Macromolecular Pluralism (forthcoming in *Philosophy of Science*)

MATTHEW H. SLATER

Different chemical species are often cited as paradigm examples of structurally-delimited natural kinds. While classificatory monism may thus seem plausible for simple molecules, it looks less attractive for complex biological macromolecules. I focus on the case of proteins which are most plausibly individuated by their *functions*. Is there an objective count of proteins? I argue that the vagaries of function individuation infect protein classification. We should be pluralists about macromolecular classification.

1. INTRODUCTION

One often hears claims like "the typical human body possesses a hundred thousand proteins". This proposition, of course, refers not to the number of *individual molecules* in the typical human body, but to the number of distinct *kinds* of proteins typically found in the human body. But how should we count? How do we individuate these biological beasts of burden so as to make the claim come out (approximately) true?[1]

Following the lead of natural kinds essentialists of old, one might suggest individuating proteins (and other biological macromolecules) on the basis of their *chemical structure*. On first glance, this stance affords a tempting monism about biochemical taxonomy. But it is a costly and ad hoc monism. For there are several plausible contenders for the concept of a protein's structure, none of which accommodates biological practice (§2). While structure clearly plays a role in biochemical research, it often plays second fiddle to biochemical *function* when it comes to individuating different natural kinds of proteins (§4). However, counting proteins by functions introduces complications which strongly legislate for taxonomic pluralism (§5).

2. AN EMBARRASSMENT OF RICHES

Proteins, recall, are polymers of amino acids — "polypeptides" —, their order "coded" by DNA. They serve as enzymes (catalysts for biochemical reactions) and structural features of our cells and bodies; they help our hearts pump and allow our blood to carry oxygen. They achieve these and countless other feats in large part because of their complex and multilevel structures.

To acquire its characteristic biochemical dispositions, a protein must usually be "folded". From a protein's initial *primary structure* (its amino acid sequence), it usually acquires regular regions of *secondary structure* (e.g., α-helices and β-pleated sheets). Its *tertiary structure* consists in the subsequent three-dimensional folding of

[1] Of course, the phrase 'the typical human body' introduces (related) complications that far exceed the scope of this paper. But as nothing substantive for the issue of macromolecular classification turns on how or if we understand this notion, I leave it unanalyzed here — or we can always substitute the phrase '*my* body'.

these structures. *Quaternary structures* form when multiple subunits with tertiary structure become complexed together.

All this structure (at various levels of complexity) might seem a structure-obsessed monist's dream. It's really a nightmare. In the first place, the monist must apparently decide which of these various levels of structure is relevant to a protein's kind.[2] But no decision looks very promising. Primary structure — the simplest, most *stable* choice — apparently won't do. The "one-gene-one-protein" dogma once reigned, but is now rejected. The same amino acid sequence can be folded into a variety of different three-dimensional structures with different biochemical dispositions and functions.[3] Nor does a polypeptide's secondary structure determine these dispositions. Tertiary and quaternary structure — the level at which the majority of a protein's enzymatic prowess is realized — seem better candidates. For it is the particular three-dimensional folding of a polypeptide that confers an enzyme's reactive specificity. As the well-worn analogy has it, enzymes "recognize" their substrates rather like locks recognize their keys.

But how are we to understand tertiary structure? We might abstract away from the peptide bonds (the links between individual amino acids) and think of a protein's three-dimensional structure as simply the relative *location* of individual amino acids as if frozen in amber (in something like the manner we think of a crystalline structure). This approach, however, fails to accommodate the *movement* of a protein's components. The whole folded mess, remember, is held together primarily by hydrogen bonds and other weak intermolecular forces, perhaps at only a few points. Imagine a long hose fixed with magnets at certain points along its length which, when folded up, stuck together. Even if those magnets were strong, the structure of the folded hose might yet "flop around" a bit.

