

## **A relational-constructionist account of protein macrostructure and function**

Gil Santos<sup>1</sup>, Gabriel Vallejos<sup>2</sup>, Davide Vecchi<sup>3</sup>

(All authors contributed equally to the present work)

**Abstract** One of the foundational problems of biochemistry concerns the conceptualisation of the relationship between the composition, structure and function of macromolecules like proteins. Part of the recent philosophical literature displays a reductionist bias, that is, the endorsement of a form of microstructuralism mirroring an out-dated biochemical conceptualisation. We shall argue that such microstructuralist approaches are ultimately committed to a potentialist form of micro-predeterminism whereby the macrostructure and function of proteins is accounted for solely in terms of the intrinsic properties and potentialities of the components of the primary structure as if they were self-contained or essentially immutable entities. We shall instead suggest that a conceptualisation of the relationship between proteins' composition, structure and function consistent with contemporary biochemical practice should account also for the causal role of the cellular, organismal and environmental relations in protein development. The analysis of the folding process we propose suggests that microstructure-laden reductionist approaches are ontologically indefensible. Rather than a potentialist form of micro-predeterminism, our

---

<sup>1</sup> Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal and Centro de Filosofia das Ciências da Universidade de Lisboa, Faculdade de Ciências, Campo Grande, 1749-016 Lisboa, Portugal.

<sup>2</sup> Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile; Departamento de Filosofía, Universidad de Santiago de Chile, Santiago, Chile

<sup>3</sup> Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; Centro de Filosofia das Ciências da Universidade de Lisboa, Faculdade de Ciências, Campo Grande, 1749-016 Lisboa, Portugal and Departamento de Filosofía, Universidad de Santiago de Chile, Santiago, Chile.

analysis ultimately supports a relational-construction-based view of protein development and potentialities formation, which requires an indispensable analysis of the dynamical interplay between the micro-level of the parts and the macro-level of the relational structures of their systems.

## **1. Characterising protein microstructuralism**

An important foundational problem in biochemistry concerns the conceptualisation of the relationship between molecular composition, macrostructure and function of proteins. Proteins are macromolecules made out of amino acidic residues linked together by peptide bonds into linear chains that, during a complex developmental process, acquire a specific three-dimensional structure that makes them functional in a relevant environment, what is called “native structure” by biochemists. Biochemically speaking, a proper conceptualisation of protein development needs to account for each of its relevant stages, from the formation of the peptide bond between residues to the acquisition of topological and architectural features through the folding process, from the formation of native structure to post-folding modifications, etc. It is within this multi-step developmental process that the characteristic structural features of proteins, such as active sites and interfaces, are formed. And it is these features that account for the protein’s functional capacities such as the catalysis of chemical reactions or the possibility to interact with other molecules or proteins. The primary focus of many recent contributions in the philosophy of biochemistry is the important issue of classification (Slater 2009, Bartol 2016, Havstad 2018). Significantly, some of these contributions betray a reductionist bias concerning the conceptualisation of the relationship between composition, macrostructure and function (Goodwin 2011, Tahko 2019). As we shall show in section 2, this bias has been inherited from out-dated and oversimplified models of protein folding. This reductionist bias can be characterised as an endorsement of some form of microstructuralism (Hendry 2006), that is, a brand of reductive physicalism according to which all systems’ properties can be taken as fully determined and explainable in terms of the individual properties and respective relations of those systems’ parts. We shall argue that this view is untenable in the case of proteins.

Microstructuralism implies the idea that the macro or global structural organisation of a system plays no role in the production of any of its macro-properties or, at least, that any global structural organisation, although not negligible, can be fully derived (metaphysically, logically, or nomologically) from the systems' microstructures alone. Either way, the fundamental idea underlying both views is that for a system to be subject to a successful microstructuralist analysis, each of its parts must instantiate the micro-properties and micro-relations it does *independently* of the specific macro-structure of relations of its system. Indeed, if reduction is a real *unilateral* or *asymmetric* relation between a reducing and a reduced term, microstructuralism can only be coherently defined as the view that all properties of a system are fully produced or determined by - thereby being fully explainable or reducible to - the properties and capacities that each part instantiates *independently* of being a part of that very system. Microstructuralism thus implies the notion that any system is micro-determined by the *intrinsic* or *system-independent* properties of its parts (Santos 2020).

This is why one can make sense of the idea that parts can carry on their fundamental properties and dispositions, despite some transformations in their global structural organizations. In this way, microstructuralism can assume that the properties and relations of a system's components either *predetermine in a necessitarian way* (in a programme-like fashion) all structural and functional properties that their system will instantiate in the future or, at least, that such parts already have in themselves, *as intrinsic dispositions or potentialities* (in a preformationist fashion), all properties that their system *may* or *may not* instantiate in the future (Santos 2015). In sum, in its diachronic version, microstructuralism can take the form of either a *necessitarian* or *potentialist predeterminism*.

What could the microstructure be in the protein case? Given what has been said above, if the protein's macrostructure is to be understood as the different global relational structures that a polypeptide exhibits in the course of its development, it seems clear that the protein's microstructure must be seen as the set of the *properties instantiated by the parts* constituting the successive polypeptide's macro-structures. The constituents of the polypeptide's primary structure are *the amino acid residues and the peptide bonds of the backbone connecting them*. Later on, this set of components will be enriched by other components, as we shall see in section 3.

In this sense, a diachronic microstructuralist view can thus argue that the *primary structure* determines, either in *necessitarian* or in *potentialist* terms, all the structural and functional properties that the developing polypeptide will instantiate. Indeed, since at least the early 60s, the set of the constituents of the protein's primary structure has been identified as the analogue of the microstructure, a reductionist view that has exerted a profound influence in biochemistry (Ben-Naim 2016) and that still permeates, as we shall see in section 4, several metaphysical accounts in the philosophy of biochemistry detached from contemporary biochemical practice. In the terminology of biochemistry, this diachronic microstructuralist view has been understood as if proteins' native structure and function are *determined* by the constituents of the polypeptide's primary structure, that is, by the amino acid residues connected by a backbone composed of peptide bonds.<sup>4</sup> In this article we shall argue that there are several good reasons, reviewed in section 3, to dismiss microstructure-based reductionist approaches to protein biosynthesis, native structure acquisition and maintenance. The analysis of the folding process we shall propose is mechanistic in ethos (Bechtel and Richardson 2010) and is aimed to show why microstructure-laden reductionist approaches are ontologically indefensible.

Let us now explain why in the 60s it was postulated that primary structure is ontologically primary. In a series of *in vitro* experiments, Anfinsen (1973) and colleagues showed that primary structure is the only necessary factor for refolding in the right conditions, i.e., the protein could be refolded *in vitro* in the absence of any additional molecular factors of the cellular environment apart from the primary structure itself. Extrapolating from the experiments, Anfinsen – who received the Nobel prize in 1972 – proposed the “thermodynamic hypothesis” (sometimes called Anfinsen's “dogma”), according to which “... *the three-dimensional structure of a native protein in its normal physiological milieu .... is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of the interatomic interactions and hence by the amino acid sequence, in a given environment*” (Anfinsen 1973, p. 223). Let us now take a closer look at this complex hypothesis.

