

# Coarsenings of the Genetic Code by the Twenty Registry Classes of a Chrysene Stack:

## A Partition Criterion, a Filter Selection, and the Double-Sided Encoding Hypothesis

Charles D. Schaper, Ph.D.

Chief Editor, *Annals of the Chrysene Formalism*

San Francisco, California, USA

editor@chrysene.com

### Abstract

The predecessor classifies the marker assignments of a chrysene stack of six sites into twenty frame classes and expressly declines to identify them with any twenty-element set, combinatorial or biological. We ask under what condition such an identification is possible, and what follows from one. Fixing a four-letter secondary alphabet with a distinguished partition into two classes of two, we show that the twenty classes admit a bijection to the twenty amino acids respecting the placement of codons by the standard genetic code precisely when that partition is amino from keto; for purine against pyrimidine the Hall deficiency is one and for Watson–Crick it is two, and in each case the obstruction is exhibited as a set of classes sharing too few neighbours, no amino acid being unmatched in every maximum matching. The result depends only on the partition and not on which of the eight admissible letter assignments is chosen, and the invariance is proved rather than observed. The admissible bijections number one hundred and twenty-eight; five of the twenty assignments are the same in all of them, one of the five being forced by a single class's own adjacency. Five physical filters select one bijection uniquely, by a margin of 0.17%, of which three of the five are individually decisive. The resulting table and the standard genetic code are related by two block operators whose composite reproduces all sixty-four codons.

What the construction does not establish is stated before it rather than after. The adjacency whose matchings are counted is built from the standard genetic code, so the table selected is a coarsening of that code onto the twenty classes and not a derivation of it: left unconstrained over all  $20!$  bijections the same filters prefer a map of higher total that violates the adjacency at sixteen of twenty classes. Neither block operator is determined by anything constructed here, and both are given by exhibition. Existence of a coarsening is not rare, holding for seventy per cent of codes carrying the standard degeneracy profile. The coincidence between the class count and the three-letter multiset count at  $N = 3$  is refuted on three grounds and used nowhere. Under a named hypothesis that the amino acid a cluster produces is attributable to the assembly rather than to the reading of it, the assignment factors through the twenty classes and code degeneracy becomes frame multiplicity; the same hypothesis then yields fibres of sizes two and four, which the standard code does not have, so the two-step relation is not generated by the frame. Five assignments are predicted independently of the filters, and one of them by a single measurement.

**Keywords:** Chrysene Formalism; registry classes; frame orbits; structured alphabet; Hall's theorem; perfect matching; permanent; genetic code; coarsening; amino/keto partition; non-implication.

## 1 Introduction

### 1.1 The problem

The predecessor [Sch26e] classifies the marker assignments of a chrysene stack of  $2N$  sites under the frame group, counts the classes as  $2^{2N-2} + 2^{N-1}$ , refines them by stabiliser, and

establishes that every quantity attributable to the assembly rather than to the reading of it is a function on the class set. At  $2N = 6$  there are twenty classes. That paper then does something unusual and, for the present purpose, decisive: it refutes the numerical coincidences its own count invites. Its Proposition 3.15 records that the agreement of twenty classes with twenty three-letter multisets holds at  $N = 3$  and at no other length and implies no

map; that its surjection of sixty-four words onto twenty classes has fibres of sizes two and four only, whereas the genetic code has fibres of sizes one through six together with three codons in no fibre, so that the two surjections are not the same map; and that no identification of the classes with any twenty-element set, combinatorial or biological, is asserted there.

Three questions are left open by that refutation, and they are not the questions the refutation answers. Whether an identification of the twenty classes with the twenty amino acids is *possible* is not settled by the observation that the two surjections differ, since an identification of classes with amino acids is a map of twenty-element sets and not a claim that two surjections of sixty-four codons agree. Whether such an identification is constrained by the structure of the secondary alphabet is not addressed at all, the predecessor assuming only that the alphabet carries some partition into two classes of two and never asking which. And what would follow from an identification, mathematically or as a physical claim, is not considered, the predecessor having no occasion to consider it.

## 1.2 The objective

To determine under which partition of a four-letter secondary alphabet the twenty classes admit a bijection to the twenty amino acids compatible with the standard genetic code, and to exhibit the obstruction where none exists; to count the compatible bijections and determine how much of such a bijection is fixed before any physical criterion is applied; to construct a physical criterion from properties of the reaction environment and determine whether it selects uniquely and how robustly; to compare the resulting table with the standard genetic code and state the comparison as an identity between exhibited objects rather than as a derivation; to state, under a named hypothesis about what a cluster produces, what the identification would mean and what it would predict; and, throughout, to state at the point of use which facts are evidence and which are not.

## 1.3 What is established

1. The partition criterion: the classes admit a bijection to the twenty amino acids compatible with the standard code precisely under the amino/keto partition, with Hall deficiencies one and two for the other two partitions and the obstructions exhibited; the result depends only on the partition, and the independence of the eight letter assignments within a partition is proved (Theorem 5.4, Proposition 5.5).
2. The letter assignment realizing it descends to the classification, and sixteen of the twenty-four bijections do not; it is fixed by class-level data only up to the exchange  $G \leftrightarrow U$ ,  $A \leftrightarrow C$  (Proposition 2.15).
3. The count: one hundred and twenty-eight compatible bijections (Proposition 5.8); five of the twenty assignments

are common to all of them, one forced by a single class's own adjacency and four by the matching condition (Proposition 5.9).

4. The selection: five filter scores, constructed from pocket geometry and residue properties, are Fr-invariant after passage to the class score, and select a unique bijection by a margin of 0.0244 in 14.42; three of the five are individually decisive and two are not (Proposition 4.10, Theorem 5.12, Proposition 5.13).
5. The identity: two block operators whose composite carries the selected table to the standard genetic code on all sixty-four codons (Theorem 6.7).
6. The consequence of the double-sided hypothesis: under frame-invariance the amino acid factors through the twenty classes, under injectivity it is a bijection, and code degeneracy is then the multiplicity of the class (Propositions 8.1, 8.3, 8.10); five assignments follow without any filter, and one by a single measurement (Theorem 8.7, Remark 8.8).
7. Six non-implications, each refuted by a named witness (Proposition 3.1).

## 1.4 What is not claimed

No axiom is introduced (Remark 2.2). Axioms A3, A4, A5, A7, A9, A10 and A11 are used nowhere (Proposition 2.5), so no statement below is a statement about sequence dependence and the order-dependence theorem of [Sch26d] is not invoked. No deferred target of [Sch26c] is discharged, and none of [Sch26d] or [Sch26e] (Remark 10.2); four are touched and the manner of the touching is recorded.

Three readings that the construction invites are refuted rather than used, and they are stated at the outset because the body of the paper would otherwise be taken to support them.

The first concerns the word *derivation*. The adjacency whose matchings are counted is defined by the standard genetic code (Definition 5.1), which therefore enters at the beginning of the construction and not at the end. The reduction from 20! candidates to one hundred and twenty-eight is done by that table; the filters supply the remaining factor. Maximising the same filters over all bijections, without the constraint, yields a map of higher total that disagrees at every one of the twenty classes and assigns tryptophan to a class whose only admissible value is alanine. The table of Section 5 is a coarsening of the standard code, and no section of this paper derives that code from the chrysene system.

The second concerns the two block operators. Neither is determined by the filters, and both are exhibited: the seven blocks of the first are seven of the eight four-fold family boxes of the target table, and the second, whose shape is intrinsic, selects its winners by no rule we have found, the best of five candidates succeeding in ten of thirteen halves and one block

being filled with an amino acid not present in it. The claim that the transformation takes exactly two steps is withdrawn for want of a move set.

The third concerns the number twenty, and is inherited. The coincidence with the three-letter multiset count holds at  $N = 3$  alone, implies no map, and rests on a reading of the multiset count that fails on three independent grounds (Section 9, Remark 3.3); the length  $N = 3$  is fixed here by the chemistry and by nothing else. And the agreement of the number of classes with the number of amino acids is a fact about two sets separately (Remark 3.4).

Finally, that any physical system realizes the selected table is  $H_{\text{dual}}$  with its identification clause, registered in Section 2 and used in no proof; the filters' physical reading is  $H_{\text{crys}}$ , likewise registered and confined to Sections 5 and 7. No synthesis is reported.

## 1.5 Logical position

Every numbered statement carries one of the four status labels of Definition 2.1. Every result using a hypothesis beyond A1–A12 names that hypothesis in its statement, and every hypothesis is registered in Section 2.5. Every molecular reading is confined to remarks marked [E] and is used in no proof. This paper falls within the framework scope condition of [Sch26b, §4.5] and is written to the orthogonal-constraint review bar of [Sch26b, §4.3].

Three features of the record are stated here rather than left to the appendix.

First, the standard genetic code changes its role between the predecessor and this paper. In [Sch26e] it is an external datum used only to refute, and that paper records that deleting the clause using it would weaken its warnings and no theorem. Here it defines the adjacency, and every statement of Sections 5 to 7 depends on it. The declaration is Remark 2.4 and the non-implication making it precise is Proposition 3.1(ii); the degeneracy profile was recomputed from the standard codon table rather than read from the citation, in the manner of [Sch26e, Rem. 3.16].

Second, the defect ledger of Appendix B is populated from the pre-drafting pass onward, in the crit/rev/ver form of [Sch26b, Def. 3], and therefore records criticisms raised against the plan of this paper and against its table file before any section was written, as well as those raised during drafting. Thirty items were disposed before drafting began and four during it; a criticism accepted before drafting is recorded as such and not retrospectively presented as a design decision. Of the thirty-four, a substantial majority are over-claims in the manuscript's own direction, which is the asymmetry [Sch26b, Rem. 1] reports and [Sch26e, Rem. B.5] reproduces, recorded here on a third manuscript.

Third, targeted verification for this manuscript re-runs the mathematical stage rather than employing an independent reviewing system. By [Sch26b, §9.4], independence of pass is not

independence of reviewing system, and the reader should discount the verification accordingly. The numerical verifications of Appendix B are of a different kind: each is an exhaustive enumeration or a closed computation over a finite set, reproducible from the definitions, and they are collected in a supplementary script which imports nothing from the tables of this paper, so that agreement is evidence and not a restatement.

## 1.6 Organization

Section 2 fixes conventions, status labels, the imported objects, the cluster and the naming discipline, the nucleobase alphabet and its partition, and the hypothesis register. Section 3 states the traps and non-implications, before the constructions they guard. Section 4 defines the five filter scores, proves that the per-codon score is not an assembly quantity and that the class score is, and records the provenance of its data. Section 5 defines the adjacency, proves the partition criterion and exhibits the obstructions, counts the coarsenings, identifies the five assignments fixed without any physical input, selects the rank-one coarsening and reports its sensitivity. Section 6 defines the two block operators and proves the identity on sixty-four codons. Section 7 treats the stop codons and records two withdrawals. Section 8 states the double-sided hypothesis in its clauses, extracts the filter-independent predictions, and gives the molecular reading. Section 9 records the arithmetic of the count coincidence. Section 10 restates the non-implications and inventories the Targets. Appendix A tabulates the twenty classes and the coding tables; Appendix B is the concordance, the domain items, the defect ledger and the verification record; Appendix C is the established literature.

# 2 Conventions, Logical Status, and the Hypothesis Register

## 2.1 Logical status

**Definition 2.1** (Status labels). Imported verbatim from [Sch26c, Defs. 1, 3, 5, 6] and used as in [Sch26a, Def. 2.7]: [A] axiomatic; [D] derived realization theorem; [T] target; [E] empirical instantiation claim.

**Remark 2.2** (No thirteenth axiom). [A] This paper introduces no axiom. Every structure below is either an object imported from [Sch26d] or [Sch26e], a definition internal to  $\{A1, \dots, A12\}$ , or a realization hypothesis in the sense of [Sch26c, Def. 3] registered in Section 2.5. The restriction is that of [Sch26d, Rem. 2.2] and [Sch26e, Rem. 2.2]: an additional axiom would reopen the non-vacuity theorem [Sch26c, Thm. 61] and the acyclicity of the axiom ladder

[Sch26c, Prop. 39], neither of which is re-established here.

**Remark 2.3** (Two counting conventions). Statements about a purely finite object — a set of binary strings, a  $20 \times 20$  zero-one matrix, a table of 64 entries — are proved without reference to any axiom and are labelled [A], in accordance with Definition 2.1, under which a statement following solely from the definitions needed to interpret the axioms is axiomatic. Each such statement is instantiated once, by a corollary labelled [D], as a statement about encoded stacks under the hypotheses named there. The two are kept apart throughout, in the manner of [Sch26d, Rem. 3.9] and [Sch26e, Rem. 2.3], so that a count of strings is never presented as a statement about a cluster.

**Remark 2.4** (The external datum, and the change in its role). The standard genetic code is an external datum. It is cited to [Cri68] for the code itself, and the codon assignment and degeneracy profile used below were recomputed from the standard codon table at the domain pass rather than read from the citation, in the manner of [Sch26e, Rem. 3.16]; the recomputation is recorded at Appendix B.

One difference from the predecessor must be declared here rather than left to the reader. In [Sch26e] the datum is used *only to refute*, and [Sch26e, Rem. 3.16] records that deleting the clause that uses it would weaken that paper's warnings and no theorem. In this paper the datum is used *constructively*: it defines the adjacency of Section 5, and every statement of Sections 5–7 depends on it. Accordingly no statement of this paper may be read as independent of the standard genetic code, and the non-implication making that precise is Proposition 3.1(ii). Statements proved against the fixed finite table are labelled [A] by Remark 2.3; the table is not itself a claim about the chrysene system and is not of type [E].

## 2.2 Imported objects

Throughout,  $\mathfrak{C}$  denotes the molecular reduct of A1 and  $(X, j_{\text{str}}, \mathcal{E}_0, w, \{C_j\}, \{\tau_j\})$  the admissibility datum of A12, with admissible domain  $\Omega_{\text{adm}}$ . The species alphabet is  $\Sigma_{\text{sp}} = \{\mathbf{s}, \mathbf{c}\}$  of [Sch26d, Def. 2.3], with  $\mathbf{s}$  carrying  $A_{\mathfrak{C}}^{\text{pri}} = \{v_3, v_6, v_9, v_{12}\}$  and  $\mathbf{c}$  carrying  $A_{\mathfrak{C}}^{\text{dual}} = \{v_2, v_5, v_8, v_{11}\}$ , and  $e_z$  is the stacking direction of [Sch26d, Def. 2.5].

The following are imported unchanged and used below without restatement.

| Object                                                                                            | Source                                   |
|---------------------------------------------------------------------------------------------------|------------------------------------------|
| marker $\text{mk}$ , values $F = 0, B = 1$                                                        | [Sch26d, Def. 3.3]                       |
| vessel; the two binaries $b_1, b_2$ ; alphabet $\Sigma_{\mathfrak{p}}$                            | [Sch26d, Defs. 3.10, 3.12, 3.13]         |
| registry $\text{reg} = b_1$ , classes $\{F\bar{F}, BB\}, \{FB, BF\}$                              | [Sch26d, Def. 3.15, Prop. 3.16]          |
| structural group $G = (\text{id}, \tau, \rho, \tau\rho)$ ; $c = \rho$                             | [Sch26d, Props. 4.7, 4.10]               |
| frame $(k_0, \varepsilon, f)$ ; frame-setters $\Phi = \Phi_{\text{dir}} \times \Phi_{\text{typ}}$ | [Sch26d, Defs. 6.1, 6.2]                 |
| frame group $\text{Fr} = (\text{flip}, \text{rev})$                                               | [Sch26d, Def. 6.13], [Sch26e, Def. 2.13] |
| registry sequence $D(m)$ , class $\bar{D}(m)$ , central entry $\bar{D}_N$                         | [Sch26e, Def. 2.11]                      |
| the class set $X$ , $ X  = 2^{2N-2} + 2^{N-1}$ , twenty at $2N = 6$                               | [Sch26e, Lem. 3.1, Thm. 3.3]             |
| orbit sizes two and four, profile eight and twelve                                                | [Sch26e, Prop. 3.6, Cor. 3.7]            |
| the twenty classes in tabulated form                                                              | [Sch26e, App. A, Tab. 2]                 |
| descent theorem: five equivalent transport conditions                                             | [Sch26e, Thm. 4.9]                       |
| eight of twenty-four bijections descend                                                           | [Sch26e, Prop. 4.14]                     |
| the four induced maps of class sets; the residual pair                                            | [Sch26e, Cor. 5.4, Thm. 5.8]             |
| the ensemble functional $\Phi$ and its genericity                                                 | [Sch26e, Def. 6.6, Thm. 6.12]            |

**Proposition 2.5** (Provenance). [A] Each constituent of the constructions of this paper is supplied by a named axiom, as follows: the site set and the translation action, together with the stacking direction, by A6; the two attachment sets, hence the species alternation, by A1; the non-collapse of the marker by A2; the oriented external perimeter cycle read by the marker, by A8; and the constraint functionals and the admissible domain by A12. No constituent is supplied by A3, A4, A5, A7, A9, A10 or A11.