This analogy suggests an alternative model. Perhaps we should understand a protein's tertiary structure as a series of nodal "connections" (if weak) at different spots on the polypeptide chain. But this too underestimates the extent of a polypeptide's flexibility. For one, tertiary structure depends very sensitively on folding order. While many proteins fold more or less spontaneously into their active forms (as a low-energy conformation), many require a delicate dance of helper proteins (playfully referred to as 'chaperonins') to assist in folding. The same amino acid sequence is often differently spliced and folded in different biochemical milieus. For two, even for unchaperoned foldings, differences in chemical context often affect significant conformational differences. A protein's tertiary structure may be influenced by external forces. Certain amino acids have hydrophobic residues

[2] There is room here for a monist to press: perhaps her monism could be, as it were, "dappled" — employing some levels of structure in some circumstances and other levels in other circumstances.

[3] Though one hears of "optimization studies" in which three-dimensional structure of a polypeptide is surmised by a computer model, these must take into account chemical context and possible interaction with intermediate molecules — of which more presently.

and hence, in aqueous environments, tend together toward the center of the protein globule. They are "held together" — like antisocial members of an agoraphobic society — only by common aversion.

It might be objected at this point that molecular biologists *do* speak of a protein's structure — they invest large sums of money and many computing clock-cycles trying to *determine* their structure, after all. Molecular biology texts and journals include scores of *pictures* and three-dimensional representations of these structures. The devil may be in the details, as usual, regarding how specifically to understand a protein's structure — but what are these representations *of* if *not* protein structure?

While molecular biologists do indeed speak this way, it's rarely (to my knowledge) in the context of describing a protein's *essence*.[4] It is well to remember that the most effective technique for elucidating a protein's structure — x-ray crystallography — requires precipitating and *crystallizing* a quantity of the protein, stopping (or anyway, *significantly slowing*) its native motion. These "snapshots" no more capture the essence of a protein than a single satellite photo does of a hurricane. More importantly, a protein's motion — the normal structural change it undergoes — is often crucial to its catalytic activity. Proteins are structures in motion.[5] Emil Fischer's long-standing static "lock-and-key" model of enzyme-substrate interaction — where target molecules fit precisely like molecular keys into the larger rigid protein lock where the reaction is catalyzed — has been fairly comprehensively replaced by Koshland's (1958) "induced-fit" model. Fischer's model might have made sense of the target specificity of proteins, but not their enzymatic activity. As Matthews and van Holde put it, "a lock does nothing to its key" (1996, 368). Consider hexokinase, the first enzyme involved in glycolysis (the digestion of glucose). When a glucose molecule binds to this enzyme's active site, it initiates a significant conformational change in hexokinase which initiates the phosphorylation of glucose to glucose-6-phosphate (i.e., adds a phosphate ion to the sixth carbon of the glucose molecule), leveraging the energy stored in an ATP molecule. Some "protein machines" like the tetrameric sodium-potassium pump employ similar coupling reactions to power a repeated "airlock"-esque motion that pushes ions across cell membranes against their concentration gradients. Other proteins in turn harness the energy of these concentration gradients like turbines to catalyze further reactions. Such examples can be multiplied *ad nauseum*. They most naturally illustrate (though of course do not decisively *show*) that structure, in any standard sense of the term, cannot accommodate the *particular* diversity of proteins that biochemists ordinarily recognize. Counting by primary and secondary structure counts too few — by tertiary structure, too many. What are the alternatives?

[4] Even for simpler molecules, the familiar ball-and-stick models depict only their *average* (sometimes only *ideal*) bond-angles, ignoring significant motion their constituents undergo.

[5] I owe this neat turn of phrase to [reference suppressed].

3. Function's Attractions

We might be tempted at this stage throw up our hands and go anti-realist or radically pluralist: perhaps there is only the number of different kinds of proteins biologists *recognize* — not the *correct* number (cf. Kitcher 1984 on the cynic's definition of biological species). But anti-realism is strong medicine not obviously indicated. On the other hand, the committed monist might claim that, strictly speaking, as the individual molecules of giant protein jiggle, they instantiate continuum-many distinct kinds which we (understandably) group into higher-order kinds; and ditto (somehow) for the other examples. So long as we are only *ignoring* structural differences that really exist, the monist can confidently sit back and allow that groupings can be as arbitrarily motivated as you like.[6]

These are certainly bullets that either the radical monist and pluralist can bite. I cannot argue here that they are wrong to bite them (so it usually goes with bullet-biting). Perhaps actual biochemical practice is confused and needs revision; perhaps there's no such thing as biochemical practice! My presumption, however, is that a metaphysics of science ought to put a high priority on accommodating the bulk of scientific practice. If the result offends against some long held, pristine piece of philosophy, perhaps *we philosophers* ought to entertain revision.

