the minimal set of values that determine the value of any biochemical property of the macromolecule. It would be easy to demonstrate how these foundational biomolecular values are extensional values, wherein the "extensional" here refer to any space-constrained property: atomic position, distance between functional groups, angular separation and so on. In the end, the astonishing effectiveness of diagrammatic representations intrinsically depends on the intrinsic spatial character of biomolecular phenomena at any level of organization (i.e. pentagonal benzene, helicoidal secondary structure); fact that can be satisfactorily exploited by well-suited graphical mechanisms.

As a brief parenthesis, it is worth mentioning that this application of the KSET (3) and (4) to the scientific domain of molecular biology could explain why diagrammatic indispensability, under the conditions specified in the previous section, holds. The fact the certain diagrams and some formulas might be homeomorphic with higher biomolecular structures and biochemical sequences, respectively, does not tell us any relevant nor explain the diagrammatic effectiveness thesis. Otherwise, it could be satisfactorily explained by the fact that the structural complexity of macromolecules is well-suited for being diagrammatically and extensionally codified on symbol: for instance, complexity of a protein's quaternary structure become codified in the graphical entanglement of a structural model. While on the other hand, it could be sequentially-intentionally codified in the syntactic complexity (namely, how the symbols like "OH" or "C" are sententially articulated) of the correspondent language, which is precisely what underlies the computational overload mentioned before.

So far, we have been referring to the representational effectiveness of diagrams (3) in molecular biology; but this kind of visual representations also play a decisive and foundational epistemic role in this scientific field, as it will be fleshed out in the next section. The epistemic effectiveness of diagrams (or in other words, the semantic-pragmatic necessity of diagrams for obtaining and manipulating biomolecular knowledge), as formulated à la Shimojima in (4), can be based on the same fact than its representational effectiveness: the suitability of geometric-topological resources for encoding and processing relevant biomolecular information. That is precisely the reason why diagrams, being semantic and epistemically superior to sentential representation, are constantly used in actual practices of molecular biology: they are required for (i) facilitating compressibility and understandability of complex objects-mechanism in pedagogical contexts, (ii) many different kinds of biomolecular explanations (structural-functional, dynamical-mechanistic or even topological), (iii) testing novel hypothesis and discovering new phenomena, and (iv) predicting biomolecular events. Such rich manifold of epistemic virtues in biomolecular diagrammatic reasoning, grounded on its indispensability and effectiveness (as it has been defended in this paper), will be properly evaluated in the following section.

4. Epistemic Virtues of Diagrams in Biomolecular Reasoning

I have tried to show in the two previous sections that if molecular biologist rather use diagrammatic representations in their scientific practices is precisely because (i) in some occasions there are no sentential-linguistic options and (ii) in most occasions diagrams are more effective in codifying biomolecular information and knowledge than non-diagrammatic means. Nevertheless, I will argue in this section that diagrammatic representations are also superior to other representational apparatuses no just in describing macromolecules but also in explaining and understanding the biological function of certain chemical compounds or the mechanism underlying certain biochemical processes and mechanism, in assessing new hypotheses or even in predicting how biomolecules will behave under specific conditions.

Visual-Diagrammatic Understanding

One of the clearest epistemic advantages of diagrammatic representations over sentential formulations of biochemical information is the high degree of understandability or comprehensibility of the former over the later

format, which is not only important at the pedagogic level (it should be noticed that biochemical textbooks can be valued according to the quality of their illustrations [7]) but also in hypothesis-testing and explanation. With certainty, it might be claimed that molecular biology is one of the scientific disciples whose epistemic agents dependent cognitively more on visual means (diagrams, schemas, pictures, etc.) of conveying relevant information. At this point, we can interpret our thesis of diagrammatic indispensability (2) not in general computational terms, as it was done in section 2, but moreover from a cognitive perspective: it would be impossible, even for a professional molecular biologist, to understand or fully grasp the informational content of any sequential-sentential representation of a complexly structured macromolecule, precisely because the resources for cognitively processing it exceed by far human capacities. Then, as Perini defends, diagrammatic formats in the biomolecular domain are cognitively essential or indispensable.

