To be presented at the PSA 2014, Chicago, 6-8 November 2014.
Please do not cite without permission.

On the Limits of Causal Modeling: Spatially-Structurally Complex Phenomena

Marie I. Kaiser

Word count: 7.170

Abstract

This paper examines the adequacy of causal graph theory as a tool for modeling biological phenomena and formalizing biological explanations. I point out that the causal graph approach reaches it limits when it comes to modeling biological phenomena that involve complex spatial and structural relations. Using a case study from molecular biology, DNA-binding and -recognition of proteins, I argue that causal graph models fail to adequately represent and explain causal phenomena in this field. The inadequacy of these models is due to their failure to include relevant spatial and structural information in a way that does not render the model non-explanatory, unmanageable, or inconsistent with basic assumptions of causal graph theory.

1. Introduction

In recent decades major advances have been made in formalizing causation and causal inference (Spirtes, Glymour, and Scheines 2000; Pearl 2000) and in using these formalisms to address traditional philosophical issues such as scientific discovery and the nature of scientific explanations (Woodward 2003, Hitchcock and Woodward 2003, Woodward and Hitchcock 2003). At the heart of these formal theories lie causal models that involve elements such as causal graphs, probability distributions, Bayesian nets, and structural equations which satisfy certain conditions, most prominently the Causal Markov Condition (more on this in Section 2). Causal models are appreciated because they allow for inferring causal relations from observed probabilistic correlations, for predicting the effects of manipulations and interventions, and because they can be used for representing and explaining causal relationships in very general, formal terms.

Proponents of the causal modeling approach usually emphasize and exemplify the wide scope of this approach. In recent years several authors have, for instance, shown that the causal modeling approach can also be applied to mechanistic explanations in biology and medicine (Casini et al. 2011; Gebharter and Kaiser 2014; Clarke, Leuridan, and Williamson forthcoming; Gebharter forthcoming). Also Woodward's interventionist theory of causation and causal explanation (2003) that makes extensive use of causal graphs is supposed to be applicable to a very wide range of causal relationships, including those in the biological sciences (2010, 2011, 2013).

In this paper I agree that causal modeling is a powerful approach to formally represent, explain, and discover causal relations. However, I also think that its scope

should not be overestimated and that it is important to recognize also the *limits* of the causal modeling approach. This paper explicates one of these limits: the explanation of spatially and structurally complex biological phenomena.[1] According to my line of criticism, formal causal models fail to offer adequate causal explanations of biological phenomena that essentially involve *complex spatial and chemical-structural relations*. This failure is due to the fact that causal graphs only provide causal difference-making information of the sort: A change in the value of *X* would under suitable conditions change the value (or probability distribution) of *Y*. The explanations of some biological phenomena, however, seem to be richer than this: these explanations do not only represent causal relations but also and prominently spatial relations and biochemical structures (such as the conformation and chemical structure of macromolecules, the spatial orientation and fitting of macromolecules to each other, and the complementarity of chemical structures). Based on the analysis of a case study from molecular biology, DNA recognition and binding by gene regulatory proteins, I show that the formal tools of causal graph theory are too impoverished to model biological processes that involve complex spatial-structural relations.

Interestingly, Woodward (2011) has basically conceded this point, but he does not see this as a limitation of his causal modeling or interventionist approach to scientific explanation. Taking a relaxed stance, he argues that complex spatiotemporal information

[1] Other limitations may be that causal models fail to account for the complex dynamics that biological phenomena such as biological clock mechanisms involve (Weber manuscript) and that causal models provide a confusing and ontologically inadequate view of the entities and activities involved in biological mechanisms (Gebharter and Kaiser 2014).

can just be added back to the backbone of causal difference-making information and can be used to "organize" (2011, 423) or "fine-tune" (2013, 55) causal difference-making information. This paper shows that things are not that easy. Some biological processes involve complex spatial and chemical-structural relations that are central to explaining these processes but that cannot be represented in causal graph models without rendering the model un-explanatory, unmanageable, or contradictory.

I proceed as follows. In Section 2 I briefly review the core notions and assumptions made in the causal modeling literature such as the notion of a causal graph, a probability distribution, and the Causal Markov Condition. In Section 3 I introduce the case study on which my analysis relies by explaining the three kinds of fit that are involved in DNA-protein recognition and binding (Section 3.1) and by specifying what exactly the phenomenon is that biologists seek to explain and which constraints on the adequacy of modeling this phenomenon follows from this (Section 3.2). In Section 4 I construct a causal graph model of DNA-protein recognition and then point out the shortcomings it has (Section 5). To elaborate my argumentation I discuss two possible objections that a causal modeler could raise: first, one might argue that the proposed causal graph model it not good enough, but that it is possible to construct an alternative model that *does* include the relevant spatial-structural information and *is* adequate (Section 5.1), second, one might object that I am not making an interesting or novel point as the explanation of DNA-protein recognition is non-causal and causal graph theory was not intended to model non-causal explanations (Section 5.2).

2. Causal Modeling

The most frequently used causal models can be grouped into two kinds: causal Bayesian networks and structural equation models (which are distinct but closely related, cf. Spirtes 2010). In this paper I focus on causal Bayesian networks (which can also be called causal graph models) and leave structural equations aside. Causal graph models combine mathematics and philosophy: the mathematical elements are directed acyclic graphs (DAGs) and probability theory (with focus on conditional independence); the philosophical elements are assumptions about the relationship between causation and probability (Spirtes, Glymour, and Scheines 2000).