Here's a candidate revision: abandon the dogma that individuating natural kinds of proteins requires cleaving to *a single* level of structure as relevant. A more opportunistic approach might thus adopt a "mixed stance", picking out primary sequences in some cases and (vague) tertiary structures in others. But one might fairly worry either that such decisions would be merely arbitrary, not reflecting anything about what natural kinds actually exist or that any "mixed" approach to what level of structure was *appropriate* for a given object would already presuppose some conception of what that object's kind was.

The appeal of this sort of "dappled structuralism" is best understood as an artifact of the importance of function for biochemical taxonomy. Suppose for the moment that we possessed a satisfyingly deep philosophical account of biochemical function which upheld ordinary scientific intuitions. Note how nicely it would solve the problems faced by the various structuralisms mentioned above. Individuating proteins by tertiary structure forecloses on grouping the various protein machines for which motion or conformational change crucially underpins their particular function. Biologists also recognize "dual-topology proteins", often observed in membranes, for which the direction of insertion and orientation of the protein subunits depend only weakly on their charge characteristics. It is hypothesized that the possibility of assuming multiple conformations within cells' plasma membranes may in fact be crucial to their characteristic function (see Rapp et al. 2006).

Individuating proteins by their primary structures, on the other hand, fails to recognize the wide variety of three-dimensional configurations (and different functions) proteins can achieve. The very same polypeptide

[6] I discuss the potential of this kind of monistic stance in my [author's essay].

sequence might get differently folded within a cell membrane to function in signal transduction and as a transcription regulator within the cell. How many kinds of proteins have we here? We hardly hesitate in answering 'two', counting by their distinct functions. Another cheer for functional individuation!

But as the glow of our above supposition subsides and we reflect on the thorny issues involved in providing an account of function, retreat to our dappled structuralism looks tempting. Perhaps pragmatics, informed by our rough-and-ready, un-philosophically-informed estimations of structure ought to dictate the relevant structural level. Even bracketing the worry that such pragmatism is simply evasion, I think this approach is also at odds with the way working biochemists classify. Take the case of alcohol dehydrogenase (ADH), which catalyzes the conversion of ethanol to the hangover-inducing acetaldehyde. Insofar as the catalysis involves motion, primary structure would appear to be the pragmatically-indicated choice: there's just one kind of protein here, not continuum-many very slightly different kinds of proteins. But even in this case, counting by primary-structure plausibly counts too many kinds of protein. For ADH comes in many "alleloforms" — very slightly different polypeptide sequences whose differences happen to be inconsequential to the protein's tertiary structure and biochemical function.[7] In most populations of the fruitfly *Drosophila melanogaster*, for example, ADH comes in two variants: called "slow" (ADH^S) and "fast" (ADH^F) (named after their speed through an electrophoretic gel). As Ridley puts it, "The *enzyme* called alcohol dehydrogenase is actually a class of two *polypeptides* with slightly *different amino acid sequences*" (2003, 83, *my emphasis*). Since normally such differences are inconsequential to the enzymatic activity of ADH, ADH^S and ADH^F are regarded as of a single kind of protein.[8]

4. Function's Perplexities

Let us turn at last to the question on everyone's mind: *If we individuate enzymes by function — and that's a big 'if'*[9] *— how do we individuate functions?!* I do not see an easy answer to this complex problem. In fact, it seems to

[7] There are various levels of variation: In *Drosophila* ADH, "the amino acid difference appears as a base difference in the DNA, but this was not the only source of variation at the DNA level. The DNA is even more variable than the protein study suggests. At the protein level, only the two main variants were found in the sample of 11 genes, but at the DNA level there were 11 different sequences with 43 different variable sites. . . . At the level of gross morphology, a *Drosophila* with two ADH^F genes is indistinguishable from one with two ADH^S genes; gel electrophoresis resolves two classes of fly; but at the DNA level, the two classes decompose into innumerable individual variants" (Ridley 2003, 84).