*Proof.* By inspection of the cited definitions against the statements of the axioms, exactly as in [Sch26e, Prop. 2.7], from which the present list does not differ. No statement of this paper refers to the word operator, to the sequence algebra or to any translation unitary, so the negative assertion holds for A7 as for the remaining six.  $\square$

**Remark 2.6** (Consequence of Proposition 2.5). This paper, like its predecessor, is a combinatorial construction over the alphabet and the site set. The sequence dependence of [Sch26d, Thm. 4.17] and the non-implication of its Proposition 4.18 are not used, and no conclusion below may be read as a statement about the order dependence of an encoded word. This is worth restating here because Sections 5–7 concern the genetic code, whose codons are ordered triples; the order in question is the order of the vessels within a cluster, supplied by the frame, and not the noncommutativity of  $\mathfrak{A}_{\mathfrak{C}}$ .

## 2.3 Clusters, and the naming discipline

**Definition 2.7** (Cluster). A *cluster* is a stack of  $2N = 6$  sites in the sense of [Sch26d, Def. 3.1], read at the offset  $o = 0$  of [Sch26e, Def. 2.9], so that it carries  $N = 3$  vessels. Its marker assignment is an element  $m \in \{0, 1\}^6$  and its class is  $\bar{D}(m) \in X$ , of which there are twenty by [Sch26e, Thm. 3.3(ii)].

**Remark 2.8** (Naming discipline). [A] The term *vessel* is reserved, as in [Sch26d, Def. 3.10], for an adjacent pair

of sites carrying one unit of each species; the six-site object is the cluster of Definition 2.7 and is never called a vessel. Drafts circulated before Pass 0 used “reaction vessel” for the six-molecule object, which collides with [Sch26d, Def. 3.10] and with the molecular reading of its Remark 3.11; the correction is recorded in the defect ledger. This paper likewise adds no fourth sense of *registry*: the three senses kept apart at [Sch26d, Rem. 2.20] and [Sch26e, Rem. 2.36] — the vessel registry  $\text{reg}$ , the  $z$ -registry  $\eta$  of A6, and the frame-setter coordinate  $\Phi_{\text{typ}}$  — are the only ones, and the registry sequence of [Sch26e, Def. 2.11] is the sequence of values of the first on adjacent pairs.

**Remark 2.9** (Molecular reading). [E] In the intended reading a cluster is six stacked substituted chrysene units alternating in species, and its three vessels are the three catalyst–substrate interfaces. Each interface produces one nucleobase, so a cluster produces one codon; this is why  $N = 3$  is the length treated, and no further justification of that length is offered or required. This remark enters no proof.

## 2.4 The nucleobase alphabet and the letter assignment

**Definition 2.10** (The secondary alphabet).  $\Sigma'_v := \{A, C, G, U\}$ , the four RNA nucleobases, with  $|\Sigma'_v| = 4$ ; this is the secondary alphabet of [Sch26d, Def. 8.1] in the instance treated here.

**Hypothesis 2.11** ( $H_{\text{base}}$ , the structured nucleobase alphabet).  $\Sigma'_v$  carries, from its own structure and not by transport from  $\Sigma_v$ , a distinguished partition into two two-element classes, namely  $\{A, C\}$  and  $\{G, U\}$ .

**Remark 2.12** (The partition, and its name). [E] The partition of Hypothesis 2.11 is the amino/keto dichotomy of the exocyclic substituent presented at the Watson–Crick face: adenine and cytosine carry an exocyclic amino group, guanine and uracil a keto group. These are the IUPAC ambiguity classes  $M$  and  $K$ . The three partitions of a four-letter nucleobase alphabet into two classes of two are exactly the three IUPAC binary classes:  $W|S = \{A, U\}|\{C, G\}$ , the Watson–Crick classes;  $R|Y = \{A, G\}|\{C, U\}$ , purine and pyrimidine; and  $M|K = \{A, C\}|\{G, U\}$ , amino and keto. Drafts before Pass 0 called the last of these “the wobble pairing”; that name is withdrawn, wobble denoting  $G\cdot U$  and inosine pairing and  $A\cdot C$  being a rare protonated mismatch. The chemical reading enters no proof; what the proofs use

is the partition alone.

**Remark 2.13** ( $H_{\text{base}}$  is the weakest form of the datum). Hypothesis 2.11 is  $H_{\text{sec}}$  of [Sch26e, Hyp. 2.23] instantiated at  $\Sigma'_v$ , together with the designation of which partition is meant. By [Sch26e, Rem. 2.24] the group  $G'$ , the invariant  $\text{reg}'$  and the fixed-point-free element  $c'$  are recovered from the partition alone and require nothing further, and the partition is the weakest of the four equivalent data. Nothing is assumed about base pairing, about hydrogen bonding, or about any biological role.

**Definition 2.14** (The letter assignment).  $L : \Sigma_v \rightarrow \Sigma'_v$  is the bijection

$$L(FF) = G, \quad L(FB) = A, \quad L(BF) = C, \quad L(BB) = U.$$

The *codon* of a cluster with word  $(\sigma_1, \sigma_2, \sigma_3)$  is  $L(\sigma_1)L(\sigma_2)L(\sigma_3)$ .

**Proposition 2.15** ( $L$  descends). [D] Assume  $H_{\text{bij}}$ ,  $H_{\text{sec}}$  and  $H_{\text{base}}$ . Then

- (i)  $L$  carries  $\text{reg}^{-1}(0) = \{FF, BB\}$  onto the keto class  $\{G, U\}$  and  $\text{reg}^{-1}(1) = \{FB, BF\}$  onto the amino class  $\{A, C\}$ ; hence  $L$  satisfies the equivalent conditions of [Sch26e, Thm. 4.9] and descends to the class sets.
- (ii) Of the twenty-four bijections  $\Sigma_v \rightarrow \Sigma'_v$ , exactly eight descend, and the remaining sixteen each carry some class of  $X$  onto a set that is not a class.
- (iii)  $L^{\times 3}$  carries the eight classes of size two onto eight classes of size two and the twelve of size four onto twelve of size four; it does not in general preserve stabiliser type.
- (iv)  $L$  is determined by class-level data only up to the pair  $\{L, c' \circ L\}$ , whose two members agree at no letter and differ by the exchange  $G \leftrightarrow U, A \leftrightarrow C$ .

*Proof.* (i) By Definition 2.14,  $L(\{FF, BB\}) = \{G, U\}$  and  $L(\{FB, BF\}) = \{A, C\}$ , which are the two classes of the partition supplied by Hypothesis 2.11; this is clause (i) of [Sch26e, Thm. 4.9] after the labelling of the two values of  $\text{reg}'$ , and the equivalence with clauses (ii)–(v) is that theorem. (ii) is [Sch26e, Prop. 4.14], the admissible bijections being exactly those carrying the partition to the partition, which number  $|\mathcal{N}| = 8$ . (iii) is the final assertion of [Sch26e, Thm. 4.9] together with [Sch26e, Rem. 4.13]. (iv) is [Sch26e,

Prop. 5.3(iii)] and [Sch26e, Thm. 5.8(i)]: the fibre of  $\Psi$  through  $L$  is  $\{L, c' \circ L\}$ , and  $c'$  is the unique fixed-point-free element of  $G'$ , which by Remark 2.13 is the involution exchanging the two members of each class of the partition, that is  $G \leftrightarrow U$  and  $A \leftrightarrow C$ .  $\square$

**Remark 2.16** (What Proposition 2.15 does not do). Three things are not established.  $L$  is not constructed from the structural data of [Sch26d, §§3–6]; it is exhibited, and its provenance is the empirical remark below. Constructing it is [Sch26d, Target 10.5] and is not discharged. And by clause (iv) the choice between  $L$  and  $c' \circ L$  is a designated zero for  $\Sigma'_0$ , which is exactly [Sch26e, Target 7.7]; exhibiting one member of the pair on empirical grounds does not discharge a target that asks for the datum from the structure of the secondary realization. Both targets are carried unchanged into Section 10.

**Remark 2.17** (Provenance of the assignment). [E] In the intended reading, a  $\pi$ -electron stack of alternating 2, 5, 8, 11-tetrasubstituted (catalyst) and 3, 6, 9, 12-tetrasubstituted (substrate) chrysene units bearing phosphoethoxy substituents synthesises one of the four nucleobases at each catalyst–substrate interface. The base produced is determined by the ordered pair of markers of the interface, read against the projection reference supplied by the initiator of [Sch26d, Def. 9.1], and the assignment so read is  $L$ . The composition of matter is claimed in US 12,195,423 and the corresponding system in US 12,473,247, which also describe the encoding, its integration with phospholipids, and the substrate role in amino-acid synthesis. The clause “read against the projection reference” is not decorative. By [Sch26d, Prop. 4.4] the relative orientation of the two units fixes  $b_1$  alone, of two values; the second coordinate, and with it the identity of the base within its class, is supplied by the frame under  $H_{\text{frm}}$ , and by [Sch26d, Cor. 6.9] the designated zero is aligned. A reading that omitted the clause would assert a four-way determination from a two-valued datum. This remark enters no proof, and no theorem below becomes false, or loses a hypothesis, if it fails.

## 2.5 The hypothesis register

Each hypothesis below is a realization hypothesis in the sense of [Sch26e, Def. 3]. No statement of this paper uses an unnamed hypothesis. The first six are imported; the remainder are introduced here.

**Hypothesis 2.18** ( $H_{\text{mk}}$ , marker realization). As [Sch26d, Hyp. 2.6].

**Hypothesis 2.19** ( $H_{\text{stack}}$ ,  $z$ -constraint). As [Sch26d, Hyp. 2.7].

**Hypothesis 2.20** ( $H_{\text{frm}}$ , external frame datum). As [Sch26d, Hyp. 2.8]: a structure external to the stack, not determined by it, supplies the origin, the direction sign  $\varepsilon$  and the projection reference  $f$ .

**Hypothesis 2.21** ( $H_{\text{ind}}$ , vessel independence). As [Sch26d, Hyp. 2.11]; the consequence used here is [Sch26d, Prop. 3.19], that the set of words is all of  $\Sigma_0^3$ , hence that all sixty-four codons occur.

**Hypothesis 2.22** ( $H_{\text{bij}}$ , bijectivity). As [Sch26d, Hyp. 2.12].

**Hypothesis 2.23** ( $H_{\text{sec}}$ , structured secondary alphabet). As [Sch26e, Hyp. 2.23], instantiated at  $\Sigma'_0$  by Hypothesis 2.11.

**Hypothesis 2.24** ( $H_{\text{stop}}$ , stop designation). Three codons, UAA, UAG and UGA, lying respectively in the classes numbered 18, 4 and 15 of Table 1, are non-coding.

**Remark 2.25** (Why  $H_{\text{stop}}$  must be named). The designation is supplied by no axiom and by neither predecessor. By [Sch26e, Prop. 3.15(ii)] the surjection of the sixty-four words onto the twenty classes has no element outside a fibre, whereas the standard genetic code has three; removing three codons from the coding set is therefore added structure and not a consequence of the classification.  $H_{\text{stop}}$  is used in Sections 5 and 7 and in no other place.

**Hypothesis 2.26** ( $H_{\text{crys}}$ , filter provenance). The five filter scores of Section 4 reflect thermodynamic preferences of the cluster.

**Remark 2.27** (Role of  $H_{\text{crys}}$ , and its position in the predecessor’s scheme).  $H_{\text{crys}}$  is used in Sections 5 and 7 and nowhere else; the combinatorial statements of Section 5 — the adjacency, its permanent, the existence of a maximiser — are proved without it, and none becomes false if it fails. Its position in the scheme of [Sch26e, §6] should be stated explicitly. By [Sch26e, Prop. 6.1] no quantity invariant under relabelling of the target selects anything, so a selection criterion must be built from a weight attached to the target’s own labels. The five filters are such a weight, and

are therefore an instance of  $H_{\text{fit}}$  of [Sch26e, Hyp. 2.29]; the per-class score of Section 4 is a function on classes, as that hypothesis requires, because it is taken as a maximum over the codons of a class. Selecting the rank-one map by maximising the total is an instance of  $H_{\text{var}}$  of [Sch26e, Hyp. 2.32]. By [Sch26e, Prop. 7.1(viii)] uniqueness of a maximiser is not selection by any process, and no such selection is asserted here.

**Hypothesis 2.28** ( $H_{\text{dual}}$ , double-sided encoding). In three clauses.

- (i) The amino acid produced by a cluster is a function of the assembly and not of the reading of it, that is, it is Fr-invariant.
- (ii) The function of clause (i) is injective on  $X$ .
- (iii) The marker state of a cluster that produces a given codon is the state that produces the cognate amino acid.

**Remark 2.29** (The clauses are not interchangeable). Clause (i) alone has mathematical force: by [Sch26e, Thm. 3.3(iii)] an Fr-invariant function of the marker assignment factors uniquely through the twenty classes, so under clause (i) the amino acid is a function on  $X$  and the degeneracy of the assignment is the orbit structure. Clause (ii) makes that function a bijection and is the clause the counting of Section 5 uses. Clause (iii) is the double-sided claim proper and is empirical throughout. The cluster-level hypothesis named  $H_{\text{vessel}}$  in drafts before Pass 0 is clause (i) conjoined with clause (ii) and is not separately named; the substitution is recorded in the concordance.

The cost of clause (i) is stated at Proposition 3.1(iv) and should be read with it: under clauses (i) and (ii) the fibres of the codon-to-amino-acid assignment are the classes, of sizes two and four, and by [Sch26e, Prop. 3.15(ii)] the standard genetic code is not such a surjection.  $H_{\text{dual}}$  therefore yields the code of Section 5 and not the standard code, and the transformation of Section 6 is not generated by the frame.

**Hypothesis 2.30** ( $H_{\text{chem}}$ , chemical instantiation). As [Sch26d, Hyp. 2.15].

**Remark 2.31** (Status of  $H_{\text{chem}}$ ). [E] As [Sch26d, Rem. 2.16]: used in no proof, and no theorem below becomes false, or loses a hypothesis, if it fails.

## 2.6 Notation concordance

| Symbol                 | Meaning                                                       |
|------------------------|---------------------------------------------------------------|
| $\Sigma'_0$            | the nucleobase alphabet $\{A, C, G, U\}$                      |
| $L$                    | the letter assignment of Definition 2.14                      |
| $M K, R Y, W S$        | the three IUPAC two-class partitions of $\Sigma'_0$           |
| cluster                | a stack of six sites, three vessels; never called a vessel    |
| $\mathcal{A}$          | the $20 \times 20$ adjacency of Section 5                     |
| $\varphi$              | the rank-one bijection $X \rightarrow$ amino acids, Section 5 |
| $\tau_1, \tau_2$       | the two block operators of Section 6                          |
| $\varphi_{\text{std}}$ | the standard genetic code as a map of 64 codons               |

**Remark 2.32** (Citation keys). [Sch26d] retains here the key it carries in [Sch26e], namely the encoded-stack paper; [Sch26e] is new in this paper and denotes the twenty-registry-classes paper, which [Sch26e] itself could not cite. Every locator into [Sch26a]–[Sch26e] was checked individually against the source at the concordance pass and not against this paper’s own earlier use of it, the precaution [Sch26e, Rem. B.2] records as necessary after the reassignment of ninety locators in the predecessor.

## 3 Traps and Non-Implications

The formalism’s governing restraint is that restoring a foundational object never licenses restoring an unproved implication of that object [Sch26b, §3.1], of which the standing non-implications [Sch26b, Eqs. (2)–(3)] are the recorded instances. This section states, in that spirit, what the constructions of Sections 4–8 do not establish. It is placed before them rather than after, for the reason given at Remark 3.5. Each clause is refuted by a witness named in its proof, so the section adds no new argument; its purpose is to make the list checkable.

**Proposition 3.1** (Non-implications). [D] Assume  $H_{\text{bij}}$ ,  $H_{\text{sec}}$  and  $H_{\text{base}}$  where the statement refers to the nucleobase alphabet. None of the following implications holds; each is refuted by the witness named in its proof.