---

<sup>4</sup> A polypeptide chain is a physical entity already with a three-dimensional structure. All molecules are three-dimensional structures in the same sense. To think otherwise is to confuse the linear organisation of the polypeptide-chain structure with its dimensionality, a confusion present in Bartol (2016) and Hüttemann and Love (2001, p. 539), when they refer to amino acids as lacking three-dimensional structure.

## 2. The dual nature of the thermodynamic hypothesis

What kind of hypothesis concerning the causal role of primary structure does the thermodynamic hypothesis represent? It is difficult to understand, but surely the causal hypothesis cannot be that primary structure is the only causal factor affecting folding. This causal exclusivity claim would be supported if Anfinsen's experiments showed, for instance, that renaturation or refolding to a functional native structure occurs without the help of other molecular (e.g., water, salts, chemical buffers) or biotic factors (e.g., enzymes).<sup>5</sup> It is clear that the thermodynamic hypothesis cannot be interpreted in this way (in analogy to the causal role of DNA molecules in development, see Vecchi 2019). We can anticipate two basic reasons for this (others will be illustrated in section 3):

a) First of all, Anfinsen restricts the validity of the hypothesis to a “*normal physiological milieu*”, i.e. a given environment where the "right conditions" are provided;<sup>6</sup> indeed, by varying above or below a certain threshold, environmental parameters affect folding; for instance, variations in pH might have such effects (Anfinsen 1973, p. 227);

b) Secondly, in a physiological context, a variety of molecular interactors might be involved in folding. In fact, even for very simple proteins like ribonuclease A, which are capable to refold *in vitro*, its autonomy to fold itself up without external aid within a cell is a myth; indeed, in their experiments, Anfinsen and colleagues discovered an enzyme (i.e., protein disulfide isomerase or PDI) that acts whenever incorrect disulphide bonds are formed (Anfinsen 1973, p. 224). More generally, it is today well known that protein folding – like many other biological processes – is a regulated process whose robustness depends exactly on the causal role played by a variety of molecular factors like chaperones and quality control systems (see section 3.1).

---

<sup>5</sup> It is important to note that the significance of Anfinsen's experiments is that they put an end to the old debate concerning the nature of the template supposedly governing folding. Anfinsen's answer was that there is no such template and that the protein only requires itself (i.e., its polypeptide chain) in order to fold into the native structure. However, this does not mean that proteins fold in a vacuum. That is why Anfinsen restricted his hypothesis to the "normal physiological milieu". Indeed, Anfinsen's experiments were carried out in aqueous environments with various abiotic factors.

<sup>6</sup> Needless to say, characterising the “normal” milieu and “right” environmental conditions generally is impossible.

What kind of causal hypothesis concerning primary structure is the thermodynamic hypothesis then? Anfinsen indicates that the most salient feature of the protein folding process is that generally a polypeptide chain develops along a specific pathway. As Anfinsen (1973, p. 228) argues: “A chain of 149 amino acid residues<sup>[SEP]</sup> with two rotatable bonds per residue<sup>[SEP]</sup>, each bond probably having two or three permissible or favored orientations, would be able to assume on the<sup>[SEP]</sup> order of  $4^{149}$  to  $9^{149}$  different conformations in solution.” But then it folds, in the normal physiological environment, extremely quickly into its typical native structure: why, among this enormous set of possible alternative conformations, folding is generally constrained along a specific pathway? And what are the relevant causal factors at play in such constraining process? The difficulty in interpreting the nature of the thermodynamic hypothesis is that it mixes two kinds of answer to the above questions. The first answer appeals to thermodynamic and energetics considerations: the native structure is the configuration with minimal free energy, which means that it is the most chemically stable configuration.<sup>7</sup> The second answer is the causal hypothesis according to which primary structure *determines* native structure. We shall call this latter *determination hypothesis*, a causal hypothesis concerning the causal role of folding factors, which are reduced to those characterisable in terms of primary structure. Thus, Anfinsen’s thermodynamic hypothesis mixes a principle of physico-chemical stability with the primary structure determination claim. But it is difficult to understand the connection between them. The former underwrites the hypothesis that the acquisition of the native structure within an energy landscape corresponds to reaching the minimal free energy conformation through some predetermined route (i.e., a specific folding pathway). This principle is tailored to make sense of the observation that folding happens quickly in a strongly constrained fashion, thus providing a solution to the computational explosion problem alluded by Anfinsen. But there remain questions both concerning the universal validity of this principle (e.g., is it an exceptionless or *ceteris paribus* law? Is it a deterministic or statistical law?), its seeming lack of a causal-mechanistic explanation and, above all, its relation to the determination hypothesis.<sup>8</sup> Supposedly, the connection between the principle of physico-

---

<sup>7</sup> The idea is that native structure is the energetically optimal folding solution, for instance representable as the unique deepest point in a folding funnel in the conformational energy landscape of the protein.

<sup>8</sup> The point is that, even if we knew all the secrets of folding in terms of thermodynamic and kinetic theories, we would still not know anything about the molecular, cellular and environmental causal

chemical stability and the determination hypothesis is that there exist some kind of privileged pathway guiding each protein's folding to the lowest free energy conformation whereby the privileged pathway and conformation must be “codified” and determined by the primary structure. From a mechanistic perspective, whether this is so cannot be assumed *a priori* just because it is underpinned by a principle of putative universal applicability. This causal determination hypothesis also underpins the promise of predicting the native structure and function of proteins from knowledge about primary structure (see also section 4). Indeed, this reductionistic view about protein structure has had an important heuristic value<sup>9</sup> in the development of biochemistry, structural biology and molecular biophysics in the second half of the 20<sup>th</sup> century. In fact, it opened an entirely new biochemical research programme, with many ramifications<sup>10</sup>, whose aim was to find the so called “protein folding code”, i.e., not only to predict native structure, but also to explain in mechanistic terms the pathways of protein folding for a given protein sequence. In this article our focus is on evaluating the determination hypothesis, particularly because it has strongly influenced recent contributions to the philosophy of biochemistry.

### 3. The determination hypothesis is a false and out-dated simplification

---

factors affecting the folding of each protein. Mechanistic analysis is causal in the latter sense, eschewing reference to subsumption under laws (as in deductive-nomological explanations) as much as possible.

<sup>9</sup> There are many reasons why primary structure acquired a central heuristic importance in protein science after Anfinsen experiments. For instance, primary structure is robust as it is not changed when tertiary structure changes; it is also manipulable by means of chemical modifications or through genetic mutation; it can be used in molecular evolution models etc. So, it is understandable that in several epistemic contexts primary structure is identified as “the” cause of certain phenomena (e.g., acquisition of tertiary structure) while other factors (e.g, environmental) have been relegated to the role of background conditions. This situation parallels the postulation, common in molecular biology, that genes are primary causes (Gannett 1999), even though genes should be more appropriately seen as ideal epistemic entry points for the analyses of biological phenomena. This means that the instrumental primacy of DNA-centric biology does not vindicate the ontological primacy of genes (Waters 2019). Analogously, the heuristic role of primary structure in protein science does not vindicate the ontological primacy of primary structure.