[8] Isn't this just the case of jadeite and nephrite? It is related — but there, function ascription played no role in answering the question of whether jade is natural kind of stuff. I have purposefully avoided entering into the debates about multiple realization here, fearing that it would take us too far afield and cause more confusion than it would solve. But clearly, this gambit deserves exploration.

[9] I have in mind an objection like this: *surely it's not an a priori necessity that a protein could have only one function! Could we not imagine a single kind of protein accomplishing two distinct functions in different circumstances?* I take this objection seriously. At this point, I'm not certain that I want to posit a one-to-one correspondence between biochemical functions and kinds of proteins. That being said, I believe this correspondence could be defended by pointing to an ambiguity in the phrase 'same kind of protein' between the functional kind and the different alleloforms that "realize" it. I return to this point in §5 below.

me that on any of our best accounts of function, there may *be* no unambiguous answer. Pluralism about macromolecular classification seems to follow.

Broadly speaking, two approaches to function have dominated the literature: the etiological or selected effects approach (championed by Millikan 1984, 1989; Neander 1991; Wright 1973) and the causal role or systemic capacity functions (developed by Amundson and Lauder 1994; Cummins 1975; Davies 2001). According to the selected effects approach, the function of a trait depends on whether performance of that function has figured in the selective history of that trait. Spelling out this intuitive thought has proved complex. How, specifically, must a particular capacity figure in a selective explanation? Must it explain the *increase* of the trait or just its *maintenance* — cannot traits that are *declining* in frequency nevertheless possess functions? And of course, settling these questions does not provide a way of gauging the contribution that a particular trait makes. Some traits are, in Gould and Vrba's (sometimes maligned 1982) term, *exaptations*: traits originally selected because of one kind of functionality but later co-opted for another. Even somehow restricting the contributions that trait has made to a lineage's "modern history" (see Godfrey-Smith 1994) leaves open the question of how important possession of that trait has been to its possessors' success.[10]

Difficulties like these — while probably not insuperable — underlie much of the attraction to the more liberal systemic capacity approach to function (upon which I shall focus). On this approach, we attempt to understand the function of a component in terms of its contribution to the capacities of the larger system. DNA polymerase, a key enzyme involved in DNA replication, has a "proofreading/error-correction" function that greatly reduces copying errors. High-fidelity DNA replication clearly contributes to the growth and development of organisms (not to mention their reproduction). It's not hard to appreciate how this systemic capacity likewise contributes to the fitness of organisms and serves as an explanation of it's own persistence in the lineage.[11]

But we can divide up systemic capacities in a variety of ways. When we ask *what is the function of DNA polymerase?*, a natural response is that it has *several functions*: DNA replication *and* error-correction. But this may be an artifact of the fact that we understood its role in replication before we understood its role in error-correction. Perhaps we ought to construe its function as *high-fidelity replication*. Which is correct? Is there a single function here or multiple functions? Is there a uniquely privileged way of dividing systemic capacity functions?

[10] Another issue that I will not pursue here is the complexity of construing the presence of a certain polypeptide as a *trait*. For if indeed the same polypeptide can carry out very different functions it may be impossible to separate the contributions of each to the organism's success.

[11] Davies (2001) has argued that the systemic capacity and selected effect theories are not distinct.

The monist has an obvious response to this kind of pointed question (inspired by the radical monism above): analyze down — divide reality as far as it can be divided. Those divisions are the natural joints. Simply ignoring these fine distinctions — in, say, grouping different alleloforms that carry out the same function or even grouping different functional kinds into a higher order kind — hardly impugns their reality. There *is* a privileged system of biochemical classification, even if it is not ours. In the case of DNA polymerase, while it is natural to include a variety of structures under the functional umbrella of *high-fidelity DNA replication*, biochemists can distinguish between various types of DNA polymerase on the basis of precisely *how* they carry out this function: "When [a mismatched DNA complex] was supplied to *E.* coli DNA polymerase I, the incorrectly hydrogen-bonded base was removed by a 3′ → 5′ exonuclease activity. . . . In DNA polymerase III, this function resides in the Θ subunit of the core polymerase" (Lodish et al. 1995, 387). The committed monist could once again argue that while strictly-speaking there are multiple functional biochemical kinds here, whether we group them as a single kind (owing to the sameness of a systemic capacity at a coarse level of analysis) or acknowledge the difference and divide them as different kinds is up to us.