A *directed acyclic graph* (DAG, also called *G*) is an ordered pair $G = \langle V, E \rangle$, where *V* is a set of variables and *E* is a set of directed edges (that are graphically represented by arrows) that have the variables in *V* as their vertices. A variable can be binary, its values representing for instance the instantiation or non-instantiation of some property, or variables can have multiple values or even be continuous. Here is an example of a DAG:

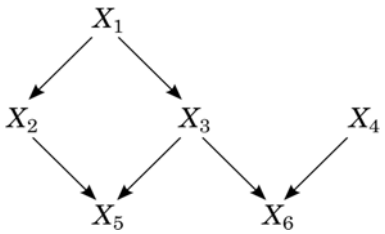

Fig. 1: An example of a directed acyclic graph (DAG)

Besides the DAG, a causal model consists of a *probability distribution P* over the variable set *V* that assigns a value to every variable in *V* such that the sum of all assigned variables in *V* equals 1. The pair of DAG *G* and probability distribution *P* over *V* is referred to as a

Bayesian network if and only if *G* and *P* satisfy the Markov Condition.[2] A DAG becomes a *causal graph* as soon as its edges are interpreted causally: an edge leading for instance from variable X_2 to variable X_5 (see Figure 1) is interpreted such that X_2 is a direct cause of X_5. When DAGs are interpreted causally the Markov condition (and d-separation) are assumed to be the correct connection between causal structure and probabilistic independence. This assumption is called the *Causal Markov Condition* and it can be formulated as follows:

Causal Markov Condition (CMC)

G and *P* satisfy the Causal Markov Condition if and only if for every variable *X* in *V*, $\text{INDEP}_{Pr}(X, \text{N-Des}(X) \mid \text{Pa}(X))$.

In words, a directed acyclic graph *G* and a probability distribution *P* over variable set *V* satisfy the Causal Markov Condition iff every variable *X* in *V* is probabilistically independent of *X*'s non-descendants (N-Des(*X*)) given *X*'s parents (Pa(*X*)). For example, X_2 is probabilistically independent from its non-descendants (i.e. X_3, X_4, and X_6) given X_1. CMC captures the intuition that conditioning on all common causes (e.g., X_1) and on intermediate causes breaks down the probabilistic influence between two formerly correlated variables (e.g., X_2 and X_3). Since causal models satisfy CMC they allow, for instance, for probabilistic prediction and manipulation.

[2] In DAGs the Markov Condition turns out to be equivalent to a more generally useful graphical relation: d-separation (Pearl 1988).

In the remaining sections of this paper I apply these formal tools of causal graph theory to a paradigmatic example of a biological explanation: the explanation of how gene regulatory proteins recognize and bind to a specific DNA region. My analysis will show that the causal modeling approach reaches its limits when it comes to representing and explaining spatially-structurally complex biological phenomena.[3]

3. DNA-Recognition and -Binding by Proteins

The regulation of the expression of genes is an important process in living beings. Differential gene expression is for instance the basis for cell differentiation during development. In eukaryotes, gene expression is regulated at different steps, for instance, a cell controls when and how often genes are transcribed (transcriptional control). The most important elements in regulating gene transcription are *gene regulatory proteins* (also called transcription factors). These proteins can recognize and bind to specific nucleotide sequences without having to open the double helix. This is due to the fact that the surface

[3] One might object that even in biological fields where spatially and structurally complex phenomena are studied (e.g., in protein folding and interaction research) graphical models turn out to be useful (cf. Balakrishnan et al. 2010, Thomas et al. 2009). I don't think, however, that cases like these are counterexamples to my thesis. First, the applied graphical models are *undirected probabilistic* graphical models, not direct causal graph models of the kind I discuss here. Second, in these studies graphical models are not used to directly *represent* or to *explain* phenomena such as protein folding or protein-protein interaction. Rather, they are used as techniques or tools, for instance, to guide the design of new protein sequences. Hence, graphical models might be useful in that domain, but not for the purposes I discuss in this paper.

of the DNA (in particular, its major grooves) presents a distinctive pattern of hydrogen bond donors and acceptors, and hydrophobic patches. A gene regulatory protein recognizes a specific DNA sequence because its surface is extensively complementary to the special surface features of the major groove of DNA. In other words, the protein *fits* well to a certain region of the DNA.

3.1 Three Kinds of Fit

This fitting of a gene regulatory protein to a specific DNA binding site can be interpreted as involving three interwoven aspects: a spatial aspect, a structural aspect, and a causal aspect. The *spatial fit* refers to the fact that the spatial conformation of the protein is such that it allows certain parts of the protein being placed in the major groove of the DNA double helix, at the DNA backbone, or in its minor groove. The *structural fit* means that the protein exhibits particular amino acid residues at certain places such that they are complementary to the functional groups that the nucleotide bases of the DNA exhibits at certain places. The spatial and the structural fit give rise to a *causal fit*. That is, the spatial orientation and the structural complementarity of the protein and the DNA enable that certain causal interactions between the function groups of the protein and those of the DNA take place and certain chemical bindings between them are established. These three kinds of fit can be summarized as follows:

(1) **Spatial fit**: The spatial conformation of the gene regulatory protein matches the double helix conformation of the DNA.