Functional-Structural Explanations

The topic of diagrammatic explanations in empirical sciences is clearly underexplored. Perini [7] argued that this is mainly because since the deductive-nomological model of scientific explanations it has was assumed that only linguistic representations can be explanatorily relevant. Against this prejudice, one of the most relevant epistemic capacities of diagrams in general biology (but also in biochemistry and molecular biology) is, also according to Perini [7], to develop functional explanations by exploiting semantic-syntactically visual resources to graphically remark the functional relations existent among the parts of a certain molecule. I strongly agree with her in the thesis that diagrammatic systems are better suited than sentential representations in carrying those particular kinds of explanations, which might be regarded as a consequence of epistemic effectiveness (3) previously defended:

> The unique capacity of visual representations to represent higher-order relations in virtue of visible relations holding among parts of the figure suits them to functional explanation in a way that representations with a linear format cannot match [7, p. 267]

One structural model or one schematic diagram, as explanans, could be functional-structurally explanatory by making salient certain features within the visual representation that encoded the relevant information about

the explanandum. Let us consider the illustrative case of the biomolecular mechanism underlying bronchodilation as *explanandum*: it could be possible to explain in a signal transduction context how binding interactions and the functional import of the receptor are mediated by electrostatic effect; and we do it by representing the macromolecule (cell-membrane spanning) β2 adrenoreceptor in a complex space-filling model or diagram. The proper explanatory power of the diagram as *explanans* depends externally on (i) a minimal background knowledge of molecular biology and enzymology (facilitating understanding of the diagram), and intrinsically on (ii) graphically reinforcing the relation between those functionally-relevant structures within the macromolecule, in this case by a chromatic codification of biomolecular electrostatic properties.

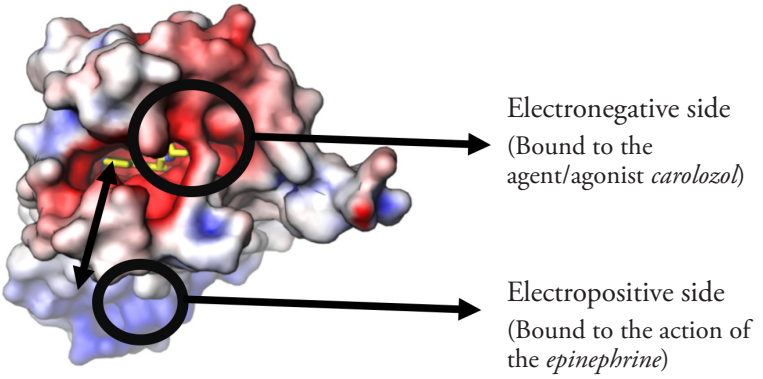


Figure 4. Explanation of the function of β2 adrenoreceptor in bronchodilation.

Mechanistic-Dynamic Explanations

The explanatory power or epistemic effectiveness of diagrammatic representations is far from being constrained to the functional-structural kind of explanations. Another closely-related kind of diagrammatic explanation very frequent within the biomolecular context is what is known within the literature as "mechanistic explanation" [1], and when time is a key factor they are also called "dynamic explanation". Sheredos [10] argues that, although it is perfectly possible to offer mechanistic-dynamic based on linguistic descriptions, diagrammatic representations of biomolecular mechanisms will always be both semantically and epistemically privileged against sentential representation. I would defend that the underlying generic reason for this privilege is precisely (3) and (4). A much more specific reason is that mechanistic explanations are usually displayed via causal chains or cause-effect sequences; in that sense, a sequential-formulaic (one-dimensional) representation could