- (i)  $|X| = |\text{Mult}_N|$  at  $N = 3 \not\Rightarrow$  any map between the twenty classes and the twenty three-letter multisets, or equality at any other length.
- (ii) The five filter scores of Section 4  $\not\Rightarrow$  a bijection  $X \rightarrow$  amino acids consistent with the standard genetic code. The scores rank candidates; they do not produce them.
- (iii) The filter scores  $\not\Rightarrow$  either operator of Section 6: neither the block–target pairs of  $\tau_1$  nor, in the thirteen halves where a choice arises, the amino acid selected by  $\tau_2$ .

- (iv)  $H_{\text{dual}}(i)$  and (ii)  $\not\Rightarrow$  the standard genetic code. Under those clauses the fibres of the codon-to-amino-acid assignment are the classes, and the standard code's are not.
- (v) Exhibition of the letter assignment  $L$  on empirical grounds  $\not\Rightarrow$  discharge of [Sch26d, Target 10.5] or of [Sch26e, Target 7.7].
- (vi) Existence of a perfect matching in the adjacency of Section 5  $\not\Rightarrow$  that the standard genetic code is distinguished among codes of its degeneracy profile.

*Proof.* (i) This is [Sch26e, Prop. 3.15(i)], imported and not re-proved: the two counts are  $2^{2N-2} + 2^{N-1}$  and  $\binom{N+3}{3}$ , which are 2, 6, 20, 72 and 4, 10, 20, 35 at  $N = 1, 2, 3, 4$ , agreeing at  $N = 3$  and at no other length, since for  $N \geq 4$  passing from  $N$  to  $N + 1$  multiplies the first by at least 34/9 and the second by at most 2. No map between the two sets is constructed there or here. Remark 3.3 records the further reason the coincidence cannot be read as support.

(ii) By witness, and a single class suffices. The adjacency  $\mathcal{A}$  of Section 5 is defined by the standard genetic code; the score of Section 4 is defined for every pair (class, amino acid) and does not refer to it. Maximising the total score over all 20! bijections, without the adjacency constraint, yields a bijection of total 21.0794, against 18.5639 for the rank-one map of Section 5; it disagrees with that map at all twenty classes and violates the adjacency at sixteen. The witness class is class 4, whose two members are GCU and UAG; the second is a stop by  $H_{\text{stop}}$ , so the class admits Ala and no other amino acid, and the unconstrained maximiser assigns it Trp. The verification is V8 of Appendix B.

(iii) By witness. Under aggregation of the scores over the four codons of a block, no weighting of the five filters makes all seven targets of  $\tau_1$  the maximising amino acid of their block: with non-negative weights the linear programme is infeasible for  $\text{GU} \rightarrow \text{Val}$ ,  $\text{AC} \rightarrow \text{Thr}$ ,  $\text{GC} \rightarrow \text{Ala}$  and  $\text{GG} \rightarrow \text{Gly}$ , and with signed weights no single common vector serves all seven. Under the weights of Section 4 none of the seven targets is the maximiser, these being Phe, Phe, Trp, Phe, Phe, Cys and Cys respectively. The structural reason is that under block aggregation two of the five filters reduce to functions of the second base alone,  $\bar{S} = \mathbf{1}(b_2 \in \{C, U\}) + \frac{1}{2}$  and  $\overline{\text{pHBD}} = \text{HBD}(b_2) + \frac{5}{4}$ , and therefore cannot separate the block CU, which  $\tau_1$  fills, from AU, which it does not. For  $\tau_2$ , of the thirteen halves in which two amino acids are present after  $\tau_1$  and one must be chosen, the best of five candidate rules formulated from the data this paper has — the per-codon score, the class score, the registry weight in either direction, and the class size — selects correctly in ten, and none in all thirteen; and the block UC is filled with serine, which is present nowhere in it, so no selection rule of any kind produces clause (b) of

Definition 6.6. The verifications are V11, V12 and V18 of Appendix B.

(iv) By fibre-size profile. Under  $H_{\text{dual}}(i)$  the amino acid is an Fr-invariant function of the marker assignment and therefore factors through  $X$ , by [Sch26e, Thm. 3.3(iii)]; under clause (ii) it factors bijectively. Its fibres are then the classes, of sizes two and four with multiplicities eight and twelve by [Sch26e, Cor. 3.7], less the three codons removed by  $H_{\text{stop}}$ , giving the profile  $\{1 \times 1, 7 \times 2, 2 \times 3, 10 \times 4\}$ . The standard code has  $\{2 \times 1, 9 \times 2, 1 \times 3, 5 \times 4, 3 \times 6\}$ . Since the multiset of fibre sizes is an invariant of a surjection up to bijections of source and target, and the two multisets differ, no pair of bijections carries one to the other; this is [Sch26e, Prop. 3.15(ii)] in the present instance. The verification is V2 of Appendix B.

(v) [Sch26d, Target 10.5] asks for a construction of the transformation from the structural data of [Sch26d, §§3–6]; Definition 2.14 exhibits one and constructs none. [Sch26e, Target 7.7] asks for a designated element of  $\Sigma'_b$ , the image of the zero the frame designates in  $\Sigma_b$ , supplied by the structure of the secondary realization; by Proposition 2.15(iv) the pair  $\{L, c' \circ L\}$  is separated exactly by that datum, and Remark 2.17 supplies it from a source external to the formalism. An external datum of the right type is not a discharge of a target that asks for the datum from the structure. Both targets are carried unchanged into Section 10.

(vi) By enumeration. A perfect matching exists in the adjacency for 70.6% of five thousand random codes carrying the standard degeneracy profile, and for 25.9% of all  $4!^3 = 13\,824$  position-wise relabellings of the bases in the standard code. Existence is therefore not automatic, and equally not rare. Moreover, among relabellings that admit a matching the number of matchings has median 128, which is the value the standard code itself attains, placing it at the fifty-seventh percentile. The verifications are V13 and V14 of Appendix B.  $\square$

**Remark 3.2** (Status of the individual clauses). Clauses (i), (ii), (iii) and (vi) are refuted by witnesses using no hypothesis beyond  $H_{\text{stop}}$ , which clause (ii) uses only to identify the admissible set of class 4, and could carry the label [A] separately. The proposition is labelled [D] because clauses (iv) and (v) refer to the nucleobase alphabet and therefore use  $H_{\text{bij}}$ ,  $H_{\text{sec}}$  and  $H_{\text{base}}$ .

**Remark 3.3** (The coincidence of counts at  $N = 3$ ). The equation  $2^{2N-2} + 2^{N-1} = \binom{N+3}{3}$  has exactly one positive integer solution,  $N = 3$ , where the common value is twenty; this was verified for  $N = 1, \dots, 20$  and is proved for all  $N$  by the growth comparison of [Sch26e, Prop. 3.15(i)]. Three things must be said about it at once, since the statement invites a reading the body of this paper would otherwise be taken to support.

First, it is not a result of this paper. It is [Sch26e, Prop. 3.15(i)], imported by citation.

Second, its polarity in the predecessor is refutation, not support: [Sch26e, Prop. 7.1(ii)] records the agreement of the two counts as evidence of nothing, and Remark 3.5 below is the reason that reading is restated here rather than at the close.

Third, the reading of the multiset count as a count of chemical environments fails on its own terms. The number of unordered pairs of flanking bases is  $\binom{5}{2} = 10$  and not twenty; a quantity determined by an unordered triple of bases is blind to codon order, whereas the standard code is not, the codons CUG, GUC and UGC carrying three different amino acids; and by [Sch26d, Rem. 4.19] and [Sch26d, Rem. A.15] the twenty multisets are precisely the image of the counterfactual in which the translation unitaries are deleted, that is, of the order-blind degeneration of the encoding. No such reading is used anywhere below, and none is asserted.

**Remark 3.4** (Trap: agreement of counts is not evidence). [A] Restated from [Sch26e, Rem. 4.7] because it applies verbatim to the present construction. Under  $H_{\text{base}}$  and  $H'_{\text{ind}}$  the class set of the nucleobase realization has twenty elements with the same split into eight and twelve, whatever transformation is given and whether or not any is. Consequently

$|X| = |X'|$  and equal size profiles  $\nRightarrow$  any correspondence  $X \rightarrow$

and in particular the agreement of the number of classes with the number of amino acids is a fact about two four-element alphabets and carries no information about a map between them. What Proposition 2.15 establishes is descent, which fails for sixteen of the twenty-four bijections; what Section 5 establishes is a selection among matchings of a specific adjacency. Neither rests on a count agreeing with a count.

**Remark 3.5** (Why the section is placed here). A reader who arrives at Section 5 with the numbers twenty and sixty-four in hand will read the construction as a derivation of the genetic code, and will do so whether or not any such derivation is offered; the reading is invited by the reading-frame analogy of [Sch26d, Rem. 6.12] and by the biogenesis thesis of [Sch26b, §7.2], neither of which this paper uses. The refutations are therefore placed before the construction, exactly as [Sch26e, Rem. 3.17] places its two, and are stated again in Section 10.

**Remark 3.6** (The three that matter most). Clauses (ii), (iv) and (vi) are the ones on which a reader is most likely to

over-read.

Clause (ii) states where the selective work is done. The adjacency reduces  $20! \approx 2.4 \times 10^{18}$  candidate bijections to one hundred and twenty-eight; the filters supply the remaining factor, by a margin of 0.0244 in 14.42, which is 0.17%. The adjacency is built from the standard genetic code. The Chrysene Code of Section 5 is therefore a coarsening of the standard code onto the twenty classes, selected among one hundred and twenty-eight such coarsenings by a physical criterion, and it is not derived from the chrysene system independently of the code. The word *derivation* is used nowhere in Sections 5–7.

Clause (iv) states what the double-sided hypothesis buys and what it costs. It buys the identification of code degeneracy with frame multiplicity: under clause (i) of  $H_{\text{dual}}$  a cluster produces two or four codons and one amino acid, and the degeneracy of the assignment is the orbit structure. It costs the standard code: the fibres so produced have sizes two and four, and the standard code's do not, so the transformation of Section 6 is not generated by the frame and requires a process the formalism does not supply.

Clause (vi) states the strength of the existence result. That the adjacency admits a perfect matching at all is a real constraint — roughly three in ten codes of the standard profile admit none, and three in four base relabellings admit none — but it is not a rare event, and the number of matchings the standard code admits is the median among relabellings that admit any. Existence is reported in Section 5 as a fact and is not offered as evidence.

A reader who grants all three retains exactly what this paper proves: a partition criterion that selects amino/keto over the other two dichotomies, a unique physically preferred coarsening of the standard code onto the twenty classes, and a two-step identity relating that coarsening to the code it came from.

**Remark 3.7** (Numerical verification). Every count and every witness of Proposition 3.1 was recomputed from the definitions — the marker alphabet, the frame group, the registry sequence, the filter formulas and the standard codon table — by a supplementary script that imports nothing from the tables of this paper, so that agreement is evidence and not a restatement, in the manner of [Sch26e, Rem. B.6]. The relevant checks are V2, V8, V11, V12, V13 and V14 of Appendix B. No random sampling is reported except at clause (vi), where the quantity of interest is a frequency and the two populations are stated with the draw counts; the second of them, the 13 824 base relabellings, is enumerated exhaustively and not sampled.

## 4 The Physical Filters

This section defines five numerical scores on pairs consisting of a codon and an amino acid. They are the weight from which Section 5 selects, and by [Sch26e, Prop. 6.1] a selection criterion must be built from a datum attached to the labels of the target, which is what they are. Nothing in this section refers to the standard genetic code; that datum enters at Definition 5.1.

Every statement of this section is [A]: the scores are functions on a finite set, defined by the formulas below, and their properties are proved by computation over that set. Their physical reading is  $H_{\text{crys}}$  of Hypothesis 2.26 and is used in no proof of this section.

### 4.1 The pocket data

**Definition 4.1** (Pocket spaciousness). Let a cluster have markers  $(m_1, \dots, m_6)$ , the odd indices carrying catalyst units and the even indices substrate units. The *pocket spaciousness* is

$$S := \mathbf{1}(m_3 = B) + \mathbf{1}(m_5 = B) \in \{0, 1, 2\},$$

the count of catalyst sites in the second and third vessels carrying the marker  $B$ .

**Lemma 4.2** (Pocket spaciousness in codon coordinates). [A] Under the letter assignment  $L$  of Definition 2.14, the purine/pyrimidine character of the base produced at a vessel is determined by the catalyst marker alone:  $m_{\text{cat}} = F$  gives a purine,  $G$  or  $A$ , and  $m_{\text{cat}} = B$  gives a pyrimidine,  $C$  or  $U$ . Consequently, for a cluster with codon  $b_1b_2b_3$ ,

$$S = \mathbf{1}(b_2 \in \{C, U\}) + \mathbf{1}(b_3 \in \{C, U\}).$$

*Proof.* By Definition 2.14,  $L(FF) = G$  and  $L(FB) = A$  are the purines and  $L(BF) = C$  and  $L(BB) = U$  the pyrimidines; the first coordinate of the marker pair is the catalyst marker by [Sch26d, Def. 3.12]. The displayed identity is Definition 4.1 rewritten, the second and third vessels carrying the markers  $(m_3, m_4)$  and  $(m_5, m_6)$ .  $\square$

**Definition 4.3** (Pocket hydrogen-bond donors). With  $\text{hbd}(G) = 2$  and  $\text{hbd}(A) = \text{hbd}(C) = \text{hbd}(U) = 1$ , the *pocket donor count* of a codon  $b_1b_2b_3$  is  $\text{pHBD} := \text{hbd}(b_2) + \text{hbd}(b_3) \in \{2, 3, 4\}$ . The corresponding acceptor count is  $\text{pHBA} = 4$  for every codon and is carried through the formulas as a constant.

**Remark 4.4** (Molecular reading of the pocket data). [E]  $S$  counts the flanking catalyst sites presenting the smaller

pyrimidine ring, so a larger  $S$  is a more spacious pocket; guanine is the only base presenting two donors at the Watson–Crick face, which is what Definition 4.3 records. This remark enters no proof.

### 4.2 The amino-acid data

The five scores use six functions on the twenty amino acids. All are external data; their provenance is Remark 4.13.

| aa | volume ( $\text{\AA}^3$ ) | rank | hbd <sub>aa</sub> | hba <sub>aa</sub> | KD   | soft / rigid |
|----|---------------------------|------|-------------------|-------------------|------|--------------|
| G  | 48.3                      | 1    | 0                 | 0                 | −0.4 | 0.0 / 0.5    |
| A  | 67.0                      | 2    | 0                 | 0                 | 1.8  | 0.0 / 1.0    |
| S  | 73.0                      | 3    | 1                 | 2                 | −0.8 | 0.0 / 0.8    |
| C  | 86.1                      | 4    | 1                 | 1                 | 2.5  | 2.0 / 1.0    |
| P  | 90.0                      | 5    | 0                 | 0                 | −1.6 | 0.0 / 3.0    |
| D  | 91.6                      | 6    | 0                 | 4                 | −3.5 | 0.0 / 0.8    |
| V  | 91.9                      | 7    | 0                 | 0                 | 4.2  | 0.0 / 1.5    |
| T  | 93.0                      | 8    | 1                 | 2                 | −0.7 | 0.0 / 0.8    |
| N  | 96.0                      | 9    | 1                 | 2                 | −3.5 | 0.0 / 0.8    |
| K  | 105.1                     | 10   | 3                 | 0                 | −3.9 | 0.0 / 0.3    |
| E  | 109.0                     | 11   | 0                 | 4                 | −3.5 | 0.0 / 0.8    |
| Q  | 114.0                     | 12   | 1                 | 2                 | −3.5 | 0.0 / 0.8    |
| H  | 118.0                     | 13   | 1                 | 2                 | −3.2 | 0.3 / 1.5    |
| M  | 124.0                     | 14   | 0                 | 1                 | 1.9  | 1.5 / 1.2    |
| I  | 124.0                     | 15   | 0                 | 0                 | 4.5  | 0.0 / 1.5    |
| L  | 124.0                     | 16   | 0                 | 0                 | 3.8  | 0.0 / 1.5    |
| F  | 135.0                     | 17   | 0                 | 0                 | 2.8  | 0.4 / 2.0    |
| Y  | 141.0                     | 18   | 1                 | 2                 | −1.3 | 0.4 / 1.8    |
| R  | 148.0                     | 19   | 5                 | 1                 | −4.5 | 0.0 / 0.8    |
| W  | 163.0                     | 20   | 1                 | 0                 | −0.9 | 0.5 / 2.0    |

### 4.3 The five scores

**Definition 4.5** (The filter scores). Let  $c = b_1b_2b_3$  be a codon lying in the class  $O$ , with  $S$  and  $\text{pHBD}$  as above,  $|D|$  the Hamming weight of the registry class  $\overline{D}(O)$ , and  $|O| \in \{2, 4\}$  the class size. For an amino acid  $aa$  set

$$\text{Sz}(c, aa) := S \cdot \text{rank}(aa) / 40,$$

$$\text{HB}(c, aa) := (\text{pHBD} \cdot \text{hba}_{aa} + \text{pHBA} \cdot \text{hbd}_{aa}) / 36,$$

$$\text{Hydr}(O, aa) := \text{KD}(aa) \cdot (5 - |D|) / (5 \cdot 4.5),$$

$$\text{Soft}(O, aa) := \text{soft}(aa) \cdot (1 \text{ if } |O| = 2, 0.3 \text{ if } |O| = 4) / 2,$$

$$\text{Rigid}(O, aa) := \text{rigid}(aa) \cdot |D| / (5 \cdot 3),$$

and let  $\sigma(c, aa)$  be their sum. The normalisers  $40 = 2 \cdot 20$  and  $36 = 4 \cdot 4 + 4 \cdot 5$  are the maxima of the two unnormalised numerators.