(2) **Structural fit**: The chemical structure of the protein (i.e., the sequence and location of its amino acids) is complementary to the nucleotide sequence of the DNA binding site.

(3) **Causal fit**: Certain amino acid residues causally interact with/make contact with certain nucleotide bases.

Consider the example of DNA recognition and binding by a *zinc finger* (ZnF). Zinc finger proteins are specific gene regulatory proteins that use zinc finger motifs to bind to DNA. The three zinc fingers of the Zif268 protein in mice, for instance, are arranged in a semicircular, C-shaped structure so that the α-helix of each zinc finger fits directly into the major groove of the DNA double strand (Pavletich and Pabo 1991). Figure 2 illustrates this spatial fit.

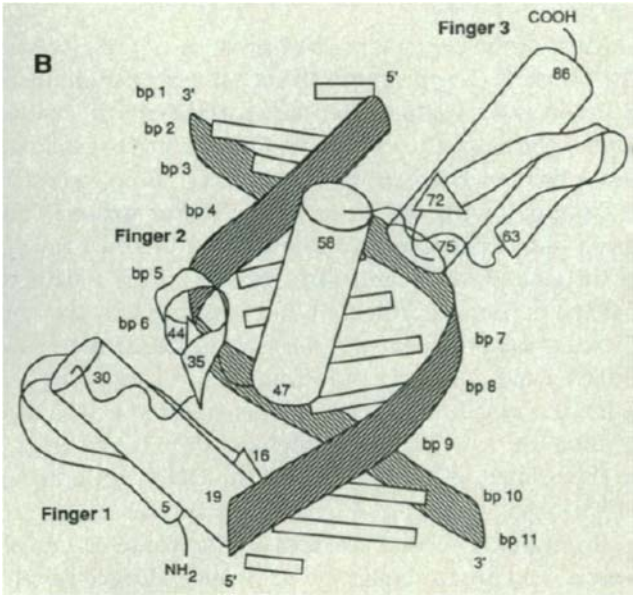

Fig. 2: Spatial fit between Zif268 and the DNA. (Pavletich/Pabo 1991, 811, Fig. 2.)

The cylinders and ribbons mark the α-helical and β-sheet regions of each finger. The shape of the Zif268 protein is such that it wraps round the DNA and the three α-helices (the cylinders) fit into the major groove.

The second kind of fit, the structural fit, is due to the fact that the chemical structure of the three zinc fingers is complementary to the nucleotide sequence of the DNA binding site. This means that the zinc finger protein possesses the "right" amino acids at the "right" places. For instance, the twenty forth amino acid of Zif268 protein is arginine, which has a positively charged residue. If the protein collides with the DNA binding site (in the right orientation) arginine is close to the nucleotide guanine, with which is can form hydrogen bonds (see Figure 3.b). Hence, the spatial and structural match between the zinc finger protein and its DNA binding region enables that the two also causally match: given that Zif268 collides with the DNA in the right orientation its three zinc fingers form extensive, characteristic contacts with the nucleotide bases (primarily along the guanine-rich DNA strand). Each finger has a similar relation to the DNA and makes its primary contacts in a three-base pair subsite. A summary of the critical base contacts is depicted in Figure 3a.

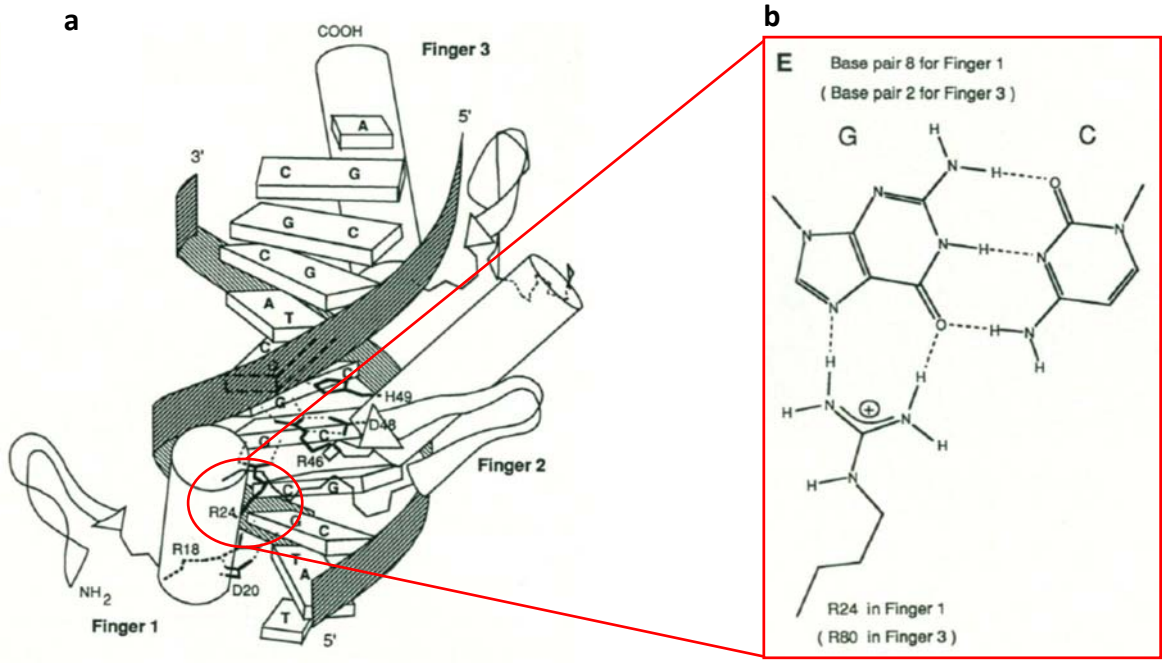

Fig. 3: Base contacts made by Zif268. (a: Summary of all critical base contacts; b: Arginine-guanine interaction that is present in finger 1 and 3; Pavletich and Pabo 1991, 812, Fig. 3.)