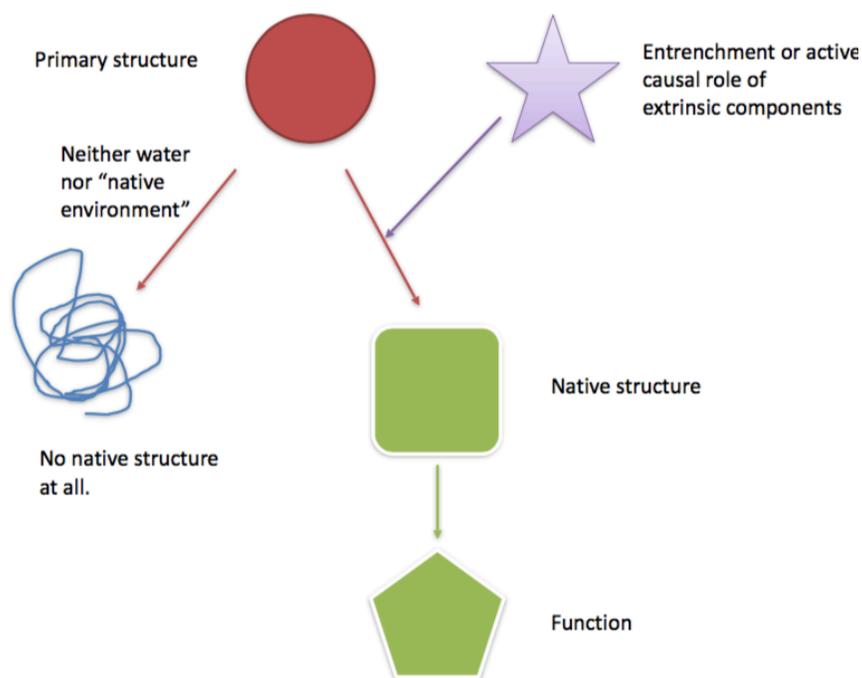
<sup>10</sup> Just to mention a few, Alan Fersht's protein folding pathways research programme (Fersht 2017), the “folding funnel” theoretical research program (Dill & Chan 1997), the “foldon” research program (Englander & Mayne 2017) etc.

The determination hypothesis suggests that primary structure determines native structure and function. Primary structure is characterisable in terms of the nature of the amino acid residue components of the polypeptide chain and the nature of the peptide bonds connecting them. The nature of the residues and peptide bonds will determine the specific spatial arrangements the polypeptide can realise. The realisable specific arrangements are accountable in terms of the properties of each residue (e.g., their physical and chemical nature) and of the properties of each peptide bond between the residues (e.g., their rotatability and chemical properties) of the polypeptide chain. Anfinsen's determination hypothesis is thus the view that the potentialities for the formation and maintenance of *the totality of the interatomic interactions* between the amino acids of the native structure are already present in the primary structure and that the causal role of the environment is just to activate these potentialities during the developmental process. Put differently, the question concerning the nature of the relevant causal factors constraining the folding process is answered by the determination hypothesis in intrinsicist, preformationist and micro-predeterminist term: all the polypeptide's realisable specific arrangements (e.g., the various transformations the polypeptide undergoes en route to reach native conformation as well as all its functional conformations) are predetermined in a *necessitarian* or *potentialist* way by the properties and relations of the primary structure's components. We shall now explain why the determination hypothesis is an out-dated simplification.

### **3.1 The causal role of extrinsic components in the development and maintenance of the native structure**

The insufficiency of primary structure for determination of native structure is particularly evident in cases of incorporation of "additives" (Mitchell and Gronenborn 2017) or environmentally available components. Often the polypeptide "entrenches" extrinsic components present in the environment in order to acquire and maintain native structure. These components are often necessary for such acquisition and maintenance. This process of entrenchment during folding and post-folding modifications is a variant of the more general process of "developmental entrenchment" (West- Eberhard 2003, pp. 500-503), which is basically the phenomenon whereby phenotypes do not "emanate" spontaneously from DNA taken by itself. The main upshot in the case of protein development is that the environment

plays a causal role. The cellular environment (even a test tube providing the “right” folding conditions) is inhabited by a variety of molecules of various sizes (e.g., water, ions, metabolites, proteins, nucleic acids, complex organic molecules, osmolytes, etc.) that are deployed during the developmental transformations the polypeptide undergoes during folding. Particularly significant in many proteins are prosthetic groups such as hemes and chlorophyll that are incorporated in a variety of proteins such as, respectively, haemoglobin and photosynthetic reaction centres. Even more ubiquitous is structural water assimilation, a process that is necessary in the folding of all proteins (Bellissent-Funel 2016). Moreover, extrinsic factors also play necessary causal roles in the folding process. One of the driving forces of folding is the hydrophobic effect, whereby the proteins "hide" hydrophobic groups due to an increase in aqueous solvent entropy. That is, without water no globular soluble protein (i.e., native structure) would fold appropriately, clearly demonstrating the constraining causal role of water. In both cases, prosthetic groups and water molecules are components alien to the primary structure that are necessary for the development and maintenance of the native structure.<sup>11</sup>



<sup>11</sup> These are just two examples of extrinsic factors playing an important role in protein folding and native structure maintenance. The presence of specific cations and anions, ionic strength, pH (protonation states), osmolytes, specific ligands, other polypeptide chains (in the case of obliged n-mers) etc. will also play a similar role to prosthetic groups and water molecules in the folding process.

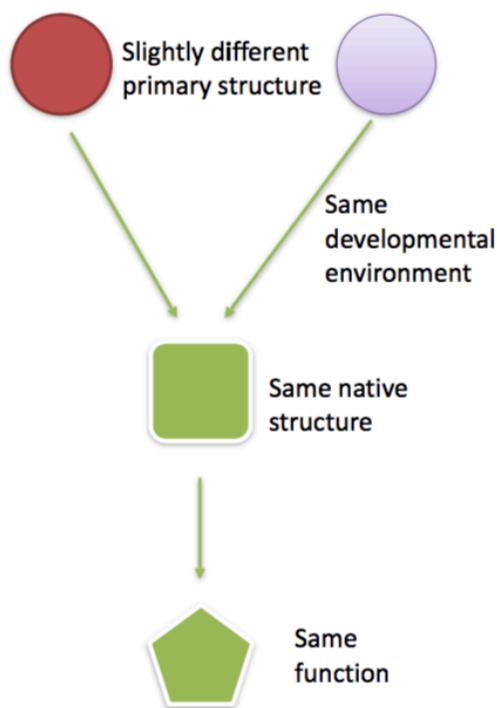
Another important example concerns the role of chaperones, which play an active causal role by generating correct microenvironments in order for proteins to reach their native structure and avoiding the formation of incorrectly folded states. As we saw in section 2, Anfinsen qualified the determination hypothesis by restricting the unique causal role of the primary structure in the context of the normal physiological environment. In this respect, additive and entrenched extrinsic factors, chaperones and quality control systems cannot be considered as abnormal elements of the physiological environment in any meaningful sense.<sup>12</sup>

### 3.2 Multiple realisation of structure and function

It is well known that many different primary structures can be folded into the same native structure. As an example, consider the human titin protein, the largest protein known, composed of more than 30.000 amino acids. Given this compositional complexity, extensive variability in amino acid composition – for instance due to transcriptional and translational “errors” (Drummond and Wilke 2009) - between token proteins even within the same organism ensues. However, many of these different token variants with different primary structures can nonetheless be considered and classified as the same kind of protein with the same function. Hence, the causal relation between primary structure and macrostructure is not deterministic in the sense that it is not the case that for every token primary structure a different native structure will result. This means that there must be additional causal factors beyond the primary structure constraining folding.