This response may work well for some cases, but I do not think it will wash in general. For whether a proffered systemic capacity represents the "lowest level" of functional analysis will sometimes depend on how we divide into kinds other portions of the world — (crucially) *including* the very biochemical world which is at issue. I mentioned chaperonins above. These proteins fold other polypeptide sequences into their active tertiary structures. Now imagine two alleloforms of a certain protein which are shaped by the same chaperonin (either structurally duplicate molecules or the very same molecule): do these shapings count as two different functions? That depends, presumably, on whether the two alleloforms are in fact the same kind of protein — whether *they* have the same lowest-level biochemical functionality. But it might be that *their functionality* is to dismantle the very chaperonins which built them! Any appeal in this case to the objectivity of a lowest-level of functional analysis simply goes round in a circle.[12]

This suggests that there will sometimes be no uniquely privileged way of exhaustively carrying out a functional analysis on the systemic capacity approach. Insofar as the individuation of biochemical kinds depends on just this sort of functional analysis (I believe the story is similar on the etiological approach), there is no uniquely privileged way of dividing up biochemical kinds. We thus arrive at macromolecular pluralism. Notice that a corollary of this line of thinking is that whether two molecules with the very same structure (at whatever level of detail) might depend on facts extrinsic to them — for example, their biochemical context, the

[12] I argued similarly in [reference suppressed] that whether different enantiomers (mirror-image molecules) are classified as of the same kind admits of a similar kind of "circular volunteerism": to claim that, say, L-glucose differs in kind from D-glucose naturally appeals to a dispositional difference (reactivity with other chiral molecules, say) which in turn depends on whether enantiomers are different kinds of molecules (the question at issue).

variety of available substrates, and so on. For it is only relative to that context (or perhaps recent selective history) that a macromolecule *has* a particular function.

5. Conclusions

We've found that the monist's blanket acceptance of "*structure*" as individuating different natural kinds, even at the level of biochemistry, forecloses on much of the inferential/explanatory role to which we put many of those "structures". Monism about function ascriptions likewise appears forlorn — though admittedly, there is much more that would need to be said in defense of this view.

However that may be, pluralism makes good sense of the actual classificatory activities of biochemists. Our classification schemes typically reflect our best inductive and explanatory strategies — and these are rarely neat and orderly. Coarse taxonomies develop in contexts where fine functional analysis serves no apparent utility. Clearly, these may be revised if circumstances demand it (if we cannot make sense of a certain effect of introducing a particular protein). Biochemists can also avail themselves of a plurality of legitimate classificatory systems *simultaneously* by invoking what I'd like to call different "modes" of classification. The example of the alleloforms of ADH highlights an important feature of how we classify the world. Notice how Ridley described this case: the same kind of *enzyme* is actually a class of two kinds of *polypeptides*. The role of these organizing terms cannot be ignored. Polypeptides are individuated by their sequence; I conjecture that enzymes (and proteins more generally) are individuated, at least in large measure, by their function. We needn't maintain a one-to-one correspondence between different kinds of proteins and functions in order to appreciate this importance. If proteins are not *purely* functional kinds, though, the details of their (pluralistic) individuation seem to me an interesting and wide-open area for future research — one that would go a long way toward "de-radicalizing" the pluralism.