An example for such a contact between an amino acid residue of the zinc finger protein and a nucleotide base of the DNA is the Arginine-guanine contact (see Figure 3b). Arginine-guanine interactions seem to be responsible for much of the specificity in the zinc finger complex (Pavletich and Pabo 1991, 816).

3.2 Constraints on an Adequate Model of DNA-Protein Recognition

In this section I point out the implications that the biological literature has for the conditions under which an explanatory causal model of DNA-protein recognition is adequate. For this purpose I, at first, specify what exactly the phenomenon to be modeled

and to be explained in this case study is. Then I identify the different sorts of information that biologist treat as relevant to explaining this phenomenon. The underlying methodological assumption is that how biologists actually study and reason about DNA-protein recognition tells us which information is relevant to explaining this phenomenon and which information thus must be included in an adequate causal model.

In the biochemical literature the phenomenon is generally characterized as "DNA recognition by proteins" (Pavletich and Pabo 1991; Somers and Phillips 1992; Klemm et al. 1994) or as "protein-DNA interaction" (Luisi 1991). The target of these biochemical studies is to reveal *why* and *how* a particular gene regulatory protein (such as the zinc finger protein ZiF268 in mice) *recognizes* and *binds* to a specific DNA region.[4]

It seems to me that the phenomenon of DNA-protein recognition has (at least) two major characteristics, both of which must be captured by any adequate model of this phenomenon. First, a model of DNA-protein recognition must account for the *regular changes* from unbound proteins to DNA-bound proteins that take place under certain conditions. These regular changes include certain sub-processes, as the process of diffusion (of the protein and the DNA strand), the process of collision of the protein and the DNA, the process of recognition, and the process of binding of the protein to the DNA. Second, an adequate model of this phenomenon requires that one accounts for the *specificity* of the binding process. That is, a model must elucidate why a certain gene regulatory protein

[4] In mechanistic terms one could say that they seek to uncover the mechanism for DNA-protein recognition and binding.

recognizes and binds to a *specific* DNA region, rather than to a different region with a different nucleotide sequence.

The biochemical literature reveals further constraints on how DNA-protein recognition is adequately modeled. In Section 3.1 I have argued that biochemists provide three kinds of information when they explain how a certain gene regulatory protein recognizes and binds to a specific DNA region: first, they disclose the three-dimensional structure of the gene regulatory protein (which α helices and which β sheets it has and how they are located to each other) and its spatial orientation on the double helix when it is bound to the DNA. This amounts to showing that there is a *spatial fit* between the protein and its DNA binding region. Second, they reveal the chemical structure of the involved macromolecules and show that the chemical structure of the protein surface is *complementary* to the chemical features of the DNA sequence, that is, that there is a *structural fit* between protein and DNA binding site. Finally, biochemists point out which functional groups of the amino acid residues causally interact with which functional groups of the nucleotide bases and what the chemical nature of these interactions is (whether they are hydrogen bonds, van der Waals interactions, salt bridges, etc.). Modeling DNA-protein recognition typically involves providing a complete list of the contacts that are made between protein and DNA (e.g., Luisi et al. 1991, 502f; Pavletich and Pabo 1991, 812-814; Klemm et al. 1994, 23-25). This contact list specifies the *causal fit* between gene regulatory protein and DNA binding site.

All three kinds of information are necessary parts of an adequate causal model that provides an understanding and explanation of how a gene regulatory protein recognizes

and binds to a specific DNA region. Neglecting some of these relevant kinds of information renders the model inadequate. At least such a model would be incomplete in a way that is disastrous for its explanatory power. If one for instance ignores complex spatial and structural information and represents only the causal interactions between the functional groups of protein and DNA, the regular changes from unbound proteins to DNA-bound proteins and the specificity of the binding process will remain obscure.

4. How to Model DNA-Protein Recognition by Causal Graphs

How can we construct a formal causal model of the binding of a gene regulatory protein (e.g., Zif268) to a specific DNA binding site? The framework of causal graph theory (recall Section 2; for an introduction see, e.g., Spirtes et al. 2000) seems to be a promising tool. One possibility to model DNA-protein recognition by causal graphs (and probability distributions) is to conceive this phenomenon as a chain of causal events that leads to the binding of the gene regulatory protein Zif268 to the corresponding DNA binding site. The preceding causal events would then be the diffusion of Zif268 into the nucleus, which leads to (or allows for) the collision between Zif268 and the DNA at the corresponding binding site, which in turn causes the binding of Zif268 to the DNA. This chain of causal events can be represented by the following causal graph model (let us call it M_1):

Fig. 4: Causal graph model M_1 for DNA-protein recognition.