---

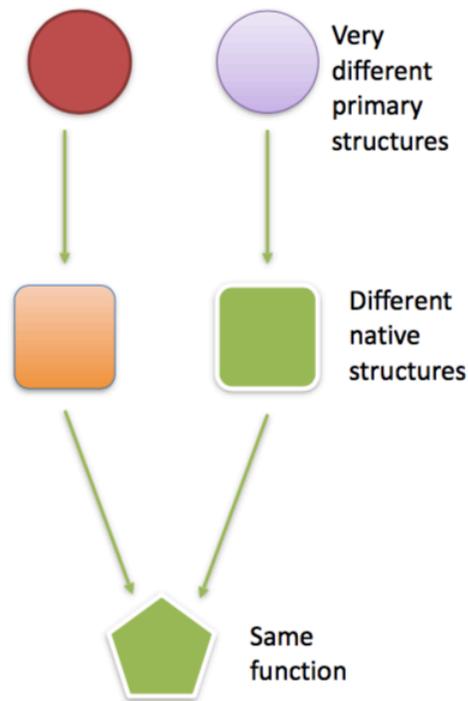
<sup>12</sup> Indeed, the same point applies also to refolding *in vitro*.



Indeed, as anticipated in section 2, Anfinsen (1973, p. 225) acknowledged that multiple realisation of native structure and function or, as he called this phenomenon, retention of “native structural *memory*” (otherwise called “robustness” of native structure in contemporary biochemical terminology), often occurs. Anfinsen (1973, p. 227) argued that “*Biological function appears to be more a correlate of macromolecular geometry than chemical detail.*” By “chemical detail” Anfinsen meant information about primary structure, i.e., about residues and peptide bonds components, their atomic properties and their location within the polypeptide chain; in brief, this is all the information necessary to capture the totality of interatomic interactions between the primary structure’s components, which should account for the formation of any kind of covalent and non-covalent bond between residues and for any transitional conformations the polypeptide assumes en route to reach native conformation. Anfinsen noted that some changes in amino acid composition (i.e., replacement, addition and deletion) are causally irrelevant in the sense of not causing structural instability and loss of functionality of proteins (Anfinsen 1973, p. 225-6). This robustness of the protein folding process – i.e., retention of “native structural *memory*” – given variation in primary structure should, thus, be explained (we return to this point in section 4). Part of the explanation is that some residues are more important than others: for example, the residues forming the hydrophobic core of the protein are fundamental for

architecture conservation while the active sites residues are fundamental for function conservation. Thus, not all amino acid changes have equal effect on native structure and function, with the majority producing no effect on them (Alexander et al. 2007) and a minority producing major changes, e.g., changes in what Anfinsen called macromolecular geometry and/or functional ones (Alexander et al. 2009).

Moreover, multiple realisation of function could be unrelated to folding robustness: there are many cases where neither primary nor native structures determine function. For instance, *E. Coli* possesses two different proteins performing phosphofructokinase activity, i.e., Pfk1 and Pfk2 (Babul, J. 1978). These proteins are not evolutionary related and are very different in terms of primary and native structure.

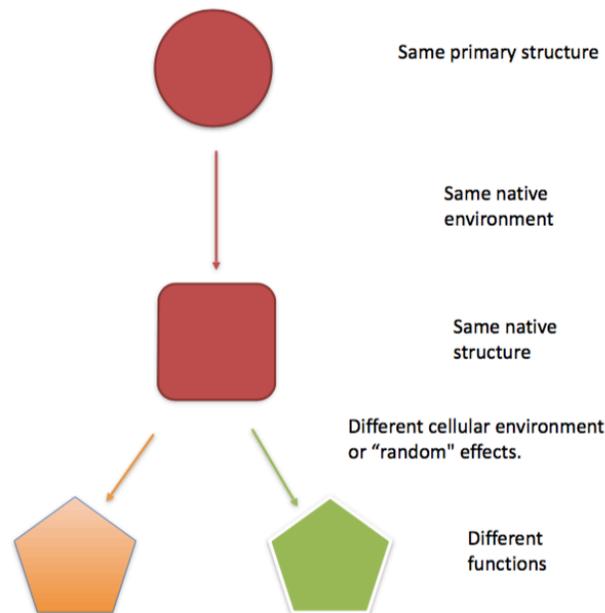


This means that function, diachronically speaking, is neither determined by the primary nor by the native structure, being rather relationally constructed by the cell or the organism according to its physiological needs within a particular organismal and environmental context.

### 3.3 Protein's plasticity

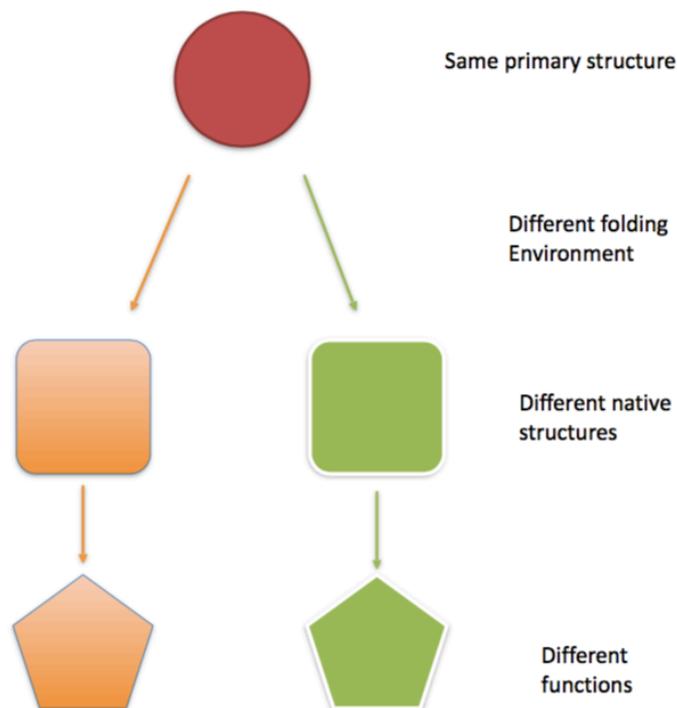
While multiple realisation refers to the fact that from different primary structures the same native structure and function can be realised (i.e., a many-one relationship), plasticity refers to the fact that the same primary structure can lead to different native structures and functions (i.e., a one-many relationship). In analogy to the biological literature centred on developmental and phenotypic plasticity (West-Eberhard 2003), the transformations undergone by a primary structure can produce plastic phenotypes, that is, different native structure variants and different functional variants. This is not surprising because, as already noted in section 1, folding and post-folding modifications are developmental processes. An interesting case concerns bifunctional proteins, whereby same primary structure develops into the same native structure but the protein, in the same environment, performs two different functions. In this case, the bifunctionality of the protein, for instance dependent on using two different active sites (or even the very same one), could be due to causal factors such as some form of physiological regulation (e.g. allosteric effect) or the mere contingent presence - and consequent enhanced probability of binding with the protein - of a particular substrate instead of another. For instance, the ADP-dependent sugar kinase of the methanogen *Methanococcus maripaludis* performs two different functions using the very same active site (Castro-Fernandez et al. 2014): the ADP-dependent phosphorylation of either glucose or Fructose-6-Phosphate, whereby the specific function is dependent on the contingent presence of the appropriate substrate in the environment. Other forms of plasticity, such as so-called moonlighting proteins, have already attracted considerable philosophical attention (Tobin 2010, Tahko 2019). These plastic proteins possess same primary structure and (sometimes) same native structure while exhibiting different functional roles in different environments (e.g., intra vs. extra-cellular environments). Unless theoretical considerations are framed in terms of Anfinsen's dogma, it would not be surprising if moonlighting happened often for the reason that proteins' catalytic and metabolic functions are distinct from, for instance, structural and signalling ones. What is most intriguing from an evolutionary perspective is that moonlighting functions are generally not conserved even between proximate species, illustrating how evolutionary novelty is highly contingent: for instance, while in ducks argininosuccinate lyase acts both as an enzyme in the urea cycle and as a structural protein, in chickens the two functions are performed by two different but homologous proteins (i.e., two proteins with different