**Remark 4.6** (Which scores see which datum).  $\text{Sz}$  and  $\text{HB}$  are the only scores depending on the codon beyond its class; they see it through the second and third bases alone.  $\text{Hydr}$ ,  $\text{Soft}$  and  $\text{Rigid}$  are functions of the class and the amino acid. The division matters twice below: at Proposition 4.8, where the first two are shown to be frame-dependent and the last

three not, and at Proposition 3.1(iii), where aggregation of the first two over a codon block collapses them to functions of the second base.

**Remark 4.7** (Molecular reading of the five). [E] Sz scores a larger residue against a more spacious pocket; HB scores complementarity of donors and acceptors between pocket and residue; Hydr scores a hydrophobic residue against an aligned class, hydrophilic residues in aligned classes receiving a negative value, which is a penalty and not a prohibition; Soft scores a soft Lewis base against the symmetry of the class; Rigid scores conformational rigidity against the periodicity of the registry sequence. This remark is  $H_{\text{crys}}$  in words and enters no proof.

## 4.4 Frame dependence, and what survives it

**Proposition 4.8** (The per-codon score is not an assembly quantity). [A] Under flip, the markers of a cluster are all complemented, so  $S \mapsto 2 - S$ ; under  $L$  this is the exchange of purines with pyrimidines at the flanking sites. Hence Sz and HB are not Fr-invariant, and  $\sigma(\cdot, aa)$  is not constant on classes. By [Sch26e, Thm. 3.3(iii)] no such quantity is attributable to the assembly rather than to the reading of it. By contrast Hydr, Soft and Rigid depend only on  $|D|$ ,  $|O|$  and the amino acid, all three of which are class functions, and are therefore Fr-invariant.

*Proof.*  $\text{flip}(m) = (1 - m_1, \dots, 1 - m_6)$  by [Sch26e, Def. 2.13], so  $\mathbf{1}(m_3 = B) + \mathbf{1}(m_5 = B)$  becomes  $2 - (\mathbf{1}(m_3 = B) + \mathbf{1}(m_5 = B))$ , giving  $S \mapsto 2 - S$ ; Lemma 4.2 then gives the statement about bases. For the second assertion,  $\overline{D}$  and  $|O|$  are constant on classes by [Sch26e, Thm. 3.3(i)] and [Sch26e, Prop. 3.6(ii)].  $\square$

**Definition 4.9** (The class score). For a class  $O$  and an amino acid  $aa$ , let  $C(O, aa)$  denote the set of codons of  $O$  that the standard genetic code assigns to  $aa$ , and set

$$\Sigma(O, aa) := \max_{c \in C(O, aa)} \sigma(c, aa) \quad \text{when } C(O, aa) \neq \emptyset.$$

**Proposition 4.10** (The class score is an assembly quantity). [A]  $\Sigma(\cdot, aa)$  is a function on  $X$ , hence Fr-invariant, and is therefore a weight of the kind [Sch26e, Hyp. 2.29] requires. Passing from  $\sigma$  to  $\Sigma$  is what repairs the failure of Proposition 4.8.

*Proof.* A maximum over the elements of a class is constant on that class by construction, and  $C(O, aa)$  is defined from  $O$

as a set.  $\square$

**Remark 4.11** (The selected codon is not an assembly quantity). The codon attaining the maximum in Definition 4.9 is a datum of the reading and not of the assembly, by Proposition 4.8 with [Sch26e, Thm. 3.3(iii)]. It is reported in the score table of Section 5 for reproducibility and carries no structural claim; in particular the “primary codon” column of the stop-codon table of Section 7 is a reading and not an attribute of the class.

**Remark 4.12** (Where the standard code enters). Definition 4.9 is the first statement of this paper to use the standard genetic code, through  $C(O, aa)$ . Everything before it, including all five scores, is defined without it. The consequences are drawn at Definition 5.1 and Proposition 3.1(ii), and the declaration of the datum’s role is Remark 2.4.

## 4.5 Provenance of the numerical data

**Remark 4.13** (The data are external, and one convention is arbitrary). [E] The van der Waals volumes, the side-chain donor and acceptor counts, the Kyte–Doolittle hydrophobicities, the soft-Lewis-base scores and the rigidity scores are external data, tabulated above so that every number entering Section 5 is visible and no score is quoted without its inputs. They were entered once and recomputed at the verification pass; the reproduction of the score table of Section 5 from this table alone is V6 of Appendix B.

One convention is arbitrary and is disclosed. Methionine, isoleucine and leucine share the volume  $124.0 \text{ \AA}^3$ , so their size ranks 14, 15 and 16 are assigned by a tie-break and not by the data. The tie-break is immaterial to every conclusion of this paper: over all six orderings of the three, the maximiser of Section 5 is the same map and the margin between the first and second maps is 0.0244 in every case, only the absolute totals shifting, between 14.3694 and 14.4194. The check is V15 of Appendix B.

## 5 The Adjacency, and the Selection of a Coarsening

### 5.1 The adjacency

**Definition 5.1** (The adjacency). Let  $\mathcal{P}$  be the set of the twenty standard proteinogenic amino acids. The *adjacency* is the relation  $\mathcal{A} \subseteq X \times \mathcal{P}$  given by

$(O, aa) \in \mathcal{A} \iff$  the standard genetic code assigns  $aa$  to some codon

the codons of  $O$  being  $L^{\times 3}(O)$ , of which there are  $|O| \in \{2, 4\}$ . A *coarsening* is a perfect matching of  $\mathcal{A}$ , that is a

bijection  $\beta : X \rightarrow \mathcal{P}$  with  $(O, \beta(O)) \in \mathcal{A}$  for every  $O$ . We write  $\mathcal{M}$  for the set of coarsenings.

**Remark 5.2** (What Definition 5.1 commits the paper to). The adjacency is built from the standard genetic code and from nothing else beyond the class structure and  $L$ . Every statement of Sections 5–7 is therefore conditional on that table, and the word *derivation* is used nowhere in them: a coarsening is a selection from the standard code, not an independent construction that reproduces it. The precise non-implication is Proposition 3.1(ii), and the declaration of the datum’s changed role relative to [Sch26e] is Remark 2.4. The name *coarsening* is chosen to keep this in view: a matching assigns to each class one of the amino acids the standard code already places inside it.

**Remark 5.3** (Trap: the counts agree whatever happens). [A] Restated from Remark 3.4. That  $|X| = |\mathcal{P}| = 20$  is a fact about two sets separately and is not evidence for the existence of a coarsening; what is substantive is that  $\mathcal{A}$  admits a perfect matching at all, which is Theorem 5.4, and how strong that fact is, which is Proposition 3.1(vi).

## 5.2 The partition theorem

The adjacency depends on  $L$ , and  $L$  depends on which partition of  $\Sigma'_0$  is posited by  $H_{\text{base}}$ . The following determines the dependence completely. It is the principal structural result of this paper.

**Theorem 5.4** (Only the amino/keto partition admits a coarsening). [A] Let  $\Pi$  be one of the three partitions of  $\Sigma'_0$  into two classes of two, and let  $\ell$  be any of the eight bijections  $\Sigma_0 \rightarrow \Sigma'_0$  carrying the registry partition to  $\Pi$ . Write  $\mathcal{A}_\ell$  for the adjacency built from  $\ell$  as in Definition 5.1. Then the maximum matching of  $\mathcal{A}_\ell$  has cardinality depending only on  $\Pi$  and not on  $\ell$ , namely

|                                   |                     |       |
|-----------------------------------|---------------------|-------|
| $\Pi = M K = \{A, C\}   \{G, U\}$ | amino / keto        | 20/20 |
| $\Pi = R Y = \{A, G\}   \{C, U\}$ | purine / pyrimidine | 19/20 |
| $\Pi = W S = \{A, U\}   \{C, G\}$ | Watson–Crick        | 18/20 |

A coarsening therefore exists precisely when  $H_{\text{base}}$  posits the amino/keto partition.

*Proof.* By exhaustive computation over all twenty-four bijections  $\Sigma_0 \rightarrow \Sigma'_0$ , of which eight carry the registry partition to each of the three  $\Pi$ ; the maximum matching was computed for each and found constant on each family of eight. This is V1 of Appendix B. The invariance within a family is not an accident of the computation: by [Sch26e, Lem. 4.5] any two bijections carrying the registry partition to  $\Pi$  differ by an element of  $\mathcal{N}$

acting on the source, which permutes the classes of  $X$  among themselves and therefore carries  $\mathcal{A}_\ell$  to  $\mathcal{A}_{\ell'}$  by a relabelling of rows, under which the maximum matching is invariant.  $\square$

**Proposition 5.5** (The obstructions). [A] The failures of Theorem 5.4 are witnessed by Hall violators, as follows, the classes being numbered as in Table 1.

- (i) For  $W|S$ : the six classes  $\{2, 4, 6, 12, 13, 17\}$  have between them the five amino acids  $\{\text{Ala, Cys, Arg, Ser, Thr}\}$  as neighbours.
- (ii) For  $R|Y$ : the nine classes  $\{2, 6, 8, 12, 13, 14, 17, 18, 20\}$  have between them the eight amino acids  $\{\text{Ala, Phe, Ile, Leu, Pro, Ser, Thr, Val}\}$ .

In neither case is any amino acid unmatched in every maximum matching; the obstruction is a property of a set of classes and not of any particular amino acid.

*Proof.* By exhibition of the two sets, whose neighbourhoods are computed from  $\mathcal{A}_\ell$  directly; the final assertion by checking, for each of the twenty amino acids in turn, that some maximum matching includes it. This is V16 of Appendix B.  $\square$

**Remark 5.6** (Correction to a circulated form of Proposition 5.5). Drafts circulated before Pass 0 described the  $W|S$  obstruction as the split codon box of serine and named asparagine and valine as the two unmatched amino acids. The first description is imprecise and the second is not well defined: by Proposition 5.5 no amino acid is unmatched in every maximum matching, so “the two unmatched amino acids” names no object. The statement above replaces both. The correction is recorded in the defect ledger.

**Corollary 5.7** (Instantiation). [D] Assume  $H_{\text{mk}}, H_{\text{stack}}, H_{\text{firm}}, H_{\text{bij}}, H_{\text{sec}}$  and  $H_{\text{base}}$ . Then the classes of a six-site cluster admit a bijection to the twenty amino acids respecting the standard code’s placement, and by Proposition 2.15 the letter assignment realizing it descends to the classification. Neither conclusion holds if  $H_{\text{base}}$  is replaced by either of the other two partitions.

## 5.3 The count, and what the adjacency alone fixes

**Proposition 5.8** (The permanent). [A]  $|\mathcal{M}| = 128$ .

*Proof.*  $|\mathcal{M}|$  is the permanent of the zero-one matrix of  $\mathcal{A}$ , computed by exhaustive enumeration with the classes taken in the order of Table 1; each class has at most four neighbours, so the enumeration is finite and immediate. This is V5 of

Appendix B. □

**Proposition 5.9** (Five assignments are fixed by the adjacency alone). *[A] The one hundred and twenty-eight coarsenings agree on exactly five classes:*

$$\beta(O_2) = \text{Asp}, \quad \beta(O_4) = \text{Ala}, \quad \beta(O_7) = \text{Ser}, \quad \beta(O_{10}) = \text{Trp},$$

for every  $\beta \in \mathcal{M}$ , and on each of the other fifteen classes at least two distinct amino acids occur across  $\mathcal{M}$ . Of the five, class 4 is fixed by its own adjacency, being adjacent to alanine alone once  $H_{\text{stop}}$  removes UAG; the other four are fixed only by the matching condition.

*Proof.* By the enumeration of Proposition 5.8, recorded at V5. For the last assertion: the members of class 4 are GCU and UAG, of which the second is non-coding by  $H_{\text{stop}}$ , leaving alanine as the sole neighbour; whereas class 2 is adjacent to {Asp, Ser} and class 7 to {Ser, Leu}, and neither is a singleton. □

**Remark 5.10** (Why Proposition 5.9 is stated here). It uses the adjacency and nothing else: no filter, no weight, and no part of  $H_{\text{crys}}$ . It is the referent of Theorem 8.7 of Section 8, where the five assignments become predictions under the double-sided hypothesis, and the separation of the two mechanisms of forcing is what makes one of them a single decisive measurement.

## 5.4 The rank-one coarsening

**Definition 5.11** (The total score). For  $\beta \in \mathcal{M}$  set  $\Sigma(\beta) := \sum_{O \in X} \Sigma(O, \beta(O))$ , the class score being that of Definition 4.9. The sum is well defined:  $C(O, \beta(O)) \neq \emptyset$  for every  $O$ , by Definition 5.1.

**Theorem 5.12** (The rank-one coarsening). *[A]  $\Sigma$  attains its maximum over  $\mathcal{M}$  at exactly one element, denoted  $\varphi$ , with*

$$\Sigma(\varphi) = 14.4194, \quad \Sigma(\varphi) - \max_{\beta \neq \varphi} \Sigma(\beta) = 0.0244,$$

a margin of 0.17% of the total. The assignment  $\varphi$  is that of Table 1, and the codon attaining each class maximum is the one tabulated in the score table.

*Proof.* By evaluation of  $\Sigma$  at each of the one hundred and twenty-eight elements of  $\mathcal{M}$  and comparison; the two largest values are 14.4194 and 14.3950. This is V7 of Appendix B, the score table itself being reproduced from Definition 4.5 at V6. □

**Proposition 5.13** (Sensitivity of the selection). *[A] Write  $\Sigma_w$  for the total score with the five scores weighted by  $w \in \mathbb{R}_{\geq 0}^5$ , so that  $\Sigma = \Sigma_{(1,1,1,1,1)}$ .*

- (i) *Setting the weight of Sz or of Hydr to zero leaves  $\varphi$  the maximiser; deleting Sz leaves the margin at 0.0244 unchanged. Setting the weight of HB, Soft or Rigid to zero changes the maximiser.*
- (ii) *Over weights drawn uniformly from  $[0.8, 1.2]^5$ ,  $\varphi$  is the maximiser in 94.5% of two thousand trials; over  $[0.5, 1.5]^5$ , in 72.6%.*
- (iii) *The maximiser and the margin are unchanged under all six orderings of the size ranks of Met, Ile and Leu, whose volumes coincide.*

*Proof.* By computation; V9 for (i) and (ii), V15 for (iii). □

**Remark 5.14** (What the sensitivity means, and what it does not). Three of the five scores are individually decisive and two are not; that Sz leaves the margin exactly unchanged means it contributes the same amount to the first and second coarsenings and separates nothing. The phrase “five combined filters”, used in drafts before Pass 0, therefore overstates by two, and the correction is recorded in the defect ledger. The selection survives a twenty per cent perturbation of the weights in the large majority of draws and a fifty per cent perturbation in a bare majority; both figures are reported here rather than the first alone.

None of this bears on whether the filters model the chemistry. That question is settled negatively outside  $\mathcal{M}$  by Proposition 3.1(ii), and within  $\mathcal{M}$  it is  $H_{\text{crys}}$  and is not discharged. By [Sch26e, Prop. 7.1(viii)] the uniqueness of a maximiser is not a selection by any process, and none is asserted.