D, *C*, and *B* are binary variables. *D* stands for the diffusion of Zif268 into the nucleus, *C* for the collision between Zif268 and the DNA at a certain region, and *B* for the binding of Zif268 to a certain DNA region. Each of the three variables can take one of the two values "taking place" and "not taking place". *D* is a direct cause of *C*, which is a direct cause of *B* (represented by the arrows). The "+" stands for a positive causal influence: raising the probability that *C* takes the value "taking place" raises the probability that *B* takes the value "taking place". Hence, the causal graph model allows for making predictions and testing them by interventions. For instance, the causal dependency relation between *C* and *B* can be tested by lowering the probability of the collision of proteins and DNA strands by dilution. According to the causal model the consequence of this intervention should be that also the probability for the binding of the protein to the DNA decreases.

A possible objection to M_1 is that it does not account for the causal fit between the gene regulatory protein Zif268 and the DNA binding site (not to mention the spatial and the structural fit). In the model there is only a single variable *B* that stands for the binding of Zif268 to a certain DNA region. No further information about *which* causal interactions between *which* amino acid residues and nucleotide bases take place and bring about the binding is included in the model. But this information seems to be crucial for an understanding of the causal fit between Zif268 and the DNA binding site. This objection can be avoided by choosing a larger number of more fine-grained variables. Instead of representing the collision of the entire protein with the entire DNA binding site by a single variable (*C*) this event is broken down to into various sub-events (e.g., amino acid residue x_1 collides with nucleotide base y_1, amino acid residue x_2 collides with nucleotide base y_2,

etc.), each of which is then represented by a distinct variable (C_1, C_2, etc.). Likewise, not the binding of the protein to the binding site in general is represented (by a single variable B). Rather, different variables (B_1, B_2, etc.) represent the binding of different amino acid residues to different nucleotide bases.[5] The resulting causal graph model M_2 is the following:

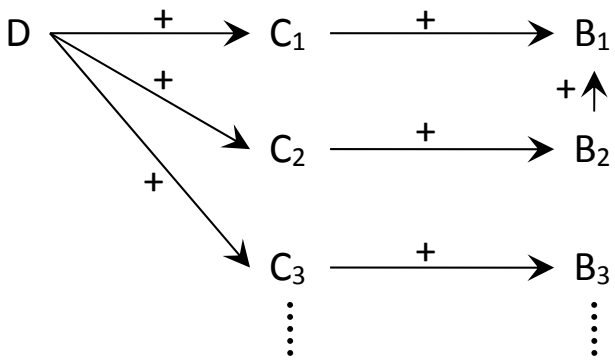

Fig. 5: Causal graph model M_2 for DNA-protein recognition.

C_1, …, C_n and B_1, …, B_n are binary variables. Each of them can take one of the two values "taking place" and "not taking place". C_1, …, C_n stand for the collision (or spatial proximity) of a certain amino acid residue of Zif268 with a certain nucleotide base, and B_1, …, B_n for the formation of a certain set of bonds/interactions between residues and bases.

[5] One might claim that these variables (C_1, C_2, C_3, B_1, B_2, and B_3) are not only more fine-grained, but also lower-level variables (i.e., variables that represent entities that are located on a lower organizational level than the entities represented by the former variables C and B). An even more "fine-grained" model would distinguish also among the different contacts that are made between one amino acid and one nucleotide.

The arrows between the variables represent that *D* is a direct cause of C_1, C_2, and C_3, which are direct causes of B_1, B_2, and B_3. Since the binding of a particular amino acid residue to a particular nucleotide base may raise the probability that another binding between the protein and the DNA is established the causal graph model also contains arrows between the *B*-variables (such as the arrow between B_1 and B_2).[6]

5. Where the Limits of Causal Graph Models Lie

In Section 3.2 I have argued that three kinds of information are crucial to adequately explaining why and how gene regulatory proteins, such as Zif268, specifically and regularly recognize and bind to certain DNA regions: information about the spatial fit, about the structural fit, and about the causal fit between protein and DNA. M_2 is superior to M_1 since it succeeds in representing the causal fit, that is, M_2 includes information about which amino acid residues causally interact with and establish chemical bindings to which nucleotide bases. However, M_2 fails to provide an understanding of the spatial fit as well as of the structural fit between Zif268 and DNA. It entails no information about the conformation of Zif268 and about how it matches the double helix shape of the DNA (spatial fit). It also does not represent the complementarity of the chemical structure of the protein (i.e., the sequence and location of its amino acids) to the nucleotide sequence of the DNA binding site (structural fit). M_2 includes information about which amino acids causally interact with which nucleotides (contained in the variables $B_1,\ldots B_n$). But this is, as such, no direct or complete information about the structural fit between protein and DNA:

[6] The same might be true regarding the *C*-variables.

The fact that protein and DNA causally fit allows only *inferences* about the complementarity of their chemical structures and it sheds light on only *some* parts of the chemical structure of DNA and protein. The causal graph model M_2 is thus inadequate because it leaves out spatial and structural information that is crucial for explaining why and how the zinc finger protein specifically and regularly recognizes and binds to DNA.