primary structures). Often, if not always, moonlighting proteins act as subunits of larger protein complexes (Huberts & van der Klei 2010) and generally undergo post-folding modifications. Thus, the same protein could be chemically modified post-folding (for example, through phosphorylation) by an enzyme recognising a structural pattern that the polypeptide only acquires by developing into the native structure; a process of this kind again shows the insufficiency of the primary structure in determining plasticity in two senses: first, because the enzyme is recognising a property of the native structure that is lacked by the primary structure; secondly, because, given two proteins with same primary structure, only the one in the appropriate environment could be phosphorylated by the appropriate enzyme. Thus, the perfect determinative correlation between primary structure and function is invalid in all such cases.



Smith et al. (2013) use the term “proteoform” in order to refer to the set of different molecular forms a protein can assume, e.g., tokens with amino acid sequence differences, splice isoforms or post-folding variants. When post-folding modifications change native structure, the modified proteoform performs a different function from the original native structure proteoform. A different case would be when the same primary structure proteoform could fold into different native structures with different functions, as in the case of the transformer protein RfaH, which can acquire two different conformations (i.e., helix hairpin

and  $\beta$ -barrel). In this case, the difference-maker in the developmental process is an environmental input (e.g., a molecular signal whose nature, as far as we know, is so far mechanistically inaccessible, cf. Burmann et al. 2012). This phenomenon contravenes the determination hypothesis in an additional sense.



There thus exist several kinds of plasticity whereby the determination hypothesis is violated, whether it is articulated as a claim concerning determination of macrostructure or function.

#### **4. Limits of microstructure-laden reductionist approaches to protein development and maintenance**

Our main goal in this article, as explained in section 1, is to evaluate whether contemporary biochemical practice vindicates the idea that the primary structure's components *predetermine*, in a necessitarian or potentialist form, all the structural and functional properties that a protein instantiates in its native or post-folded structure. In

section 1 and 2 we emphasized that Anfinsen himself acknowledged that such determinism requires the existence of appropriate contextual conditions (i.e., the “normal physiological milieu”), thus dismissing a strict necessitarian form of predeterminism. We are therefore left with the aim of showing that a potentialist form of predeterminism is unfeasible. In section 3 we argued that the causal role of extrinsic components in the development and maintenance of the native structure, the multiple realisability of structure and function and protein’s plasticity undermine the determination hypothesis. We shall now argue that these phenomena pose a serious challenge to microstructuralist approaches to protein structure and function endorsing a predeterminist potentialism. This predeterminist potentialism is indeed the version of microstructuralism that some philosophers still defend. For instance, Goodwin (2011, p. 542-3) seems to argue that the primary structure’s constituents set or fix *all possible* structural and functional properties that the developing polypeptide will instantiate:

“Moonlighting of this sort shows that there is no one-to-one correlation between the tertiary structure of a protein and its biological function; the molecular environment of the protein can have a central role in determining what function(s) the protein performs. However, this does not undermine the fundamental role of primary structure in classifying and understanding the behavior of these proteins. Indeed, the capacity to perform any of their multiple possible functions is, for proteins of this sort, a consequence of their tertiary structure. The capacity to assume this tertiary structure (in the right biological circumstances) is, in turn, a consequence of the primary structure of the protein. So, though there are a lot of environmental and contextual factors to consider as well, *the diverse functions of these sorts of moonlighting proteins are ultimately understood to issue, as potentials or capacities, from their primary structure.*” [italics added]

According to this view, the causal role of cellular and organismal contexts and environmental conditions is only to *select* the potentialities that will be *actually instantiated* by the developing polypeptide, while the potentialities are fixed from the outset by the primary structure. After all we have been arguing in section 3, what could this mean? In order to add some metaphysical bite to the predeterminist potentialist view, let us consider Tahko’s explanation of moonlighting. After admitting that the environment plays a role in determining whether amphoteric substances will behave as acids or bases (thus dismissing a

necessitarian form of predeterminism), he ventures, in a rather speculative fashion, to compare amphoteric substances to moonlighting proteins:

“Yet, it would be odd to claim that the capacity of an amphoteric substance to react as either an acid or a base is something over and above its microstructural properties, a capacity that the substance gains only when the relevant environmental circumstances are in place. After all, both losing a proton (reacting as an acid) and gaining a proton (reacting as a base) are simple chemical reactions that are *determined by the relevant molecular structures* and we understand these reactions very well. There is no reason to think that we could not accurately capture acid-base interactions in terms of molecular structure, as *the capacity to act as an acid or a base is contained in that structure*. We arrive at the climax. Amino acids, which have a carboxylic acid group and an amino group (base) are also amphoteric. As we know, proteins are made up of amino acids and accordingly they can also react as amphoteric substances. *So, it turns out that at least some of the interesting capacities of proteins, some of their functional promiscuities, derive precisely from their amphoteric nature, which we have just explained in terms of molecular structure. This looks like a straightforward case of ontological reduction*” (Tahko 2019, p. 19 – italics added).