## 5.5 The Chrysene Code

**Definition 5.15** (The Chrysene Code). The *Chrysene Code* is the assignment of amino acids to the sixty-four codons obtained by giving every codon of a class  $O$  the amino acid  $\varphi(O)$ , except for the three codons designated by  $H_{\text{stop}}$ , which are non-coding. It is tabulated in Table 4.

**Proposition 5.16** (Degeneracy of the Chrysene Code). *[A] The Chrysene Code has degeneracy profile  $\{1 \times 1, 7 \times 2, 2 \times 3, 10 \times 4\}$  over sixty-one coding codons, with alanine alone on one codon. Every one of the sixteen first-two-position blocks contains four distinct entries except the block UA, which contains three, two of its four codons being stops. In particular no block is uniform.*

*Proof.* The fibres are the classes, of sizes two and four with multiplicities eight and twelve by [Sch26e, Cor. 3.7]; the three codons of  $H_{\text{stop}}$  lie in classes 18, 4 and 15, of sizes four, two and four, reducing one fibre of size two to one and two fibres of size four to three. The block statement is read off Table 4 and was recomputed at V2 and V3 of Appendix B.  $\square$

**Remark 5.17** (The Chrysene Code is not the standard code). By Proposition 5.16 and [Sch26e, Prop. 3.15(ii)] the two are not equivalent as surjections, their fibre-size multisets differing; the standard profile is  $\{2 \times 1, 9 \times 2, 1 \times 3, 5 \times 4, 3 \times 6\}$ . This is not a defect of the construction but its premise: the predecessor proves that a coarsening of the sixty-four codons onto twenty classes cannot be the standard code, and Section 6 measures the distance between them. The block statement of Proposition 5.16 is what makes that measurement possible, since a uniform block would already be a family box.

**Remark 5.18** (Four-distinctness is not evidence). [A] That fifteen of the sixteen blocks contain four distinct entries is a consequence of the class structure and carries no information about the standard code: the four codons of a block differ only in the third base, while the members of a class differ in all three, so no two codons of a block lie in a common class except where a stop intervenes. In particular the observation that each Step-1 block of Section 6 contains exactly one codon already carrying its target is a consequence of four-distinctness together with Definition 5.1, and is not independent evidence for the block assignment. The point is made again at Section 6 where it is needed.

## 6 Transformation Equivalence

Section 5 exhibits a coarsening  $\varphi$  of the standard genetic code onto the twenty classes, and Proposition 5.16 shows that the resulting table is not the standard code, its fibre-size profile differing. This section measures the distance between the two. The measurement is an identity between two tables of sixty-four entries, proved by exhibition; it is not a derivation of either table from the other, and the word is used nowhere below.

### 6.1 Blocks and block operators

**Definition 6.1** (Block). A *block* is the set of four codons sharing their first two bases. There are sixteen, written  $XY$ . The *pyrimidine half* of a block is the pair with third base in  $\{U, C\}$ , the *purine half* the pair with third base in  $\{A, G\}$ .

**Definition 6.2** (Block operator). A *block operator* is a map on assignments of amino acids and stops to the sixty-four codons which alters the entries of one or more blocks and fixes every other codon.

**Remark 6.3** (Blocks are transverse to classes). [A] A block and a class meet in at most one codon, except where a stop intervenes: the four codons of a block differ only in the third base, while the four marker assignments of a class differ under flip and rev in every position. This is the content of Remark 5.18, and it is why the two-step transformation is not a re-reading of the classification: a block operator acts across classes, and no block operator is induced by any operation on  $X$ .

### 6.2 The two operators

**Definition 6.4** ( $\tau_1$ ).  $\tau_1$  is the block operator which fills each of the following seven blocks with the amino acid named, and fixes the remaining nine blocks:

$$\begin{aligned} GU &\rightarrow \text{Val}, & CU &\rightarrow \text{Leu}, & CC &\rightarrow \text{Pro}, & AC &\rightarrow \text{Thr}, \\ GC &\rightarrow \text{Ala}, & CG &\rightarrow \text{Arg}, & GG &\rightarrow \text{Gly}. \end{aligned}$$

The seven block–target pairs are given by exhibition; see Remark 6.8.

**Proposition 6.5** (State after  $\tau_1$ ). [A] In  $\tau_1(\varphi)$  exactly seven amino acids occupy a single codon each, namely Cys at UGU, Phe at UUU, His at CAC, Ile at AUC, Lys at AAG, Met at AUG and Ser at AGC.

*Proof.* By computation from Definition 5.15 and Definition 6.4; this is V17 of Appendix B.  $\square$

**Definition 6.6** ( $\tau_2$ ).  $\tau_2$  is the block operator acting on the nine blocks  $\tau_1$  fixes, as follows.

- Six blocks split at the third base.* In each of UU, UA, CA, AA, AG and GA, each half of Definition 6.1 is made uniform, the amino acid filling a half being one of those already present in it and named by exhibition:  $UU \rightarrow (\text{Phe}, \text{Leu})$ ,  $UA \rightarrow (\text{Tyr}, \text{stop})$ ,  $CA \rightarrow (\text{His}, \text{Gln})$ ,  $AA \rightarrow (\text{Asn}, \text{Lys})$ ,  $AG \rightarrow (\text{Ser}, \text{Arg})$ ,  $GA \rightarrow (\text{Asp}, \text{Glu})$ , pyrimidine half first.
- One block is filled from outside.* UC is filled with Ser, which is present nowhere in that block after  $\tau_1$ , its four entries being Glu, Asn, Thr and Asp.
- Two blocks are exceptions to (a).* UG becomes (Cys, Cys, stop, Trp), which is not uniform on its

purine half; AU becomes (Ile, Ile, Ile, Met), which splits three to one and not two to two.

**Theorem 6.7** (The identity). *[A]  $\tau_2(\tau_1(\varphi)) = \varphi_{\text{std}}$ , the standard genetic code, on all sixty-four codons.*

*Proof.* By exhibition: the sixty-four entries of  $\tau_2(\tau_1(\varphi))$  were computed from Definitions 5.15, 6.4 and 6.6 and compared entry by entry with the standard codon table, with no disagreement. This is V10 of Appendix B.  $\square$

### 6.3 What the identity does not establish

**Remark 6.8** (Both operators are exhibited, not derived). *[A]* Neither operator is determined by anything this paper constructs.

For  $\tau_1$ , the seven block–target pairs are not fixed by the filter scores: no weighting of the five, non-negative or signed, makes all seven targets the maximising amino acid of their block, and under the weights of Section 4 none of the seven is, the maximisers being Phe, Phe, Trp, Phe, Phe, Cys and Cys. This is Proposition 3.1(iii).

For  $\tau_2$ , the *shape* of clause (a) is intrinsic — the third base is split by the purine/pyrimidine dichotomy, which is a property of  $\Sigma'_0$  and not of the code — but the *choice* of which of the two amino acids present fills each half is not. Of the thirteen halves in which a genuine choice arises, the best of five candidate rules formulated from the data available to this paper — the per-codon score, the class score, the registry weight  $|D|$  in either direction, and the class size — selects the correct amino acid in ten, and none selects all thirteen. Clause (b) is worse than a rule that fails: serine is not present in the block it fills, so no selection rule of any kind could produce it. Clause (c) names two further departures.

Both operators are therefore given by exhibition, and Theorem 6.7 is an identity between two exhibited tables.

**Remark 6.9** (The anchors carry no information). *[A]* In each of the seven blocks of  $\tau_1$  exactly one codon already carries the target amino acid in  $\varphi$ . Drafts before Pass 0 offered this as motivating the block assignment. It does not. Existence of such a codon is forced by the construction: the amino acids Val, Pro, Thr, Ala and Gly have all their standard codons inside the block in question, so by Definition 5.1 the class carrying such an amino acid under  $\varphi$  contains one of those codons; and uniqueness is forced by Remark 6.3, a block and a class meeting in at most one codon. The observation is a consequence of four-distinctness and the adjacency, not evidence for the target, and it is recorded

here only because it was previously presented as evidence.

**Remark 6.10** (“Exactly two steps” is withdrawn). *[A]* Drafts before Pass 0 asserted that the Chrysene Code transforms into the standard code in exactly two steps. No minimality is proved, and with block operators admitting arbitrary replacement any table of sixty-four entries reaches any other in two steps; the word “exactly” is withdrawn absent a move set and a minimality argument, which is Target 6.16.

What the decomposition does say is worth stating in its own right, and does not depend on minimality. The discrepancy between  $\varphi$  and  $\varphi_{\text{std}}$  is confined to: seven blocks made uniform; a third-position purine/pyrimidine dichotomy in six of the remaining nine; one block filled from outside; and two named exceptions. That the residual structure is a dichotomy at the third base, in six blocks of nine, is a constraint on the discrepancy and not a restatement of it.

### 6.4 Codon redistribution

**Proposition 6.11** (Gains and losses). *[A]* Passing from  $\varphi$  to  $\varphi_{\text{std}}$ , the codon counts change as follows:

|            |                                                                           |
|------------|---------------------------------------------------------------------------|
| gain + 4 : | Ser (2 $\rightarrow$ 6)                                                   |
| gain + 3 : | Ala (1 $\rightarrow$ 4)                                                   |
| gain + 2 : | Leu (4 $\rightarrow$ 6), Pro (2 $\rightarrow$ 4), Arg (4 $\rightarrow$ 6) |
| gain + 1 : | Ile (2 $\rightarrow$ 3)                                                   |
| unchanged: | Val, Thr, His, Gly, Phe, Asp, Cys                                         |
| loss − 1 : | Lys (3 $\rightarrow$ 2)                                                   |
| loss − 2 : | Tyr, Gln, Asn, Met, Glu                                                   |
| loss − 3 : | Trp (4 $\rightarrow$ 1)                                                   |

*In particular the two largest gains are serine and alanine, and leucine, proline and arginine are tied at +2.*

*Proof.* By counting the fibres of the two tables; this is V2 of Appendix B.  $\square$

**Remark 6.12** (Correction to a circulated claim). *[A]* Drafts before Pass 0 named serine, leucine and arginine as the three amino acids gaining the most codons. Proposition 6.11 shows otherwise: alanine gains more than either leucine or arginine, and proline ties them. The correction is recorded in the defect ledger.

**Remark 6.13** (On prebiotic abundance). *[E]* Drafts before Pass 0 attached to the gains a rider that the amino acids gaining most are biosynthetically early and prebiotically abundant. The rider is withdrawn. On the standard chronologies arginine is a late addition, essentially absent

from spark-discharge and meteoritic inventories and reached by a long biosynthetic path, and leucine is mid-to-late; the six-fold degeneracy of leucine, serine and arginine in the standard code is conventionally referred to codon capture rather than to abundance. Serine and alanine, which Proposition 6.11 identifies as the two largest gains, are prebiotically accessible, and a rider restricted to them would be defensible; none is asserted here, and the question is referred to the domain pass and to Target 6.19.

## 6.5 Disclosure, and the residual question

**Remark 6.14** (The seven blocks are family boxes). [A] The seven blocks of  $\tau_1$  are seven of the eight four-fold family boxes of the standard genetic code, the eighth being UC, which  $\tau_2$  fills at clause (b). This is disclosed rather than presented as a finding: it is the reason  $\tau_1$  cannot be given an intrinsic definition, since identifying those seven blocks is reading the target table.

**Remark 6.15** (The dichotomy changes between the two tables). [A] The classification of Section 5 respects the amino/keto dichotomy: flip is the fixed-point-free element of  $G$ , and under  $L$  it is the involution  $G \leftrightarrow U$ ,  $A \leftrightarrow C$ , which preserves  $M|K$  by Proposition 2.15(i). The residual structure of  $\tau_2$  respects the purine/pyrimidine dichotomy: clause (a) of Definition 6.6 splits the third base by  $R|Y$ . The two dichotomies are distinct, and neither is carried to the other by anything constructed here. This is the sharpest open question the paper raises and it is Target 6.18. Drafts before Pass 0 described it as a transition from wobble to Watson–Crick pairing; that description is withdrawn, the Watson–Crick dichotomy  $W|S$  appearing nowhere in either table.

**Target 6.16** (Minimality). [T] Fix a move set for block operators under which the number of steps is a well-defined cost, and determine the minimum number of moves carrying  $\varphi$  to  $\varphi_{\text{std}}$ . Until this is done the number two is a property of the decomposition exhibited and not of the pair of tables. Hypotheses required: a cost, which neither the formalism nor this paper supplies.

**Target 6.17** (An intrinsic rule). [T] Determine whether either operator admits a definition from the class structure, the filter scores and  $\varphi$  alone. Remark 6.8 records that five natural candidate rules fail for  $\tau_2$ , the best selecting ten of thirteen halves, and that linear reweighting cannot rescue  $\tau_1$ . The space not yet searched is that of rules using data the

filters ignore: stabiliser type, class size, and the distribution of registry classes across a block.

**Target 6.18** (The change of dichotomy). [T] Identify the physical process, if any, that replaces the amino/keto dichotomy carried by the frame group with the purine/pyrimidine dichotomy at the third codon position. Hypotheses required, at minimum: a description of the secondary realization beyond the partition of  $H_{\text{base}}$ , which is [Sch26e, Target 7.5]; and a driving force, which is [Sch26e, Target 7.4] in the form that paper states it.

**Target 6.19** (The abundance rider). [T] Determine whether the gains of Proposition 6.11 correlate with any independent measure of prebiotic accessibility, using inventories not selected after the fact. Remark 6.13 withdraws the claim in its circulated form; what would replace it is a test, not a restatement.

## 7 The Stop Codons

$H_{\text{stop}}$  removes three codons from the coding set. This section records where they sit in the classification, what the filter scores say about the classes containing them, and — at greater length than the positive content warrants — what was previously claimed about them and is here withdrawn.

### 7.1 Position in the classification

**Proposition 7.1** (The three stops lie in three distinct classes). [A] Under the letter assignment  $L$ , the codons UAG, UGA and UAA lie respectively in the classes numbered 4, 15 and 18 of Table 1, of sizes two, four and four. No class contains two of them. Consequently  $H_{\text{stop}}$  reduces one fibre of size two to one and two fibres of size four to three, which is the profile computed at Proposition 5.16.

*Proof.* By inspection of Table 1, recomputed at V3 of Appendix B.  $\square$

**Proposition 7.2** (UAG and GCU are flip-related). [A] Class 4 has size two, so its two marker assignments are  $m$  and  $\text{flip}(m)$  by [Sch26e, Prop. 3.6(i)]; they are  $FFBFBB$  and  $BBFBFF$ , with codons GCU and UAG. Hence UAG is the image of GCU under the letterwise action of flip, which by Proposition 2.15 is the involution  $G \leftrightarrow U$ ,  $A \leftrightarrow C$ .

*Proof.* Immediate from Table 1 and Definition 2.14:  $G \mapsto U$ ,  $C \mapsto A$ ,  $U \mapsto G$ .  $\square$

**Remark 7.3** (This is why class 4 is decisive). Proposition 7.1 is what makes class 4 the single class whose assignment is fixed by its own adjacency, since removing UAG leaves GCU alone and alanine as the sole neighbour. That is Proposition 5.9 and, under the double-sided hypothesis, Target 8.17. The stop designation therefore does more work in this paper as a constraint on the adjacency than as a subject in its own right.

## 7.2 The score ordering

**Proposition 7.4** (Ordering of the three classes). [A] Under  $H_{\text{crys}}$  and with the scores of Definition 4.5, the class scores of the three classes containing a stop, at their assignments under  $\varphi$ , are

$$\Sigma(O_{15}, \text{Met}) = 1.072, \quad \Sigma(O_4, \text{Ala}) = 0.460, \quad \Sigma(O_{18}, \text{Lys})$$

so that  $O_{15} > O_4 > O_{18}$ .

*Proof.* By evaluation of Definition 4.9; the three rows are those of the score table of Section 5, reproduced at V6 of Appendix B.  $\square$

**Remark 7.5** (What the ordering is, and what it is not). Proposition 7.4 is a computation and not a theorem about the formalism. It ranks three numbers produced by a weighted composite whose weights are stipulated in Section 4 and whose physical reading is  $H_{\text{crys}}$ ; by Proposition 5.13 the composite is not robust to arbitrary reweighting, and three of its five components are individually decisive for the selection of  $\varphi$  itself. Nothing in A1–A12, and nothing in [Sch26d] or [Sch26e], bears on the ordering.