In the next two subsections I elaborate on my argument by addressing two possible objections that the causal modeler could raise. First, one might admit that M_2 is inadequate, but argue that it is possible to construct an *alternative causal graph model* that accounts for the spatial and structural fit between Zif268 and DNA and thus is adequate (Section 5.1). Second, one might agree that causal graph theory is not the appropriate tool to model spatially and structurally complex phenomena, but object that this is neither an interesting nor innovative insight as the explanations in these cases are *non-causal* and causal graph theory was not intended to model non-causal explanations (Section 5.2).

5.1 Shortcomings of Alternative Modeling Strategies

The first objection says that even if the causal graph model M_2 is inadequate it is possible to revise M_2 in a way that renders it adequate. In other words, an opponent might argue that it is possible to construct an alternative causal graph model that includes all the complex spatial and structural information that is relevant to explaining DNA-protein recognition into M_2. But how could that be done? I see two different ways one could go.

One option is to include information about the spatial conformation of Zif268 into the characterization of the variables $C_1, \ldots, C_n$ or $B_1, \ldots, B_n$. For instance, one could

characterize B_2 not as the "formation of a certain set of contacts between amino acid residue x_2 and nucleotide base y_2" but as the "formation of a certain set of contacts between amino acid residue x_2 and nucleotide base y_2, where x_2 is covalently bound to x_1 and x_3, forms a salt bridge to x_{14}, has a close distance also to bases y_3 and y_9, and so on".

But this strategy encounters several problems. First, storing complex spatial and structural information into the characterization of the variables renders M_2 *unmanageable* because the measurement of the values of the variables becomes very complicated or even unfeasible.[7] Second, this strategy results in a causal model in which the spatial and structural information is *highly fragmented* because information about the spatial conformation of Zif268 and of the DNA binding site, about their spatial orientation to each other, about the chemical structures of protein and DNA, and about the complementarity of these structures is distributed over the many variables $C_1, \ldots, C_n$ and $B_1, \ldots, B_n$. This fragmentation is devastating for the explanatory power of the causal model since the resulting model fails to elucidate why and how Zif268 spatially and structurally fits to the DNA bind site (e.g. that the whole protein Zif268 has a C-shaped structure so that the α-helix of each zinc finger fits directly into the major groove of the DNA double strand). Third, this strategy of storing complex spatial and structural information into M_2 gives rise to a causal model in which a great deal of the explanatorily relevant information is contained in the characterization of the variables, not in the represented causal dependency relations. This suggests that the causal dependency relations are less informative and bear

[7] One might even argue that spatial and structural information is so complex that adding them to the causal model is *not feasible in practice*, but merely possible in principle.

less explanatory weight than the characterization of the variables. But this seems to *conflict* with the typical way of how causal graph models are conceived and reinforces the impression that one tries to add something that does not really fit.

A second option is to add one or more variables to the causal model that are supposed to represent the spatial and structural fit between protein and DNA. One could, for instance, add the variable *S* that stands for the protein Zif268 having a certain conformation and chemical structure, the DNA having a certain shape and chemical structure, Zif268 being oriented towards the DNA binding site in a certain way, and the structures of protein and DNA being complementary.[8] In its simplest form *S* would be a binary variable, which can take one of the two values "being realized" and "not being realized". The resulting causal model M_3 would look as follows:

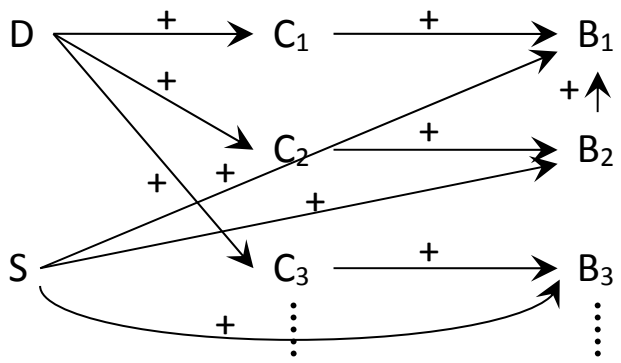

Fig. 6: Causal graph model M_3 for DNA-protein recognition.

[8] This option implies that one accepts that the conformations of, spatial relations between, and chemical structures of protein and DNA are *difference makers* with respect to their binding and that – given an interventionist, counterfactual theory of causation – information about the spatial and structural fit also is *causal* information (see Section 5.2).

M_3 accounts for the fact that whether or not a certain amino acid base makes contact to a certain nucleotide base depends not only on whether protein and DNA collide, but also on whether they spatially and structurally fit (i.e. on whether the amino acid base and the nucleotide base are complementary and whether they are proximate enough).

The revised causal model M_3, however, is still not satisfactory. First, it has the feel of *putting* the missing relevant spatial and structural information *in "by hand"*:[9] something that does not smoothly, automatically fit must be added under additional, atypical efforts. What reinforces this feeling is that all different elements of the spatial and of the structural fit are represented by a single variable *S*. It would be more adequate to add different variables (S_1, …, S_n) each for protein conformation, DNA shape, spatial orientation of protein towards DNA, amino acid sequence of Zif268, chemical structure of the DNA binding site, and for the complementarity of protein and DNA structure. These different variables could then also be quantitative variables (rather than binary ones), which might represent, for instance, relative distances among sets of protein molecules and sets of DNA molecules and possible combinations of proteins with a certain sequence binding to DNA molecules with a certain sequence. But this strategy encounters the problem that the variables are no longer *conceptually independent* from each other (e.g., a specific protein conformation requires a specific amino acid sequence, or the complementarity of protein and DNA binding site means that certain kinds of causal interactions can take place), which violates a central requirement of causal graph theory. Finally, even if we could add different variables for each element of spatial and structural fit the characterization of the