only codify information about causal paths in a serial fashion[6], while a diagrammatic (two/three-dimensional) representation would be able to encode data simultaneously about many possible causal chains within the particular mechanism. Another important observational advantage of diagram-based mechanistic explanation, particularly in metabolic pathways and enzymatic networks, is that graphical resources can be used for effectively representing those molecular structures changing during a biological process as well as those that remain invariant. For instance, we can use a metabolic pathway diagram (containing Haworth projections as sub-diagrammatic units) as the one depicted on image 5, for developing a properly mechanistic explanation of how pyruvate can be obtained from glucose by the enzyme-catalyzed biological reaction of glycolysis. A set of causal chains is diagrammatically depicted (the so-called EMP pathway) supported by a chromatic code, wherein red means changing structures, blue arrows indicate ATP consumption and yellow ones do the same with ATP production.

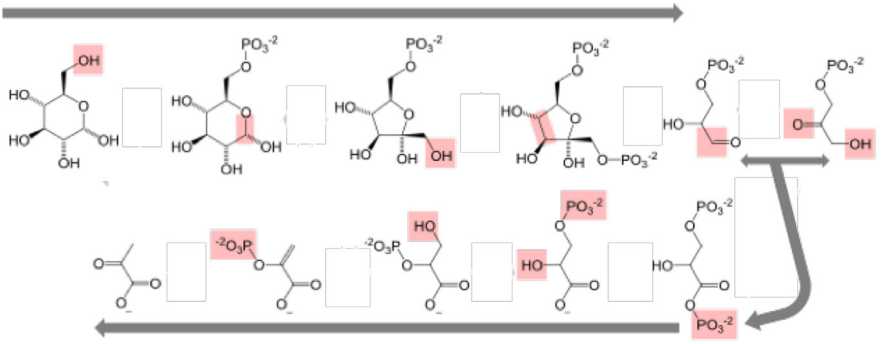


Figure 5. Dynamic explanation of glycolysis process based on Haworth projections.

Hypothesis Testing and Discovering

As a conclusion of this section, it worth mentioning briefly the highly efficient use of diagrams in the evaluation of biomolecular. Bueno [6] reported a case wherein computer-generated structural models of a particular protein were statistically compared with structural data obtained from x-ray crystallographic measurement of the same protein. Due to a poor 63 % correlation between the empirical image and the structural model (created under the theoretical assumption that the spatial configuration of the protein's surface cannot change), a new hypothesis was introduced:

[6] Of course, you can use a formulaic representation to represent multiple causal path, but always in a serial way.

> The researchers then returned to the theoretical image, and made a novel hypothesis. (…) given that the surface molecules have incomplete bonds, and are in contact with the environment, they tend to interact with it, thus changing the packing arrangement. (...) The researchers then changed the theoretical image by incorporating the assumption that surface reconstruction took place. The correlation now between the theoretical and the experimental images was 93%—a far more significant result [1, p. 663]

Due to the huge impact of the received view in the general philosophy of science, theoretical content and hypotheses were assumed to be only conveyed in symbolic-logic formulas or in ordinary language sentences; since the 90s, philosophical interest on real scientific practices lead to also regards non-linguistic formats, like the computer-rendered structural models of the example, as perfectly legitimate vehicles for abducing theoretical content. In this sense, I would strongly argue that (widely-conceived) diagrammatic activities should be regarded as the main form of creating, manipulating and rejecting biomolecular hypothesis. Thus, as the above case shows, diagrammatic procedures cannot be merely conceived as convenient semantic resources, but moreover as essential epistemic tools which also allow scientist to discover novel phenomena (i.e. rearrangement of surface configuration took place in organic crystals), and therefore as essential for the disciplinary development of molecular biology.