## 7.3 Two withdrawals

**Remark 7.6** (The two-and-one partition is withdrawn). [A] Drafts before Pass 0 partitioned the three stops as two mitochondrial and one universal, the division falling between UAG and UAA. The scores do not support that division. The gap between 1.072 and 0.460 is 0.612; the gap between 0.460 and 0.240 is 0.220. The larger gap separates UGA from the other two, so the clustering the numbers themselves suggest is one and two, not two and one, and no threshold is stated in either case. The partition claim is withdrawn; the ordering of Proposition 7.4 is retained.

**Remark 7.7** (The mitochondrial classification is withdrawn). [E] Drafts before Pass 0 classified UGA as a reassignment to methionine, describing it as the most common vertebrate

mitochondrial stop reassignment, UAG as a reassignment to alanine, and UAA as the universal stop. Each of the three is incorrect as a statement about biological systems.

The near-universal mitochondrial reassignment of UGA is to tryptophan, not methionine; it is found in vertebrate, invertebrate, yeast and mould mitochondria and in *Mycoplasma*. The vertebrate mitochondrial code does reassign AUA to methionine, but that is a sense-to-sense change and not a stop reassignment, and it reassigns AGA and AGG to stop. A reassignment of UAG to alanine is not attested; the attested reassignments of UAG are to glutamine in several ciliates, to leucine in some alternative codes, and to pyrrolysine. And UAA is not universal: it is read as glutamine in ciliates.

The classification column is withdrawn from the stop table. The item is referred to the domain pass, which did not dispose it; the present remark records the correction and makes no replacement claim about any organism.

**Remark 7.8** (What survives). Three statements survive the withdrawals. The three stops occupy three distinct classes, of sizes two, four and four (Proposition 7.1). UAG is the flip-image of GCU, which is a structural fact about a size-two class and not a claim about release factors (Proposition 7.2). And the three classes are ordered by class score as  $O_{15} > O_4 > O_{18}$  under  $H_{\text{crys}}$  (Proposition 7.4). No inference from any of the three to the behaviour of any translation system is drawn here, and the composite score is not a binding energy: it carries no physical dimension, and reading it as one would be the type mismatch [Sch26a, Rem. B.2] records and [Sch26d, Rem. 9.10] applies to the constraint functional of A12.

## 7.4 The status of the designation

**Remark 7.9** ( $H_{\text{stop}}$  is not derived). [A] Which three codons are non-coding is supplied by  $H_{\text{stop}}$  and by nothing else. The classification of [Sch26e] supplies no distinguished codons: by [Sch26e, Prop. 3.15(ii)] the surjection of the sixty-four words onto the twenty classes has no element outside a fibre, and the presence of three such elements is one of the two differences that proposition uses to separate the two surjections. Nor does the present paper supply them: the three classes containing stops have sizes two, four and four and registry weights  $|D| = 3, 2$  and 4, and no property distinguishing them from the remaining seventeen has been identified.

**Target 7.10** (A structural characterisation of the stops). [T] Determine whether the three codons of  $H_{\text{stop}}$ , or the three classes containing them, admit a characterisation in terms

of the classification — stabiliser type, class size, registry weight, or the action of  $K$  on  $X$  of [Sch26e, Prop. 5.5] — rather than by designation. A negative answer is as informative as a positive one and would fix  $H_{\text{stop}}$  permanently as an imported datum. Hypotheses required: none beyond those already registered; the question is combinatorial and decidable by enumeration.

**Target 7.11** (The ordering as a physical statement). [T] Determine whether the ordering of Proposition 7.4 corresponds to any measurable property of the three classes. This requires discharging  $H_{\text{crys}}$  in the form [Sch26e, Target 7.5] states it, namely constructing the weight from the structure of the secondary realization rather than stipulating it; until then the ordering is an ordering of three stipulated numbers.

## 8 The Double-Sided Encoding

Sections 5–7 construct and compare two coding tables. Neither construction asserts that any physical system realizes either of them. This section states the hypothesis under which one of them would be realized, determines exactly what that hypothesis supplies mathematically and what it does not, and confines the molecular reading to remarks marked [E] which enter no proof.

The hypothesis is  $H_{\text{dual}}$  of Hypothesis 2.28, in the three clauses registered there. The order of this section follows the clauses: Section 8.1 treats clause (i), which alone has mathematical force; Section 8.2 treats clauses (i) and (ii) together and extracts the predictions that do not depend on the filters; Section 8.3 treats clause (iii); Section 8.4 records what the hypothesis does not supply; Section 8.5 is the molecular reading; Section 8.6 opens the Targets.

### 8.1 Clause (i): the amino acid is a function on the class set

**Proposition 8.1** (Factorization). [D] Assume  $H_{\text{mk}}$ ,  $H_{\text{stack}}$ ,  $H_{\text{frm}}$  and  $H_{\text{dual}}(i)$ , and let  $\alpha : \{0, 1\}^6 \rightarrow \mathcal{P}$  assign to the marker assignment of a cluster the amino acid it produces,  $\mathcal{P}$  being the set of the twenty standard proteinogenic amino acids. Then there is a unique  $\bar{\alpha} : X \rightarrow \mathcal{P}$  with  $\alpha = \bar{\alpha} \circ \bar{D}$ ; that is,  $\alpha$  is a function of the class of the cluster and of nothing else. The fibres of  $\alpha$  are unions of classes, hence of cardinality divisible into parts of size two and four in the profile of [Sch26e, Cor. 3.7].

*Proof.* Clause (i) states that  $\alpha$  is Fr-invariant. By [Sch26e, Thm. 3.3(i)] the map  $m \mapsto \bar{D}(m)$  is the orbit map of Fr on  $\{0, 1\}^6$ , and by [Sch26e, Thm. 3.3(iii)] every frame-invariant function of the marker assignment factors uniquely through it. Uniqueness is the surjectivity of the orbit map. The fibres of

$\alpha$  are preimages under  $\bar{D}$  of the fibres of  $\bar{\alpha}$ , hence unions of classes, and the class sizes are two and four with multiplicities eight and twelve by [Sch26e, Cor. 3.7].  $\square$

**Remark 8.2** (What clause (i) is, stated without the chemistry). Clause (i) says that the amino acid is attributable to the assembly and not to the reading of it. It is exactly the admissibility condition [Sch26e, Prop. 9.13] of the predecessor imposes on any candidate assembly invariant, applied to a quantity the predecessor does not consider. It is a hypothesis and not a theorem: nothing in A1–A12 asserts that a chemical product of a cluster is frame-invariant, and a product that depended on the frame would be a product depending on a datum external to the stack, which  $H_{\text{frm}}$  supplies and the stack does not determine.

### 8.2 Clauses (i) and (ii): the filter-independent predictions

**Proposition 8.3** (Bijection and degeneracy). [D] Assume the hypotheses of Proposition 8.1 together with  $H_{\text{dual}}(ii)$  and  $H_{\text{stop}}$ . Then  $\bar{\alpha} : X \rightarrow \mathcal{P}$  is a bijection of twenty-element sets, and the induced assignment of amino acids to the sixty-one coding codons is the composite of the class map with  $\bar{\alpha}$ . Its degeneracy profile is  $\{1 \times 1, 7 \times 2, 2 \times 3, 10 \times 4\}$ , and it is not the standard genetic code.

*Proof.* A map between finite sets of equal cardinality that is injective is bijective;  $|X| = 20$  by [Sch26e, Thm. 3.3(ii)] and  $|\mathcal{P}| = 20$ . The fibres over the amino acids are the classes by Proposition 8.1, of sizes two and four with multiplicities eight and twelve; removing the three codons designated by  $H_{\text{stop}}$ , which by Table 1 lie in three distinct classes of sizes two, four and four, reduces one fibre of size two to one and two fibres of size four to three, giving the stated profile. The standard code has profile  $\{2 \times 1, 9 \times 2, 1 \times 3, 5 \times 4, 3 \times 6\}$ , and the multiset of fibre sizes is an invariant of a surjection up to bijections of source and target; this is Proposition 3.1(iv).  $\square$

Clauses (i) and (ii) assert that *some* bijection is realized. They do not say which, and in particular they do not say that the realized bijection is one of the one hundred and twenty-eight of Section 5, still less that it is the rank-one map. Those are two further assertions and are separated here.

**Definition 8.4** (The consistency clause).  $H_{\text{dual}}(ii')$ , the *consistency clause*: the bijection  $\bar{\alpha}$  of Proposition 8.3 lies in the set  $\mathcal{M}$  of perfect matchings of the adjacency  $\mathcal{A}$  of Section 5; equivalently, for every class  $O$ , the amino acid  $\bar{\alpha}(O)$  is one that the standard genetic code assigns to some codon of  $O$ .

**Definition 8.5** (The identification clause).  $H_{\text{dual}}(\text{ii}'')$ , the identification clause:  $\bar{\alpha} = \varphi$ , the rank-one map of Section 5.

**Proposition 8.6** (The two clauses are not equivalent). *[A] The identification clause implies the consistency clause and is not implied by it:  $|\mathcal{M}| = 128$ , so the consistency clause admits one hundred and twenty-eight bijections and the identification clause exactly one. Neither is implied by clauses (i) and (ii) of  $H_{\text{dual}}$ , which constrain  $\bar{\alpha}$  only to be a bijection; the bijections  $X \rightarrow \mathcal{P}$  number  $20!$ , of which 128 satisfy the consistency clause, a fraction of about  $5 \times 10^{-17}$ .*

*Proof.*  $\varphi \in \mathcal{M}$  by its construction in Section 5, giving the implication. The count  $|\mathcal{M}| = 128$  is the permanent of  $\mathcal{A}$ , computed at V5 of Appendix B; a set of one hundred and twenty-eight elements is not a set of one, giving the failure of the converse. The last assertion is  $128/20!$ .  $\square$

The consistency clause, weak as it is relative to identification, already fixes five of the twenty assignments, and it fixes them without reference to  $H_{\text{crys}}$  or to any filter weight. This is the strongest empirical content the construction has, and it is stated separately for that reason.

**Theorem 8.7** (Filter-independent assignments). *[D] Assume  $H_{\text{dual}}(\text{i})$ , (ii) and the consistency clause, together with  $H_{\text{stop}}$ . Then, whichever of the one hundred and twenty-eight bijections is realized,*

$$\bar{\alpha}(O_2) = \text{Asp}, \quad \bar{\alpha}(O_4) = \text{Ala}, \quad \bar{\alpha}(O_7) = \text{Ser}, \quad \bar{\alpha}(O_{10}) = \text{Trp}, \quad \bar{\alpha}(O_{15}) = \text{Met},$$

*the classes being numbered as in Table 1. The remaining fifteen assignments are not fixed by the consistency clause: each of the fifteen classes receives at least two distinct amino acids across  $\mathcal{M}$ . No filter, and hence no part of  $H_{\text{crys}}$ , is used.*

*Proof.* This is Proposition 5.9, whose enumeration is recorded at V5 of Appendix B: the one hundred and twenty-eight coarsenings agree on exactly five classes, which are those listed, and disagree on the other fifteen. The enumeration uses the adjacency  $\mathcal{A}$  and nothing else, and  $\mathcal{A}$  is defined from the class structure, the letter assignment  $L$  and the standard codon table; the filter scores of Section 4 enter no step of it.  $\square$

**Remark 8.8** (Two mechanisms of forcing, and one decisive measurement). The five assignments of Theorem 8.7 are forced in two different ways, and the distinction matters for what a measurement would test.

Class 4 is forced *locally*. Its two members are GCU and UAG; the second is designated non-coding by  $H_{\text{stop}}$ , so the

class is adjacent to alanine and to no other amino acid, and no matching is involved. Any realized  $\bar{\alpha}$  satisfying the consistency clause assigns alanine to class 4, and a single cluster of class 4 producing anything other than alanine refutes the conjunction of clauses (i), (ii), (ii') and  $H_{\text{stop}}$ . Classes 2, 7, 10 and 15 are forced *globally*, by the matching structure rather than by their own adjacency: each is adjacent to two or three amino acids, and the assignment is pinned only because the alternatives are required elsewhere. Class 2 is adjacent to {Asp, Ser} and class 7 to {Ser, Leu}; serine is forced to class 7 because leucine is required by class 11, and aspartate then follows at class 2. A measurement disagreeing with one of these four refutes the conjunction but does not localise the failure to that class.

**Remark 8.9** (The prediction structure in two tiers). Theorem 8.7 and Definition 8.5 give the paper's empirical content a two-tier structure which should be stated plainly in any experimental discussion. Five assignments follow from the consistency clause alone and are independent of the filters, of their weights, and of  $H_{\text{crys}}$  entirely. The other fifteen follow only from the identification clause and therefore inherit everything  $H_{\text{crys}}$  carries, including the margin of 0.17% between the rank-one and rank-two maps recorded in Section 5 and the sensitivity reported there. A measurement agreeing with the first tier and disagreeing with the second would leave clauses (i), (ii) and (ii') standing and refute the identification clause alone, which is the outcome the filter margin makes most likely a priori.

### 8.3 Clause (iii): the codon and the amino acid from one state

**Proposition 8.10** (Degeneracy is frame multiplicity). *[D] Assume  $H_{\text{mk}}$ ,  $H_{\text{stack}}$ ,  $H_{\text{frm}}$ ,  $H_{\text{bij}}$ ,  $H_{\text{base}}$  and  $H_{\text{dual}}(\text{i})$  and (iii). Let a cluster have class  $O$ . Then the codons read from it under the four frame-setters of [Sch26d, Def. 6.2] are the  $|O|$  distinct elements of  $L^{\times 3}(O)$ , which number two or four, while the amino acid produced is the single value  $\bar{\alpha}(O)$ . Consequently the induced map from codons to amino acids is constant on classes, and the degeneracy of a codon is the cardinality of its class.*

*Proof.* By [Sch26d, Prop. A.7] the four frame-setters read four words from one and the same stack, and by [Sch26e, Lem. 2.15] the words so read are the Fr-orbit of the word under the identification  $W$ ; applying  $L^{\times 3}$ , which is Fr-equivariant by Proposition 2.15, the codons read are the image of the class, of cardinality  $|O| \in \{2, 4\}$  by [Sch26e, Prop. 3.6(i)]. The amino acid is  $\bar{\alpha}(O)$  by Proposition 8.1, clause (iii) making it the product of the same marker state. Constancy on classes

and the statement about degeneracy follow.  $\square$

**Remark 8.11** (What Proposition 8.10 says and does not say). It says that under clauses (i) and (iii) the redundancy of the coding table is not a property of the table but of the reading: one physical cluster, read under different projection and direction conventions, presents two or four codons and synthesises one amino acid. That is a structural account of degeneracy, and it is the substantive content of the double-sided hypothesis.

It does not say that the degeneracy so accounted for is the degeneracy of the standard genetic code. By Proposition 8.3 the profile produced is  $\{1 \times 1, 7 \times 2, 2 \times 3, 10 \times 4\}$  and the standard profile is not that; the discrepancy is exactly what Section 6 measures, and by Proposition 3.1(iv) it is not generated by the frame. Whatever closes it is a process the formalism does not supply.

It also makes the coding table a frame-relative object. The codon is not attributable to the assembly, by [Sch26d, Prop. A.7] with [Sch26e, Thm. 3.3(iii)]; the amino acid is, under clause (i). A map from codons to amino acids is therefore well defined only after a frame is fixed, and the frame is supplied by the initiator under  $H_{\text{frm}}$  and not by the cluster. Nothing below depends on this observation; it is recorded because a reader may otherwise take the coding table to be an intrinsic object.

## 8.4 What the hypothesis does not supply

**Remark 8.12** (Four limitations). [A] First,  $H_{\text{dual}}$  does not supply the transformation of Section 6, by Proposition 3.1(iv); under its clauses the fibres are the classes, and the standard code's are not.

Second, it does not discharge any target. The letter assignment is exhibited and not constructed, which is [Sch26d, Target 10.5]; the designated zero of  $\Sigma'_0$  is supplied externally, which is [Sch26e, Target 7.7]; and Proposition 3.1(v) records both.

Third, it bears on no chirality claim. The marker binary is not a chirality, by [Sch26d, Rem. 3.7], its exclusion being kinematic under  $H_{\text{stack}}$  rather than a property of the symmetry stabiliser of  $A_2$ , and [Sch26d, Prop. 10.1(ii)] records the non-implication in both directions. Emergent assembly chirality is [Sch26c, Thm. 47] and [Sch26d, Target 10.4], both open, and no statement of this paper bears on either.