[9] Thanks to an anonymous referee for pointing this out.

variables would be very complex. For instance, the variable that stands for the complementarity of protein and DNA structure would have to be characterized in a way that shows not only *that* the protein structure is complementary to the structure of the DNA binding site, but that also illuminates *why* they are complementary and *what* this amounts to. This requires storing detailed information about which functional groups of which amino acid residues fit (for which reasons) to which functional groups of which nucleotide bases into the characterization of variables. The second option of including complex spatial and structural information into the causal graph model thus encounters the same objections as the first option: storing complex information into the characterization of the variables renders the causal model *unmanageable* and it *contradicts* the basic assumption of causal graph theory that causal dependency relations between variables are a *central* element of causal models (i.e. that they are not less informative and bear less explanatory weight than the variables themselves).

To conclude, even if it might be in principle or technically possible to include information about complex spatial relations and chemical structures into a causal graph model, this can only be done at the expense of the adequacy of the causal model as it renders the causal model non-explanatory (because the information gets highly fragmented), unmanageable, and entails inconsistencies with basic assumptions of causal graph theory (e.g. that variables must be conceptually independent and that causal

dependency relations are central). Hence, the causal graph approach reaches its limits when it comes to explaining spatially and structurally complex biological phenomena.[10]

5.2 Just a Matter of Non-causal Explanations?

A second line of criticism agrees with me that biological phenomena that involve complex spatial relations and chemical structures cannot be adequately modeled by the tools of causal graph theory, but argues that this is neither an interesting nor novel insight. Everybody agrees, so the criticism goes, that there are non-causal explanations in science such as the explanation of DNA-protein recognition and that *causal* graph theory is not the appropriate tool for representing such *non-causal* explanations. So what is the big news?

I agree that if the explanation of DNA-protein recognition were non-causal it would be weird to investigate why *causal* graph theory fails to adequately represent this *non-causal* explanation. But I think the process of DNA-protein recognition clearly is a *causal* process, which involves causal interactions between the functional groups of the protein and of the DNA binding site, and that this process must be explained causally, too.[11] What is interesting about this explanation is that, besides causal relations, it also and prominently

[10] Jantzen and Danks (2008) have recently argued that topological properties of complex molecules can be represented by graphical models. One might suggest that this challenges the argument that I provide in this paper. Note, however, that the graph models I discuss are very different from the ones that Jantzen and Danks propose. They use *undirected* graphs to represent topological *properties* of molecules, whereas I use *directed*, causal graphs to represent *processes* of DNA-protein binding.

[11] I regard explanations as causal if they explicitly (but not necessarily only) represent causal relations.

represents relations that are *non-causal* (or at least not directly causal): the shape of the gene regulatory protein Zif268, how Zif268 is spatially oriented to the DNA double helix (i.e. their spatial fit), the chemical structures of Zif268 and DNA binding site (i.e. the sequence of amino acids and nucleotides), and the complementarity of their chemical structures. The question of whether these kinds of non-causal information can be adequately represented in causal graph model or whether they constitute a limit of the applicability of the causal modeling approach is neither uninteresting nor has it already been sufficiently addressed.[12]

One might question whether information about the spatial and structural fit between DNA and protein is in fact non-causal as both kinds of fit *make a difference* to the binding of DNA and protein (in other words, the binding counterfactually depends on there being a spatial and structural fit). Assuming a counterfactual view of causation one could then argue that (besides the collision) the spatial and structural fit between DNA and protein *cause* their binding. It is important to note that my analysis of the limits of causal graph theory is *compatible* with such an interpretation. Characterizing complex spatial and structural information as causal makes it even more urgent and interesting to analyze whether these kinds of information can be included in causal models.

The discussion about the allegedly non-causal character of the explanation of DNA-protein recognition points to another important issue: in explanations of spatially and

[12] This question has already been discussed with regard to constitutive or part-whole relations (e.g., Casini et al. 2013; Gebharter and Kaiser 2014). This paper extends the discussion to spatial and structural relations.

structurally complex biological phenomena causal and non-causal information often cannot be easily separated, but rather are *interwoven* and highly integrated. For example, in my case study the structural fit between a gene regulatory protein and the DNA binding site is specified by referring to non-causal information such as information about the sequence of amino acids and nucleotides and to information about the complementarity of protein and DNA structures, which seem to be implicitly causal or to be closely connected to causal information. The complementarity of protein and DNA can be spelled out either dispositionally or counterfactually: to say that protein and DNA are complementary means to say that protein and DNA have the disposition to causally interact/form certain kinds of bounds (if they collide or are proximate enough) or that if protein and DNA collided or were proximate enough they would form certain kinds of bounds. Both explications refer to causal information about the formation of bounds between protein and DNA (i.e. information about the causal fit). Furthermore, the causal part of the explanation of DNA-protein recognition seems to rely on its non-causal part because to understand the causal fit between protein and DNA binding site *requires* understanding how they spatially fit together and that their structures match. Hence, even if it is possible to entangle three respects how a gene regulatory protein fits to a specific DNA binding site (recall Section 3.1) in fact these three kinds of fit and the causal and non-causal information they invoke are interwoven.