Thus, according to potentialist predeterminism, the causal role of molecular, cellular and organismal contexts and environmental conditions is only to *select* the potentialities that will *actually* be *instantiated* by the developing polypeptide. However, the potentialities are given from the outset by the primary structure. For instance, in the case of functionally promiscuous proteins such as moonlighting proteins, some moonlighting potentialities are fixed from the outset by the amphoteric capacities of the amino acids components of the primary structure. Predeterminist potentialists argue that, as there is no good reason to deny that the amphoteric behaviour is “determined” by the microstructure of the amphoteric substance - as its amphoteric capacities are “contained” in the microstructure -, so there is no good reason to deny that moonlighting behaviour is just the result of the predetermined amphoteric dispositions of the constituent amino acids.<sup>13</sup>

---

<sup>13</sup> Tahko’s explanation is speculative because the amphoteric properties of amino acids are rarely used in order to explain proteins’ behavior and have generally little relevance in the biochemistry of moonlighting proteins. Tahko overlooks more relevant aspects of protein structure such as the existence, for instance, of multiple active sites, that is, structural properties of the whole protein difficult to account in terms of the potentialities of the amino acidic residues. However, let us

At first sight, Hüttemann and Love (2011) make some progress towards the elaboration of a *non-predeterminist potentialism*. According to Hüttemann and Love (2011, p. 539), who merely focus on the structural transformations of the polypeptide (i.e., ignoring functional ones), the determination hypothesis states that the “... *properties of a linear polypeptide confer a disposition to fold into a three-dimensional protein; the disposition manifestation of folding is purely a function of the intrinsic properties (causal powers) of the linear polypeptide*”. If the determination hypothesis is framed in these terms, they argue, chaperones are significant in the evaluation of the determination hypothesis because:

“The intrinsic properties of the linear polypeptide arising from its amino acid residue parts are not sufficient to explain the protein-folding manifestation in the cell. *Chaperones do not merely trigger the disposition manifestation but contribute specifically to its manifestation*. The temporally extended process of folding not only requires appropriate environmental conditions but also the contribution of extrinsic chaperones; i.e. there is a failure with respect to the aspect of intrinsicity” (Hüttemann and Love 2011, p. 539 – italics added).

The problem of this argument is that it is easily interpretable in terms of the potentialist predeterminism Goodwin and Tahko endorse. Hüttemann and Love argue that chaperones *do not merely trigger* the disposition manifestation of the primary structure but *contribute specifically* to its manifestation. But what does “to contribute specifically to some disposition’s manifestation” mean? More specifically, does the failure of the *intrinsicity* thesis and the “specific contribution” role of chaperones concern the generation of certain dispositions, or only the *selective manifestation* of dispositions already possessed by the

---

pretend that the amphoteric character of amino acid residues is indeed relevant in order to explain a protein change of function in different environments. Even in this case, it is very difficult to make sense of the details of this hypothesis. Let us take the example, illustrated in section 3.3, of the protein argininosuccinate lyase that, in ducks, might act both as an enzyme in the urea cycle and as a structural protein in crystalline formation. In this case, we could make sense of Tahko’s hypothesis by framing it in these terms: if some amino acids act as acids, then argininosuccinate lyase will behave as an enzyme; conversely, if they act as bases, then argininosuccinate lyase will behave as a structural protein. Note that this hypothesis relies on a variety of additional assumptions: how many amino acids must behave in particular ways in order to elicit the appropriate behaviour in that specific environmental context? In which part of the protein are they located? More generally, is it amphoteric dispositions in the first place that trigger moonlighting behaviour or, rather, the fact that the protein possesses structural properties (e.g., multiple active sites) or that it is accidentally located in a specific context where it interacts with other enzymes or structural proteins, thus triggering new, previously inexistent, dispositions?

primary structure? The least we can say is that Hüttemann and Love are *not completely clear* about this issue. Still, on a closer look, they do seem to favour the latter view, since they explicitly say that chaperons contribute specifically to the “*manifestation*” of the disposition, and *not* to the very *coming into existence* or *possession* of that disposition.<sup>14</sup> Being so, Goodwin and Tahko could dismiss their *specific contribution argument* by arguing that chaperones *merely activate* a property that is already possessed intrinsically by the primary structure’s constituents. As an alternative to all these versions of potentialist predeterminism, we now propose an antithetical metaphysical view that we label “relational-construction-based potentialism”. After articulating this view metaphysically, we shall substantiate it by referring to the analysis provided in section 3.

According to relational-construction-based potentialism, *new potentialities* may come into existence by virtue of certain *compositional or structural changes*. We have already seen that polypeptides undergo a series of compositional and structural transformations through development. We add that any potentiality must be a potentiality of some *actual* structural property or complex of structural properties and that the actualisation or manifestation of any of these potentialities is, in turn, dependent on appropriate relations between each *actualised structure* and its respective environment. The point of disagreement is that, while potentialist predeterminists argue that the potentialities of each new structure instantiated by the polypeptide during development is predetermined by the potentialities of the micro-constituents of the primary structure, we argue that each new structure instantiated by the polypeptide during development might bring with it *new* potentialities not previously existing. The metaphysical rationale of relational-construction-based potentialism is: if any potentiality must necessarily be based on some concrete structure (otherwise, potentialities would be nothing but abstract or mysteriously free-floating entities), each new structure may bring with it new potentialities. One way in which they might come into existence is when

---

<sup>14</sup> Hüttemann and Love observe that there is an ongoing debate concerning the nature of the causal contribution of chaperones to folding. According to Ellis (1998) for instance, chaperones provide additional “steric information”, while Buchner & Walter (2005) deny this. The ethos of their narrative implies that Hüttemann and Love side with Ellis (1998), even though the reasons are unclear. We would say that, if the concept of steric information is characterised in terms of the information encoded in the sequence of the primary structure, then the generation of a *new* disposition – rather than the *selective manifestation* of a disposition already possessed by the primary structure - on the part of the developing polypeptide through interaction with a chaperone would show that steric information additional to that of the primary structure has been added. The analysis we hereby provide is not couched in informational terms.

the compositional and structural transformations undergone by the developing polypeptide forms a new structure that generates, in a particular cellular, organismal, and environmental context, a *new* potentiality, even though it might not be manifested; additionally, when a potentiality becomes actually *manifested* it may in turn create the conditions for the acquisition of further new potentialities depending on particular cellular, organismal, and environmental contexts. In this sense, the *intrinsicity* thesis fails not only for the manifestation of already possessed potentialities, but also for the generation of new potentialities. No predeterminism, in either necessitarian or potentialist terms, is thus defended. The domain of the future potentialities is not predetermined and fixed from the beginning. In sum, new structural organisations or systems of relations do not merely activate or trigger the manifestation of potentialities but are also either their very *generators*. In summary, we do not deny the crucial importance of the micro-constituents for the explanation of the development of any system. Rather, we criticise the predeterminist potentialism associated to microstructuralism insofar as it is committed to the view that the (structural and functional) nature of a developing polypeptide can be reduced to the set of its micro-constituents independently, first, of the fact that at least some of those micro-constituents are affected by the macro-structure of the relations of their systems and, secondly, of their specific global or macro-structural organisation.