Fourth, the filters are not a free-standing model of the cluster chemistry, and should not be offered as a proxy for  $\bar{\alpha}$  in the absence of the consistency clause. By Proposition 3.1(ii), maximising the filter total over all  $20!$  bijections yields a map of higher total that disagrees with the rank-one map at every one of the twenty classes and violates the adjacency at sixteen, assigning tryptophan to a class whose

only admissible value is alanine. Within  $\mathcal{M}$  the filters are informative; outside it they point elsewhere.

## 8.5 The physical reading

The remarks of this subsection are of type [E] throughout. Each is cited to the issued patents and the companion monographs rather than proved, and none enters a proof of this paper. The provenance of the letter assignment  $L$  is Remark 2.17 and is not repeated.

**Remark 8.13** (The template). [E] Chrysene is a planar four-ring polycyclic aromatic hydrocarbon. Two substitution patterns are relevant: the catalyst species, carrying 2, 5, 8, 11-substitution, hence the dual attachment set  $A_{\mathbb{C}}^{\text{dual}}$  of  $A_1$ ; and the substrate species, carrying 3, 6, 9, 12-substitution, hence the primary set  $A_{\mathbb{C}}^{\text{pri}}$ . The substituent in both is phosphoethoxy,  $-\text{O}-\text{C}-\text{C}-\text{OPO}_3$ . The assignment of species to attachment sets is that of [Sch26d, Def. 2.3] and [Sch26d, Rem. B.3] and is not reopened here. In a  $\pi$ -electron stack the two species alternate, which is the alternation of [Sch26d, Def. 3.1]; six such units are a cluster in the sense of Definition 2.7.

**Remark 8.14** (Amino-acid production). [E] The same cluster that encodes a codon also produces an L-amino acid, by a pathway distinct from the nucleobase pathway. The reaction proceeds from the phosphoethoxy substituent under nitric oxide: the attack is on the substituent arm at C12 of the substrate unit, the nitrogen coupling to the C12- $\alpha$  position, which then engages the unsubstituted C11 of the same substrate unit to close the residue. The catalyst fixes the direction of attack.

Two qualifications are attached. The C11 named here is the substrate's own unsubstituted ring position and not the catalyst's attachment locant; the distinction is necessary because  $A_8$  of [Sch26c] fixes  $\delta_{\text{dc}}(12) = 11$ , so that  $v_{12}$  and  $v_{11}$  are a perimeter-adjacent primary-dual cipher pair,  $\{v_{12}, v_{\delta_{\text{dc}}(12)}\} \in E(\partial G_{\mathbb{C}})$ . That the two positions the reaction engages are exactly a cipher pair is recorded here and used in no proof. And the clause that the catalyst fixes the direction so that the L configuration results is an assertion about configuration at the assembly level; by Remark 8.12 it bears on [Sch26d, Target 10.4] and is not used.

**Remark 8.15** (The double-sided reading). [E] In the intended reading the chrysene cluster is a dual template: the catalyst-substrate interfaces carry the nucleobase encoding, and the substrate ring system carries the amino-acid synthesis, the same orientational state determining both. This

is the physical content of  $H_{\text{dual}}(\text{iii})$ , and it is the reading under which the template writes a codon and synthesises its cognate amino acid without a separate adaptor. The reading is not used in any proof, and nothing in Sections 4–7 depends on it.

**Remark 8.16** (What the reading would explain, and what it would not). [E] Under Proposition 8.10 the reading accounts for degeneracy of the coding table as frame multiplicity, in the profile two and four. It does not account for the degeneracy of the standard code, which is not that profile; it makes no claim about the temporal order of events, about evolutionary history, or about the uniqueness or originality of the chrysene template, of which the paper asserts only that it is a consistent one; and it reports no synthesis. The absence of a reported synthesis under the conditions proposed is the limitation [Sch26b, §9.2] records for the corpus as a whole, and this paper does not remove it.

## 8.6 Targets

**Target 8.17** (Discharge of the consistency clause). [T] Determine, experimentally or by molecular simulation, the amino acid produced by a cluster of class 4, whose two marker assignments are  $FFBFBB$  and  $BBFBFF$ . Under clauses (i), (ii), (ii') and  $H_{\text{stop}}$  the answer is alanine, by Remark 8.8, and the prediction uses no filter and no weight. This is the single measurement that most cheaply bears on the construction, and a negative result refutes the conjunction named. Hypotheses required: a preparation in which the class of a cluster is known independently of its product, which neither this paper nor [Sch26d] supplies.

**Target 8.18** (Discharge of the identification clause). [T] Determine the products of clusters in the fifteen classes on which the elements of  $\mathcal{M}$  disagree, and thereby test  $\bar{\alpha} = \varphi$ . By Remark 8.9 a result agreeing with Theorem 8.7 and disagreeing here refutes the identification clause while leaving clauses (i), (ii) and (ii') standing, and would localise the failure in  $H_{\text{crys}}$  rather than in  $H_{\text{dual}}$ .

**Target 8.19** (Clause (i) as a physical statement). [T] Determine whether the product of a cluster is in fact invariant under the frame group, that is, whether two preparations differing by global marker reversal or by reversal of the site index yield the same amino acid. By Proposition 8.1 clause (i) is what licenses the factorization through the twenty classes, and every statement of this section rests on it. The corresponding question for the nucleobase path-

way is settled in the negative by [Sch26d, Prop. 4.4]: the base depends on the projection reference. Whether the two pathways differ in this respect is not known and is not asserted.

## 9 The Count at $N = 3$

This section records the arithmetic of a coincidence that the classification invites and that this paper does not use. Its refutation is Proposition 3.1(i) and Remark 3.3, both stated in Section 3 and not repeated here; what follows is only the computation they refer to, placed after the constructions so that no reader meets it before the warning.

**Proposition 9.1** (Uniqueness of the solution). [A] *The equation*

$$2^{2N-2} + 2^{N-1} = \binom{N+3}{3}$$

*has exactly one solution in positive integers, namely  $N = 3$ , at which the common value is twenty.*

*Proof.* This is [Sch26e, Prop. 3.15(i)], imported and not reproved; the argument is recalled for convenience. At  $N = 1, 2, 3, 4$  the two sides are 2, 6, 20, 72 and 4, 10, 20, 35, so the left side is smaller at  $N = 1, 2$ , equal at  $N = 3$ , and larger at  $N = 4$ . For  $N \geq 4$  write  $a = 2^{2N-2}$  and  $b = 2^{N-1}$ , so that  $a \geq 8b$ ; passing from  $N$  to  $N + 1$  multiplies the left side by  $(4a + 2b)/(a + b) \geq 34/9 > 3$  and the right side by  $(N + 4)/(N + 1) \leq 2$ . The left side therefore remains strictly larger for every  $N \geq 4$ , and the solution at  $N = 3$  is unique.  $\square$

**Remark 9.2** (Numerical verification). The two sides were evaluated for  $N = 1, \dots, 20$ ; the difference is  $-2, -4, 0$ , and thereafter strictly positive and increasing, reaching  $2.7 \times 10^{11}$  at  $N = 20$  against a right side of 1771. This is V19 of Appendix B.

**Corollary 9.3** (Instantiation). [D] *Assume  $H_{\text{mk}}$ ,  $H_{\text{stack}}$  and  $H_{\text{frm}}$ . Then the number of frame classes of a stack of  $2N$  sites equals the number of multisets of  $N$  letters drawn from  $\Sigma_{\text{v}}$  exactly when  $N = 3$ , at which both are twenty.*

[Sch26e, Thm. 3.3(ii). for the first count, the standard multiset formula for the second, and Proposition 9.1.  $\square$

**Remark 9.4** (What this section does not say). Corollary 9.3 asserts an equality of two integers and nothing more. No map between the two sets is constructed here or in [Sch26e]; the length  $N = 3$  is fixed in this paper by Remark 2.9, a

cluster of three vessels producing a codon of three bases, and not by any counting argument; and the reading of the multiset count as a count of chemical environments is refuted at Remark 3.3 on three independent grounds. Drafts before Pass 0 offered the equality as a third independent reason for the length  $N = 3$ . It is not a reason for anything, and the offer is withdrawn.

## 10 Non-Implications and Targets

Section 3 states the non-implications before the constructions they guard. This section restates them in one line each, so that the list can be checked against the completed text, and collects the Targets in one place. It adds no argument.

### 10.1 The non-implications restated

- (i) Agreement of the class count with the multiset count at  $N = 3$  implies no map between the two sets and no agreement at any other length (Proposition 3.1(i), Section 9).
- (ii) The filter scores do not produce a bijection consistent with the standard genetic code; they rank one hundred and twenty-eight candidates that the standard code itself supplies (Proposition 3.1(ii), Remark 5.2).
- (iii) The filter scores determine neither operator of Section 6 (Proposition 3.1(iii), Remark 6.8).
- (iv) The double-sided hypothesis does not yield the standard genetic code: under its clauses the fibres are the classes, and the standard code's are not (Proposition 3.1(iv), Remark 8.11).
- (v) Exhibiting the letter assignment on empirical grounds discharges neither [Sch26d, Target 10.5] nor [Sch26e, Target 7.7] (Proposition 3.1(v), Remark 2.16).
- (vi) Existence of a coarsening does not distinguish the standard code among codes of its degeneracy profile (Proposition 3.1(vi), Remark 3.6).

Four further non-implications are stated at the point of use and are not repeated as clauses: that agreement of the two class counts is evidence of nothing about a map between them (Remark 3.4, Remark 5.3); that four-distinctness of the blocks is a consequence of the class structure and not evidence for any block assignment (Remark 5.18, Remark 6.9); that uniqueness of a maximiser is not selection by any process ([Sch26e, Prop. 7.1(viii)], Remark 5.14); and that the composite score of Section 4 is not a binding energy and carries no physical dimension (Remark 7.8).

### 10.2 Inventory of Targets

Nine Targets are opened in this paper.

| Target | Section | Asks for                                                                                  |
|--------|---------|-------------------------------------------------------------------------------------------|
| 6.16   | 6       | a move set and a minimality proof for the two-step decomposition                          |
| 6.17   | 6       | an intrinsic definition of either block operator                                          |
| 6.18   | 6       | the process replacing the amino/keto dichotomy by purine/pyrimidine at the third position |
| 6.19   | 6       | a test of the codon-gain pattern against independent prebiotic inventories                |
| 7.10   | 7       | a structural characterisation of the three stop classes, or a proof that none exists      |
| 7.11   | 7       | a measurable correlate of the score ordering                                              |
| 8.17   | 8       | the product of a cluster of class 4, predicted to be alanine                              |
| 8.18   | 8       | the products of the fifteen classes on which the coarsenings disagree                     |
| 8.19   | 8       | whether the amino-acid pathway is in fact frame-invariant                                 |

**Remark 10.1** (Which are load-bearing). Their number is not offered as evidence of anything. Three are load-bearing, in the sense that the paper's connection to its own physical reading is incomplete without them.

Target 8.19 is the deepest: clause (i) of  $H_{\text{dual}}$  is what licenses the factorization through the twenty classes, and every statement of Section 8 rests on it. The corresponding question for the nucleobase pathway is settled in the negative by [Sch26d, Prop. 4.4], the base depending on the projection reference; whether the two pathways differ in this respect is not known.

Target 8.17 is the cheapest decisive test the construction admits, and is the reason Proposition 5.9 is stated separately from the selection of  $\varphi$ : one cluster of class 4 producing anything other than alanine refutes the conjunction of  $H_{\text{dual}}$ (i), (ii), the consistency clause and  $H_{\text{stop}}$ , and does so without reference to any filter or weight.

Target 6.18 is the sharpest open question the construction raises. The classification respects one dichotomy of the base alphabet and the residual structure of Section 6 respects another, and nothing constructed here carries one to the other.

The remaining six are not load-bearing. Targets 6.16, 6.17 and 7.10 are combinatorial and decidable by enumeration; Target 6.19 and Target 7.11 are domain questions; and Target 8.18 tests a clause whose failure would leave the rest of Section 8 standing.

### 10.3 Status of the predecessors' deferred program

**Remark 10.2** (No target is discharged). No deferred target of [Sch26c] is discharged here, none of the nine targets of [Sch26d] is discharged, and none of the eight of [Sch26e] is discharged. Four are touched and the manner of the touching is recorded so that it is not mistaken for a discharge.

[Sch26d, Target 10.5], the construction of the transformation, acquires in Definition 2.14 an exhibited instance and in Proposition 2.15 a proof that the instance satisfies the criterion of [Sch26e, Thm. 4.9]. An instance is not a construction, and the provenance of the instance is the empirical remark 2.17.

[Sch26e, Target 7.7], a designated zero for the secondary alphabet, is supplied here from outside the formalism, by the same remark. The target asks for the datum from the structure of the secondary realization; supplying it externally leaves the target open, and Proposition 3.1(v) records this.

[Sch26e, Target 7.5], the provenance of the weight, is engaged by Section 4: the five filters are a weight of the kind [Sch26e, Hyp. 2.29] requires, and by Proposition 4.10 the class score is a function on classes as that hypothesis demands. They are stipulated and not constructed from the secondary realization, so the target stands; Target 7.11 restates what discharging it would give.

[Sch26d, Target 10.4], the marker binary and emergent chirality, is touched only negatively: Remark 8.14 asserts a configuration at the assembly level, and Remark 8.12 confines that assertion to an [E] remark used in no proof, so that no statement of this paper bears on [Sch26c, Thm. 47].

**Remark 10.3** (What this paper adds to the inventory). Two of the Targets above are of a kind the corpus has not carried before. Targets 8.17 and 8.18 ask for a measurement rather than for a proof or a datum internal to the formalism, and they are the first statements of the corpus whose discharge would be an experiment. They are stated in the two tiers of Remark 8.9 precisely so that a partial result is informative: agreement with the first tier and failure at the second localises the failure in  $H_{\text{crys}}$  and leaves  $H_{\text{dual}}$  standing.

## 11 Conclusion

The question this paper set out to answer was whether the twenty frame classes of a six-site chrysene stack can be put in correspondence with the twenty standard amino acids, and if so what such a correspondence would and would not establish. The answer to the first part is affirmative and, at one point, sharply constrained; the answer to the second is that rather less follows than the correspondence invites, and the paper is organised so

that the second answer is available to the reader before the first.

**The partition theorem.** The classification of [Sch26e] admits a bijection to the twenty amino acids respecting the standard code's placement of codons precisely when the nucleobase alphabet is partitioned amino from keto. For the purine/pyrimidine partition the Hall deficiency is one and for the Watson-Crick partition it is two, and in each case the obstruction is a set of classes sharing too few neighbours, exhibited at Proposition 5.5; no amino acid is unmatched in every maximum matching, so the obstruction is a property of a set of classes and not of any amino acid (Theorem 5.4, Remark 5.6). The result depends only on the partition and not on which of the eight letter assignments carrying the registry partition to it is chosen, and the invariance is proved and not merely observed. This is the principal structural result of the paper, and it is the one statement here that could not have been anticipated from either predecessor.

**Selection, not derivation.** The adjacency whose matchings are counted is built from the standard genetic code (Definition 5.1). It has permanent one hundred and twenty-eight (Proposition 5.8), and the five physical filters of Section 4 select one of those, uniquely, by a margin of 0.0244 in 14.42 (Theorem 5.12). The division of labour is stated rather than concealed: the adjacency reduces 20! candidates to one hundred and twenty-eight, and the filters supply the remaining factor. The resulting table is therefore a coarsening of the standard code onto the twenty classes, selected among coarsenings on physical grounds, and not a derivation of the code from the chrysene system. Left to themselves, over all 20! bijections, the filters prefer a map of higher total that disagrees at every one of the twenty classes and assigns tryptophan to a class whose only admissible value is alanine (Proposition 3.1(ii)). Within the adjacency they are informative; outside it they point elsewhere.

**The identity, and its two exhibited operators.** The coarsening and the standard code are related by two block operators whose composite reproduces all sixty-four codons (Theorem 6.7). Neither operator is determined by anything constructed here. The seven block-target pairs of  $\tau_1$  are seven of the eight four-fold family boxes of the target table, and no weighting of the filters makes their targets win their blocks; the shape of  $\tau_2$  is intrinsic, the third base being split by the purine/pyrimidine dichotomy in six of nine blocks, but the choice of winner in each half is not, the best candidate rule selecting ten of thirteen, and one block is filled with an amino acid that is not present in it at all (Remark 6.8). The claim that the transformation takes *exactly* two steps is withdrawn for want of a move set (Remark 6.10). What survives is a statement about where the discrepancy lives: seven uniform blocks, a third-position dichotomy in six of the remaining nine, one import and two named exceptions. The anchors that earlier drafts offered as motivating the block assignments are a consequence of four-distinctness and the adjacency

and carry no information (Remark 6.9).