This entanglement of causal (difference-making) information with non-causal (spatial-structural) information poses a challenge to Woodward's argument that spatial information can simply be added to the backbone of causal difference-making information

and can be used to fine-tune causal information (2011, 2013). His argument requires that causal difference-making information and spatial-structural information are easily separable. But biological practice shows that in some cases information about causal and non-causal relations is closely related and interwoven in such complex ways that it is not possible to clearly tell them apart.

6. Conclusions

I agree that causal modeling is central to scientific practice and that formal theories of causal modeling and explanation, such as causal graph theory, are powerful. However, I think that their significance is not universal and that it is important to notice also the *limits* of causal graph theory. In this paper I have used an example from molecular biology to reveal one of these limitations: spatially-structurally complex phenomena. I have shown that causal graph models fail to provide explanations of biological processes that involve complex *spatial-structural* relations (such as DNA-protein recognition).

The goal of this paper has not been to argue that there exist kinds of explanation in the biological sciences that are non-causal. The process of how a gene regulatory protein regularly recognizes and binds to a specific DNA region clearly is a *causal* process. But these causal relations are not the only kind of relations that is relevant to understanding and explaining the phenomenon of DNA-protein recognition. Besides causal difference making information, any adequate causal model of this phenomenon must also and prominently include *spatial and structural information* (i.e. information about protein and DNA conformation and chemical structure, about how they spatially fit, about why their

structures are complementary). This non-causal (spatial and structural) information is often entangled with causal difference making information. But exactly this is, as I have argued, the point at which the *limits* of causal graph models become apparent.

My central argument in this paper has been that the formal tools of causal graph theory are inappropriate to model and to explain spatially and structurally complex biological phenomena (e.g. DNA-protein recognition) because they result in causal models which either ignore the importance of spatial and structural relations all together (such as M_1 and M_2) or which try to include the relevant spatial and structural information but, in so doing, render the causal graph models non-explanatory, unmanageable, or inadequate because they conflict with basic assumptions of causal graph theory (such as M_3). This argument does not erode the significance of formal approaches to causal modeling, but it demonstrates that their scope is limited.

References

Balakrishnan, S., Kamisetty, K., Carbonell, J. G., Lee, S.-I., Langmead, C. J. 2010. "Learning generative models for protein fold families." Proteins 79(4): 1061-1078.

Casini, L., P. M. Illary, F. Russo, and J. Williamson 2011. "Models for Prediction, Explanation and Control: Recursive Bayesian Networks." Theoria. An International Journal for Theory, History and Foundations of Science 70(1): 5-33.

Gebharter, A. and M. I. Kaiser 2014. "Causal Graphs and Biological Mechanisms." In Explanation in the Special Sciences. The Case of Biology and History by M. I. Kaiser, O. Scholz, D. Plenge and A. Hüttemann. Berlin: Springer, 55-85.

Gebharter, A. forthcoming. "A formal framework for representing mechanisms?" Philosophy of Science, 81(1).

Jantzen, B., Danks, D. 2008. "Biological Codes and Topological Causation." Philosophy of Science 75(3): 259-277.

Hitchcock, C. and J. Woodward 2003. "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." Noûs 37(2): 181-199.

Klemm, Juli D., Mark A. Rould, Rajeev Aurora, Winship Herr, and Carl O. Pabo 1994. "Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules". Cell 77: 21-32.

Luisi, B.F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, P. B. Sigler 1991. "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA". Nature 352:497-505.

Pavletich, Nikola P. and Carl O. Pabo 1991."Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A". Science, New Series 252:809-817.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems. San Mateo CA.: Morgan and Kaufman.

Pearl, J. 2000. Causality. Models, Reasoning, and Inference. Cambridge: Cambridge University Press.

Somers, W. S. and S. E. Phillips 1992. "Crystal structure of the met repressor-operator complex at 2.8 A resolution reveals DNA recognition by beta-strands". Nature 359:387-393.

Spirtes, P. 2010. "Introduction to Causal Inference." Journal of Machine Learning Research 11: 1643-1662.

Spirtes, P., Glymour, C. and R. Scheines. 2000. Causation, Prediction, and Search. Cambridge, Mass.: MIT Press.

Thomas, J., Ramakrishnan, N., Bailey-Kellogg, C. 2009. "Graphical models of protein-protein interaction specificity from correlated mutations and interaction data." Proteins 76(4): 911-929.

Woodward, J. 2003. Making Things Happen: A Theory of Causal Explanation. New York: Oxford University Press.

Woodward, J. 2010. "Causation in biology: stability, specificity, and the choice of levels of explanation." Biology & Philosophy 25: 287-318.

Woodward, J. 2011. "Mechanisms Revisited." Synthese 183(3): 409-427.

Woodward, J. 2013. "Mechanistic Explanation: Its Scope and Limits." Proceedings of the Aristotelian Society Supplementary Volume 87(1): 39-65.

Woodward, J. and C. Hitchcock 2003. "Explanatory Generalizations, Part I: A Counterfactual Account." Noûs 37(1): 1-24.

Weber, Marcel (manuscript): "On the Incompatibility of Dynamical Biological Mechanisms and Causal Graph Theory", presented at the PSA 2014, Chicago, 6-8 November 2014.