Relational-construction-based potentialism can be supported by the analysis provided in section 3. Let us give some examples. The acquisition of new parts by the developing polypeptide, (i.e., a part that is extrinsic to the primary structure, see section 3.1) such as the entrenchment of prosthetic groups and structural water assimilation, is a compositional change that clearly produces new macro-structural organisations with additional components. Does this count as the generation of new potentialities? Or do entrenched components *merely activate* properties that are already possessed intrinsically and in potency by the primary structure's constituents? Consider a heme group. In order to incorporate hemes, the potentiality to functionally incorporate them is only attributable to the polypeptide chain after it has undergone specific structural modifications (i.e., when it has acquired some kind of tertiary structure). From the relational perspective we endorse, this is consistent with considering the new structural organisation or new system of relations constituted by the haemoglobin tertiary structure as generating a new potentiality on the part

of the developing polypeptide.<sup>15</sup> The same can be said about structural water assimilation. The only way in which water can serve a function in polypeptide development and native structure maintenance is if it interacts with an appropriate structure of relations that the developing polypeptide instantiates after it has already undergone some structural modifications, i.e., a structural organisation very different and not “contained” in the primary structure. Thus, from the relational perspective we endorse, this is consistent with considering the new structural organisation of relations of the secondary protein structure as generating a new potentiality on the part of the developing polypeptide. Many post-folding modifications (section 3.3) generating transformer, plastic and bifunctional proteins produce compositional and structural changes to the developing polypeptide that are not in any sense predetermined or “contained” in the primary structure. The transformer protein RfaH acquires the potentiality of being able to instantiate two different structural conformations (i.e., helix hairpin and  $\beta$ -barrell) not because these are predetermined or “contained” in the primary structure, but because the new structural organisation of relations of the native structure generates, by interacting with specific environmental inputs in specific cellular and organismal contexts, a new potentiality on the part of the developing polypeptide. Plastic proteins do not have their functions determined and contained in the primary structure but, rather, these are generated anew by interacting with, for instance, phosphorylating enzymes in specific and potentially new environmental conditions (e.g., unusual extra-cellular environments), where such enzymes recognise a structural property of the developing polypeptide that is not in any sense contained in the primary structure (it is indeed lacked by it). The bifunctionality of the same protein is neither potentiality determined nor contained in the primary structure, but it can only be meaningfully ascribed to the native structure that, by interacting with, for instance, contingently present substrates, generates a new potentiality on the part of the developing polypeptide. The crucial point is that all these capacities are only attributable to the polypeptide at various stages of development rather than to the primary structure. From the relational perspective we endorse, this is consistent with postulating the generation of new potentialities on the part of the developing polypeptide.

---

<sup>15</sup> This concept of novelty is developmental, not evolutionary. It just means that the tertiary structure is new vis-à-vis the primary structure. We shall come back to the evolutionary concept later on in the section.

There are three basic interrelated reasons for endorsing a relational-construction-based potentialism instead of potentialist microdeterminism.

The first important reason is that according to potentialist predeterminism, everything that a protein's can or cannot do is fixed, *ab initio*, by the set of the properties and causal capacities of the primary structure's constituents. To use again Tahko's explanation of moonlighting: it is the predetermined potentialities of the amino acids of the primary structure (i.e., their amphoteric properties and causal capacities) that cause moonlighting behaviour. The possibility that moonlighting is due to the way in which a folded protein interacts with cellular, organismal and environmental factors (e.g., the fact that the protein has multiple active sites or that it is accidentally located in a specific context where it interacts with other enzymes or structural proteins) is not seriously taken into consideration in its implications. But, as we already noted (note 6), even though it is understandable why many philosophers tend to believe that primary structure is ontologically primary, this ontological claim does not follow from the heuristic role primary structure plays in protein science. Indeed, seriously endorsing a predeterminist potentialist view would render inexplicable the lack of predictability of protein folding models based on primary structure information (Mitchell and Gronenborn 2017). Ultimately, it is much simpler to hypothesise, as the relational-construction-based potentialism we endorse makes clear, that this prediction failure is due to the fact that the often unpredictable cellular, organismal and, more generally, environmental context is causally relevant in polypeptide development in generating new potentialities on the part of the developing polypeptide.

Secondly, privileging causal explanations couched in terms of the intrinsic properties of the components of the primary structure as if they were isolated, self-contained or essentially immutable entities is alien to the practice of biochemistry and biology in a further sense not merely concerning prediction but having to do with the nature of the biochemical and biological explanation proposed, in particular in terms of the crucial causal role attributed to the cellular and organismal context. This causal role is not only relevant in order to make sense of the process of generation of new potentialities on the part of the developing polypeptide, but also to explain the generation of new organismal potentialities. In particular, phenomena such as structural robustness and multiple realisation of function

(section 3.2) generate new organismal potentialities.<sup>16</sup> The existence of structural robustness is a precondition for the development of the capacity of the cell or organism to tolerate phenotypically insignificant amino acid changes and nonetheless make use of the variant developing polypeptides according to its physiological needs. The existence of multiple realisation of function is a precondition for the development of the capacity of the cell or organism to co-opt variant developing polypeptides for the same function in order to fulfil its physiological needs within a particular environmental context. The potentialities to resist amino acid changes or to co-opt proteins for the same function are only ascribable to the cell or organism in a particular environmental context. Thus, from the perspective of relational-construction-based potentialism, to consider such organismal capacities as intrinsic to the organism is biologically odd. In order to avoid this oddity, the alternative is to argue that they are relationally constructed by the cell or the organism according to their physiological needs within a particular environmental context.<sup>17</sup>

The third reason why a relational-construction-based potentialism is preferable to a predeterminist potentialism is evolutionary. Relational-construction-based potentialism has the advantage of making sense of the concept of evolutionary innovation that, from a potentialist predeterminist perspective, is meaningless. Every time a developing polypeptide instantiates a potentiality that has never been available in evolutionary time, an evolutionary innovation might, if heritable, occur. The first time in evolutionary time a heme group was incorporated into a protein structure, the common ancestor of haemoglobin was born. The first time argininosuccinate lyase performed a structural role additional to the, by

---

<sup>16</sup> Structural robustness often depends on the conservation of specific amino acidic residues that are crucially important for folding and folded structure maintenance. Multiple realisation of function often depends on the existence of multiple active sites in proteins' surfaces and the possibility of undergoing conformational changes. Such properties are generated during the developmental process and are dependent on the macro-structure of the developing polypeptide rather than being fully reducible to the potentialities of the micro-constituents of the primary structure. For example, the conserved residues accounting for structural robustness are causally important because of the interactions they form within the native structure or in transient intermediaries in the folding process, but not because they determine the folded structure just for being in a certain location in the primary structure. Likewise, the existence of multiple active sites is a property of the fully folded protein. So, a relational-construction-based potentialist perspective is much more appropriate to account for these features.

<sup>17</sup> As Mitchell and Gronenborn (2017) succinctly put it "Function is always the result of a protein interacting with other components of the cell and its structure-in-isolation will not always depict its functional structure-in-context." The same considerations apply when folding and refolding are performed in vitro.

assumption, ancestral enzymatic one, a new function was born. Whenever a novel change in amino acid composition occurs and the quality control mechanisms of the cell fail to edit the “error”, a novel protein from a structural point of view is born; eventually, a protein with a novel function might be generated.<sup>18</sup> Predeterminist potentialism would have to argue that all such structural and functional evolutionary innovations are, as a matter of fact, preformed in the ancestral proteins from which all extant ones have evolved. These ancestral proteins already possessed all the set of potentialities that extant proteins manifest, where such manifestation is just triggered selectively by the relevant cellular, organismal and environmental contexts. Predeterminist potentialism is thus committed to the view that, in the course of evolution, no genuine novelty occurs.