5. Conclusion

Summarizing, all along this paper it has been defended a twofold answer to the question of why might molecular biologist prefer diagrammatic representations of information rather than sentential ones. On one hand, I have argued against Perini [5] that it is not even possible to represent structurally complex biomolecules by means of sentential or formula-based vehicles precisely because it would be required (i) a logarithmic quantity of computational resources to uncodify the informational content of such semantic object or (i) non-human cognitive abilities for an epistemic agent to be able to understand it. Then, diagrammatic forms of conveying information would be semantic and epistemically indispensable at this level of biomolecular structural complexity. On the other hand, I have also defended the thesis that, even when diagrammatic systems are dispensable, they are much more effective than linguistic representations in both codifying biomolecular information and in obtaining knowledge from that empirical domain. The reason for this semantic-epistemic effectiveness could be found in the "representational

immediacy" [4] of geometry-exploiting diagrammatic mechanisms (according to Shimojima [9]) while codifying relevant structural biomolecular data. Finally, I have pointed out that both the defended indispensability and effectiveness of this kind of syntactically sophisticated visual representation might satisfactorily explain its epistemic virtuosity in many real molecular biology practices: diagrammatic understanding enable to grasp complex biochemical concepts as well as make possible to carry functional-structural explanations, additionally they can also be used to develop other explanatory modalities, like mechanistic-dynamic explanations of enzymatic processes or topological explanations of how proteins unfold, as well as to evaluate novel hypotheses and discover new phenomena. In conclusion, diagrams should not be conceived as mere useful and illustrative tools for depicting molecules or biochemical reactions; moreover, these representational systems are foundational, essential and pivotal in a vast plurality of ways (e.g. cognitively, computationally, representationally, explanatorily, predictively…) to this scientific domain as indispensable and effective vehicle of biomolecular knowledge.

Acknowledgements

Firstly, I am really grateful to an anonymous reviewer for her quite useful comments. I should also thank Javier Suarez, Laura Nuño de la Rosa, Roger Deulofeu and Álvaro Ropero for fruitful discussions. This paper was originally presented during the "Eight Philosophy of Biology and Cognitive Sciences Research Workshop" at the Complutense University of Madrid (May 2018) and it was developed with the academic support of LOGOS (research group in analytic philosophy) and the reach group "Laws, explanation and realism in physical and biomedical sciences" (FFI2016-76799-P). The research crystalized in this paper was also economically funded by the Spanish Ministry of Education (FPU16/00774).

References

[1] - Bechtel, W & Abrahamsen, A. 2005, "Explanation: a mechanist alternative", *Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421-41.

[2] - Bueno, O. 2012, "Styles of reasoning: A pluralist view", *Studies in History and Philosophy of Science Part A* 43 (4): 657-65.

[3] - Jones, N. 2014, "Bowtie Structures, Pathway Diagrams, and Topological Explanation", *Erkenntnis* 79 (5):1135-55.

[4] - Kulvicki, J. 2010, "Knowing with images: Medium and message", *Philosophy of Science* 77 (2): 295-313.

[5] - Larkin, J. & Simon H. 1987, "Why a diagram is (sometimes) worth ten thousand words", *Cognitive Science* 11: 65-99.

[6] - Suárez, M. 2002, "An inferential conception of scientific representation", *Philosophy of Science* 71 (5): 767-79.

[7] - Perini, L. 2005, "Explanation in two dimensions: Diagrams and biological explanation", *Biology and Philosophy* 20 (2-3): 257-69.

[8] - Perini, L. 2012, "Form and Function: A Semiotic Analysis of Figures in Biology Textbooks", N. Anderson & M. Dietrich (ed.), *The Educated Eye Visual Culture and Pedagogy in the Life Sciences*, 235-54.

[9] - Perini, L. 2013, "Diagrams in Biology", *The Knowledge Engineering Review* 28 (3): 273-86.

[10] - Sheredos, B.; Burnston, D.; Abrahamsen, A. & Bechtel, W. 2013, "Why Do Biologists Use So Many Diagrams?", *Philosophy of Science* 80 (5): 931-44.

[11] - Stapleton, G.; Jamnik, M. & Shimojima, A. 2017, "What Makes an Effective Representation of Information: A Formal Account of Observational Advantages". *Journal of Logic, Language and Information* 26 (2): 143-77.

[12] - Taylor, P. & Blum, A. 1991, "Pictorial Representation in Biology", *Biology and Philosophy* 6: 125-34.