I believe the notion of different "modes" of dividing the world may help with this as well. While these different modes may cross-classify reality (or merely reveal different "layers" of structure), they need not be regarded as forcing hard choices. Perhaps the ambiguity of our nomenclature (as perhaps 'protein' is ambiguous between enzymatic function and the various levels of chemical structure) simply eases transitions between these modes, discourse in which is triggered by other univocal organizing terms. *Drosophila* ADH is a member of what Benach *et al.* call "a broad and heterogeneous family of alcohol dehydrogenases, named short-chain dehydrogenases/reductases (SDR), and is the only member that utilizes small alcohols as substrates" (2000, 3613). But though heterogeneous, SDRs possess certain structural commonalities; they describe one ADH's membership in this class thusly: "the subunit shows the typical fold of the short-chain dehydrogenase/reductases enzyme cores thus exhibiting a three-dimensional structure similar to that reported for other SDR proteins whose three-dimensional structure has been solved to date. The subunit of [*Drosophila*

ADH] has an α/β single domain structure with a characteristic Rossmann fold dinucleotide-binding motif…." (3614). Our rich system of different interleaved levels of functional and structural classification, I think, is best understood as the result of coming to terms with the genuine plurality of systems of functional and structural kinds. I suggest that we look to the role of organizing terms in our conceptualization of the world. There is no dividing the world into *the natural kinds*, perhaps — but that doesn't mean that there aren't natural kinds. There are natural kinds of enzymes, natural kinds of polypeptides, natural kinds of organisms, natural kinds of celestial objects, and so on — each dependent on norms of classification devolving from our best inductive and explanatory practices.[13]

REFERENCES

Amundson, Ron, and G. V. Lauder (1994), "Function without Purpose: The Uses of Causal Role Functions in Evolutionary Biology", *Biology and Philosophy* 9 (443–470).

Benach, Jordi, Sílvia Atrian, Joan Fibla, Roser Gonzàlez-Duarte, and Rudolf Ladenstein (2000), "Structure–function relationships in *Drosophila melanogaster* alcohol dehydrogenase allozymes ADH^{S}, ADH^{F} and ADH^{UF}, and distantly related forms", *European Journal of Biochemistry* 267:3613–3622.

Cummins, Robert (1975), "Functional Analysis", *Journal of Philosophy* 72:741–765.

Davies, Paul Sheldon (2001), *Norms of Nature*. Cambridge: MIT Press.

Ereshefsky, Marc (1992), "Eliminative Pluralism", *Philosophy of Science* 59:671–690.

Godfrey-Smith, Peter (1994), "A Modern History Theory of Functions", *Noûs* 28 (3):344–362.

Gould, Stephen Jay, and Elizabeth Vrba (1982), "Exaptation — a Missing Term in the Science of Form", *Paleobiology* 8:4–15.

Kitcher, Philip (1984), "Species", *Philosophy of Science* 51:308-333.

Koshland, D. E., Jr. (1958), "Application of a Theory of Enzyme Specificity to Protein Synthesis", *PNAS* 44 (2):98–104.

Lodish, Harvey, David Baltimore, Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira, and James Darnell (1995), *Molecular Cell Biology*. New York: Scientific American Books.

Mathews, Christopher K., and K. E. van Holde (1996), *Biochemistry*. Menlo Park: The Benjamin/Cummings Publishing Company.

Millikan, Ruth Garrett (1984), *Language, Thought, and Other Biological Categories*. Cambridge: MIT Press.

——— (1989), "In Defense of Proper Functions", *Philosophy of Science* 56 (2):288–302.

Neander, Karen (1991), "Functions as Selected Effects: The Conceptual Analysist's Defense", *Philosophy of Science* 58:168–184.

Rapp, Mikaela, Eric Granseth, Susanna Seppälä, and Gunnar von Heijne (2006), "Identification and Evolution of Dual-Topolgy Membrane Proteins", *Nature Structural & Molecular Biology* 13:112–116.

Ridley, Mark (2003), *Evolution*. 3rd ed. Boston: Blackwell Scientific.

Stanford, P. Kyle (1995), "For Pluralism and Against Realism About Species", *Philosophy of Science* 62:70–91.

Wright, Larry (1973), "Functions", *Philosophical Review* 82:139–168.

[13] Similar conclusions have been reached by Kitcher (1984), Ereshefsky (1992), and Stanford (1995) regarding species.