An interesting analogy might succinctly illustrate our view. The Weismann-Roux mosaic model of development (see Amundson 2005) postulated that the fertilised embryo possesses all potentialities in a preformed state. During differentiation, individual cells actualise some of these potentialities through selective manifestation (while concomitantly losing other potentialities). Development is thus a process whose possible pathways and outcomes are prefixed or predetermined *ab initio*, in the sense that no new possibilities or potentialities can be generated. The fertilised embryo is, thus, analogous to the protein's primary structure. Developmental biology has abandoned mosaic models while relational-constructionist epigenetic accounts have prospered. In this article we have basically been articulating an epigenetic account of protein's development.

---

<sup>18</sup> A phenotypically silent structural change is equivalent, in the metaphysical terminology we use, to a new developmental potentiality. Suppose functional manifestation occurs after other phenotypically silent structural changes in amino acid composition are accumulated. The first structural change would act as a scaffold for other structural changes or, in the metaphysical terminology used, the first new potentiality is the condition of possibility for second-degree new potentialities. Evolutionary novelty might ensue if all these structural changes, initially phenotypically silent, in some specific environmental conditions produce a functional protein that has never existed before.

## Acknowledgments

Gil Santos acknowledges the financial support of FCT - Fundação para a Ciência e a Tecnologia (Stimulus of Scientific Employment, Individual Support 2017: CEECIND/03316/2017). Davide Vecchi acknowledges the financial support of the FCT - Fundação para a Ciência e a Tecnologia (DL57/2016/CP1479/CT0072). Gabriel Vallejos and Davide Vecchi acknowledge the financial support of the Fondo Nacional de Desarrollo Científico y Tecnológico de Chile (Grant N. 1171017). Gil Santos and Davide Vecchi acknowledge the financial support of the FCT - Fundação para a Ciência e a Tecnologia (Grant N. UIDB/00678/2020; R&D Project Grant PTDC/FER-HFC/30665/2017 “Emergence in the Natural Sciences: Towards a New Paradigm”).

## References

- Alexander, P.A. 2007. The design and characterization of two proteins with 88% sequence identity but different structure and function. *PNAS* 104(29):11963–11968
- Alexander, P.A. 2009. A minimal sequence code for switching protein structure and function. *PNAS* 106(50):21149–21154.
- Amundson, R. 2005. *The Changing role of the embryo in evolutionary thought*. Cambridge, Cambridge University Press.
- Anfinsen, C.B. 1973. Principles that Govern the Folding of Protein Chains. *Science* 181(4096): 223-230.
- Babul, J. 1978. Phosphofructokinases from *Escherichia coli*. Purification and characterization of the nonallosteric isozyme. *J Biol Chem* 253(12):4350-5.
- Bartol, J. 2016. Biochemical Kinds. *British Journal for the Philosophy of Science*, 67: 531–51.
- Bechtel, W. & Richardson, R.C. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge: MIT Press.
- Bellissent-Funel, M. et al. 2016. Water Determines the Structure and Dynamics of Proteins. *Chemical Reviews* 116:7673–7697. DOI: 10.1021/acs.chemrev.5b00664
- Ben-Naim, A. 2016. *Myths and Verities in Protein Folding Theories*. World Scientific.

Buchner, J. & Walter, S. 2005. Analysis of Chaperone Function in vitro. In J. Buchner and T. Kiefhaber (eds), Protein Folding Handbook. Part II (Volume 1), Weinheim: WILEY-VCH Verlag, pp. 162–96.

Burmann, B.M. et al. 2012. An  $\alpha$ -helix to  $\beta$ -barrel domain switch transforms the transcription factor RfaH into a translationfactor. *Cell* 150(2):291–303. doi:10.1016/j.cell.2012.05.042

Castro-Fernandez V, Bravo-Moraga F, Herrera-Morande A, Guixe V. 2014. Bifunctional ADP-dependent phosphofructokinase/ glucokinase activity in the order *Methanococcales* – biochemical characterization of the mesophilic enzyme from *Methanococcus maripaludis*. *FEBS Journal* 281: 2017–2029.

Dill, K.A. & Chan, H.S. 1997. From Levinthal to pathways to funnels. *Nature Structural Biology*. 4(1): 10–9.

Ellis, R. J. 1998. Steric Chaperones. *Trends in Biochemical Sciences* 23:43–5.

Englander, S.W. & Mayne, L. 2017. The case for defined protein folding pathways. *PNAS* 114(31):8253-8258.

Fersht, A. 2017. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. World Scientific.

Gannett, L. 1999. What's in a cause?: The pragmatic dimensions of genetic explanations. *Biology and Philosophy* 14(3):349-373.

Goodwin, W. 2011. Structure, Function, and Protein Taxonomy. *Biology and Philosophy*, 26: 533–45.

Havstad, J. C. 2018. Messy Chemical Kinds. *British Journal for the Philosophy of Science* 69:719–743.

Hüttemann, A. & Love, A.C. 2011. Aspects of Reductive Explanation in Biological Science: Intrinsicity, Fundamentality, and Temporality. *British Journal for the Philosophy of Science* 62:519:549.

Huberts, D.H.E.W. & van der Klei, I.J. 2010. Moonlighting proteins: An intriguing mode of multitasking. *Biochimica et Biophysica Acta* 1803:520-525.

Mitchell, S.D. & Gronenborn, A.G. 2017. After Fifty Years, Why are Protein X-ray Crystallographers Still in Business? *The British Journal for the Philosophy of Science* 68(3):703–723.

Santos, G. 2015. Ontological Emergence: How is That Possible? Towards a new Relational Ontology”, *Foundations of Science*, 20(4): 429-446.

Santos, G. 2020. Integrated-structure emergence and its mechanistic explanation. *Synthese*. doi: 10.1007/s11229-020-02594-3.

Slater, M. 2009. Macromolecular pluralism. *Philosophy of Science* 76(5):851–863. [SEP]

Smith, L. M., Kelleher, N. L., & Consortium for Top Down Proteomics. 2013. Proteoform: a single term describing protein complexity. *Nature Methods*, 10(3):186–187. doi:10.1038/nmeth.2369

Tahko, T.E. 2019. Where Do You Get Your Protein? Or: Biochemical Realization. *The British Journal for the Philosophy of Science*, axy044, <https://doi.org/10.1093/bjps/axy044>

Tobin, E. 2010. Microstructuralism and Macromolecules: The Case of Moonlighting Proteins. *Foundations of Chemistry*, 12, pp. 41–54.

Vecchi, D. 2019. DNA is not an ontologically distinctive developmental cause. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. <https://doi.org/10.1016/j.shpsc.2019.101245>

Waters, C.K. 2019. An Epistemology of Scientific Practices. *Philosophy of Science* 86: 585–611.

West-Eberhard, M. J. 2003. *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.