**The double-sided hypothesis, and what it would predict.** If the amino acid a cluster produces is attributable to the assembly rather than to the reading of it, it factors through the twenty classes (Proposition 8.1); if it is in addition injective, it is a bijection and the degeneracy of the coding table is the multiplicity of the class, one cluster presenting two or four codons and synthesising one amino acid (Proposition 8.10). That is a structural account of degeneracy, and it is the substantive content of the hypothesis. It is also where the hypothesis stops: the fibres it produces have sizes two and four, the standard code's do not, and so the transformation of Section 6 is not generated by the frame (Proposition 3.1(iv)).

The hypothesis is nevertheless testable, and in two tiers. Five of the twenty assignments are fixed by the adjacency alone, with no filter and no weight: classes 2, 4, 7, 10 and 15 carry aspartate, alanine, serine, tryptophan and methionine in every one of the one hundred and twenty-eight coarsenings (Proposition 5.9, Theorem 8.7). One of the five is fixed locally rather than by the matching: class 4 contains GCU and UAG, the second designated non-coding, so alanine is its only neighbour, and a single cluster of class 4 producing anything else refutes the conjunction (Remark 8.8, Target 8.17). The other fifteen assignments follow only from the identification of the realized bijection with  $\varphi$ , and inherit the 0.17% margin.

**What is not done.** No axiom is introduced (Remark 2.2), and A3, A4, A5, A7, A9, A10 and A11 are used nowhere (Proposition 2.5), so nothing here is a statement about sequence dependence. No deferred target of [Sch26c], [Sch26d] or [Sch26e] is discharged (Remark 10.2); four are touched, and the manner of the touching is recorded so that it is not mistaken for a discharge. The letter assignment is exhibited and not constructed, and the designated zero it carries is supplied from outside the formalism. The stop designation is an imported datum for which no structural characterisation is known (Remark 7.9). The mitochondrial classification of the stops is withdrawn, its three claims being incorrect as statements about biological systems (Remark 7.7), as is the two-and-one partition, which the scores themselves cluster the other way (Remark 7.6). The codon-gain rider is withdrawn, the two largest gains being serine and alanine and not the three amino acids earlier named (Proposition 6.11, Remark 6.13). And the coincidence between the class count and the multiset count at  $N = 3$ , which earlier drafts offered as a third reason for that length, is refuted on three grounds and used nowhere (Section 9, Remark 3.3); the length is fixed by the chemistry, a cluster of three vessels producing a codon of three bases, and by nothing else.

Six non-implications are recorded (Proposition 3.1) and nine Targets opened (Section 10), of which three are load-bearing. Two of the nine ask for a measurement rather than a proof, which is new in this corpus (Remark 10.3). No synthesis is reported,

and the limitation [Sch26b, §9.2] records for the corpus as a whole is not removed here.

**What remains.** A partition criterion that selects amino from keto and refutes the other two dichotomies, with the obstructions exhibited; a count of one hundred and twenty-eight admissible coarsenings, five of whose assignments are forced before any physical criterion is applied; a unique coarsening selected by five filters of which three discriminate; an identity on sixty-four codons between that coarsening and the standard code, by two operators given by exhibition; and a hypothesis under which the whole becomes a physical claim, with one measurement named that would refute it.

## A The Classes and the Coding Tables

Table 1 is a computation and not a construction, and is included so that every statement of Sections 4–8 may be checked against a finite object rather than against a formula. Tables 4–6 exhibit the coarsening  $\varphi$  of Section 5 and the two block operators of Section 6. Table 8 exhibits the filter scores, and Table 9 the three classes containing a stop.

| #                                                                                                                                 | D     | Sz | Members    | (marker → codon) | Stab        | aa           |         |   |
|-----------------------------------------------------------------------------------------------------------------------------------|-------|----|------------|------------------|-------------|--------------|---------|---|
| Size-2 — symmetric: 4 palindromic (stab = rev, $\overline{D}_N = 0$ ) and 4 anti-palindromic (stab = anti, $\overline{D}_N = 1$ ) |       |    |            |                  |             |              |         |   |
| 1                                                                                                                                 | 00000 | 2  | FFFFFF→GGG | BBBBBB→UUU       | rev         | F            |         |   |
| 2                                                                                                                                 | 00100 | 2  | FFBBB→GAU  | BBBFF→UCG        | anti        | D            |         |   |
| 3                                                                                                                                 | 01010 | 2  | FBFBFF→GUG | BBFBFB→UGU       | rev         | C            |         |   |
| 4                                                                                                                                 | 01110 | 2  | FBFBFB→GCU | BBFBFF→UAG !     | anti        | A            |         |   |
| 5                                                                                                                                 | 10001 | 2  | FBBBBF→AUC | BFFFFB→CGA       | rev         | I            |         |   |
| 6                                                                                                                                 | 10101 | 2  | FBFBFB→ACA | BFBFBF→CAC       | anti        | H            |         |   |
| 7                                                                                                                                 | 11011 | 2  | FBFBFB→AGC | BFBFBF→CUA       | rev         | S            |         |   |
| 8                                                                                                                                 | 11111 | 2  | FBFBFB→AAA | BFBFBF→CCC       | anti        | P            |         |   |
| Size-4 — generic (trivial stabiliser): all four frame-setters yield distinct words (12 classes)                                   |       |    |            |                  |             |              |         |   |
| 9                                                                                                                                 | 00001 | 4  | FFFFFB→GGA | FBBBBB→AUU       | BFFFFF→CGG  | BBBBBF→UUC   | trivial | G |
| 10                                                                                                                                | 00010 | 4  | FFFFFB→GGU | FFBBBB→GUU       | BBFFFF→UGG  | BBBBBF→UUG   | trivial | W |
| 11                                                                                                                                | 00011 | 4  | FFFFBF→GGC | FBFFFF→AGG       | BFB BBB→CUU | BBBBBF→UUA   | trivial | L |
| 12                                                                                                                                | 00101 | 4  | FFFBFB→GAC | FBFFFF→ACG       | BFB BBB→CAU | BBBFBF→UCA   | trivial | T |
| 13                                                                                                                                | 00110 | 4  | FFFBFB→GAG | FBFFFF→GCG       | BBFB BB→UAU | BBBFBF→UCU   | trivial | E |
| 14                                                                                                                                | 00111 | 4  | FFFBFB→GAA | FBFB BB→AAU      | BFBFFFF→CCG | BBBFBF→UCC   | trivial | N |
| 15                                                                                                                                | 01001 | 4  | FFBBBF→GUC | FBBBFF→AUG       | BFFFBB→CGU  | BBFFFB→UGA ! | trivial | M |
| 16                                                                                                                                | 01011 | 4  | FFBBBF→GUA | FBFBFB→AGU       | BFBFBF→CUG  | BBFBFB→UGC   | trivial | V |
| 17                                                                                                                                | 01101 | 4  | FFBFBF→GCA | FBBFBF→ACU       | BFBFBF→CAG  | BBFBFB→UAC   | trivial | Y |
| 18                                                                                                                                | 01111 | 4  | FFBFBF→GCC | FBFBFB→AAG       | BFBFBF→CCU  | BBFBFB→UAA ! | trivial | K |
| 19                                                                                                                                | 11001 | 4  | FBFFFB→AGA | FBBBFB→AUA       | BFFFBF→CGC  | BFB BFB→CUC  | trivial | R |
| 20                                                                                                                                | 11101 | 4  | FBFBFB→AAC | FBBFBF→ACC       | BFBFBF→CAA  | BFBFBF→CCA   | trivial | Q |

**Table 1: The twenty registry classes.** Letter assignment  $L$ :  $FB = A$ ,  $BF = C$ ,  $FF = G$ ,  $BB = U$  (Definition 2.14); the induced partition of the bases is amino | keto, IUPAC  $M|K$  (Remark 2.12). A class is an equivalence class of marker strings under the frame group  $\text{Fr} = \langle \text{flip}, \text{rev} \rangle \cong (\mathbb{Z}/2)^2$ ; all members describe the same physical stack read under different projection or direction conventions ([Sch26e, Thm. 3.3]).  $D$  = registry sequence of the first member; stab = stabiliser type. ! marks the three codons designated non-coding by  $H_{\text{stop}}$  (Hypothesis 2.24); no reassignment is claimed for them, see Remark 7.7. The final column is the coarsening  $\varphi$  of Theorem 5.12; five of its twenty entries — classes 2, 4, 7, 10, 15 — are common to all 128 coarsenings and are fixed without any filter, by Proposition 5.9. One-letter codes: A Ala, C Cys, D Asp, E Glu, F Phe, G Gly, H His, I Ile, K Lys, L Leu, M Met, N Asn, P Pro, Q Gln, R Arg, S Ser, T Thr, V Val, W Trp, Y Tyr.

| $N$ | $2N$ sites | $2^{2N}$ strings | Classes = $2^{2N-2} + 2^{N-1}$ |
|-----|------------|------------------|--------------------------------|
| 1   | 2          | 4                | 2                              |
| 2   | 4          | 16               | 6                              |
| 3   | 6          | 64               | <b>20</b>                      |
| 4   | 8          | 256              | 72                             |
| 5   | 10         | 1024             | 272                            |

**Table 2: Class count**  $2^{2N-2} + 2^{N-1}$ , from Burnside’s lemma applied to the frame group ([Sch26e, Lem. 3.1]). The length treated in this paper is  $N = 3$ , fixed by Remark 2.9: a cluster of three catalyst–substrate vessels produces a codon of three bases. No counting argument is offered for that length. In particular the agreement of the count at  $N = 3$  with the number of amino acids, and with the number of three-letter multisets, is not a reason for it and is refuted at Remark 3.3 and Section 9.

| Partition                      | flip acts as                               | Max match    | Hall def. | Obstruction                                                                     |
|--------------------------------|--------------------------------------------|--------------|-----------|---------------------------------------------------------------------------------|
| $\{A, U\} \mid \{C, G\}$ $W S$ | $A \leftrightarrow U, C \leftrightarrow G$ | 18/20        | 2         | classes $\{2, 4, 6, 12, 13, 17\}$ have five neighbours: Ala, Cys, Arg, Ser, Thr |
| $\{A, G\} \mid \{C, U\}$ $R Y$ | $A \leftrightarrow G, C \leftrightarrow U$ | 19/20        | 1         | classes $\{2, 6, 8, 12, 13, 14, 17, 18, 20\}$ have eight neighbours             |
| $\{A, C\} \mid \{G, U\}$ $M K$ | $A \leftrightarrow C, G \leftrightarrow U$ | <b>20/20</b> | <b>0</b>  | none                                                                            |

**Table 3: The partition criterion** (Theorem 5.4). The flip generator of Fr induces one of the three involutions of the four-base alphabet whose orbits are the classes of a two-class partition; these are exactly the three IUPAC binary classes, Watson–Crick  $W|S$ , purine/pyrimidine  $R|Y$ , and amino/keto  $M|K$ . A coarsening exists precisely under  $M|K$ . The maximum matching depends only on the partition and not on which of the eight letter assignments carrying the registry partition to it is chosen, and the invariance is proved and not merely computed. Obstructions are Hall violators, by Proposition 5.5: in neither failing case is any amino acid unmatched in every maximum matching, so the obstruction is a property of a set of classes and not of any amino acid. Earlier drafts described the  $W|S$  obstruction as serine’s split codon box and named two unmatched amino acids; both descriptions are withdrawn, see Remark 5.6.

| 1st | 3rd | U   |    | C   |    | A   |    | G   |    | 3rd |
|-----|-----|-----|----|-----|----|-----|----|-----|----|-----|
|     |     | cod | aa | cod | aa | cod | aa | cod | aa |     |
| U   | U   | UUU | F  | UCU | E  | UAU | E  | UGU | C  | U   |
|     | C   | UUC | G  | UCC | N  | UAC | Y  | UGC | V  | C   |
|     | A   | UUA | L  | UCA | T  | UAA | !  | UGA | !  | A   |
|     | G   | UUG | W  | UCG | D  | UAG | !  | UGG | W  | G   |
| C   | U   | CUU | L  | CCU | K  | CAU | T  | CGU | M  | U   |
|     | C   | CUC | R  | CCC | P  | CAC | H  | CGC | R  | C   |
|     | A   | CUA | S  | CCA | Q  | CAA | Q  | CGA | I  | A   |
|     | G   | CUG | V  | CCG | N  | CAG | Y  | CGG | G  | G   |
| A   | U   | AUU | G  | ACU | Y  | AAU | N  | AGU | V  | U   |
|     | C   | AUC | I  | ACC | Q  | AAC | Q  | AGC | S  | C   |
|     | A   | AUA | R  | ACA | H  | AAA | P  | AGA | R  | A   |
|     | G   | AUG | M  | ACG | T  | AAG | K  | AGG | L  | G   |
| G   | U   | GUU | W  | GCU | A  | GAU | D  | GGU | W  | U   |
|     | C   | GUC | M  | GCC | K  | GAC | T  | GGC | L  | C   |
|     | A   | GUA | V  | GCA | Y  | GAA | N  | GGA | G  | A   |
|     | G   | GUG | C  | GCG | E  | GAG | E  | GGG | F  | G   |

**Table 4: The Chrysene Code  $\varphi$**  (Definition 5.15): every codon of a class carries that class's amino acid under the rank-one coarsening. Degeneracy:  $10 \times 4, 2 \times 3, 7 \times 2, 1 \times 1$  (Ala). Fifteen of the sixteen blocks contain four distinct entries, the exception being UA, which contains three because two of its codons are stops. By Remark 6.3 this is a consequence of blocks being transverse to classes and is not evidence for any block assignment.

| 1st | 3rd | U   |                    | C   |                    | A   |    | G   |                    | 3rd |
|-----|-----|-----|--------------------|-----|--------------------|-----|----|-----|--------------------|-----|
|     |     | cod | aa                 | cod | aa                 | cod | aa | cod | aa                 |     |
| U   | U   | UUU | F                  | UCU | E                  | UAU | E  | UGU | C                  | U   |
|     | C   | UUC | G                  | UCC | N                  | UAC | Y  | UGC | V                  | C   |
|     | A   | UUA | L                  | UCA | T                  | UAA | !  | UGA | !                  | A   |
|     | G   | UUG | W                  | UCG | D                  | UAG | !  | UGG | W                  | G   |
| C   | U   | CUU | L                  | CCU | $P_{\leftarrow K}$ | CAU | T  | CGU | $R_{\leftarrow M}$ | U   |
|     | C   | CUC | $L_{\leftarrow R}$ | CCC | P                  | CAC | H  | CGC | R                  | C   |
|     | A   | CUA | $L_{\leftarrow S}$ | CCA | $P_{\leftarrow Q}$ | CAA | Q  | CGA | $R_{\leftarrow I}$ | A   |
|     | G   | CUG | $L_{\leftarrow V}$ | CCG | $P_{\leftarrow N}$ | CAG | Y  | CGG | $R_{\leftarrow G}$ | G   |
| A   | U   | AUU | G                  | ACU | $T_{\leftarrow Y}$ | AAU | N  | AGU | V                  | U   |
|     | C   | AUC | I                  | ACC | $T_{\leftarrow Q}$ | AAC | Q  | AGC | S                  | C   |
|     | A   | AUA | R                  | ACA | $T_{\leftarrow H}$ | AAA | P  | AGA | R                  | A   |
|     | G   | AUG | M                  | ACG | T                  | AAG | K  | AGG | L                  | G   |
| G   | U   | GUU | $V_{\leftarrow W}$ | GCU | A                  | GAU | D  | GGU | $G_{\leftarrow W}$ | U   |
|     | C   | GUC | $V_{\leftarrow M}$ | GCC | $A_{\leftarrow K}$ | GAC | T  | GGC | $G_{\leftarrow L}$ | C   |
|     | A   | GUA | V                  | GCA | $A_{\leftarrow Y}$ | GAA | N  | GGA | G                  | A   |
|     | G   | GUG | $V_{\leftarrow C}$ | GCG | $A_{\leftarrow E}$ | GAG | E  | GGG | $G_{\leftarrow F}$ | G   |

**Table 5:** After  $\tau_1$  (Definition 6.4): the seven blocks  $GU \rightarrow \text{Val}$ ,  $CU \rightarrow \text{Leu}$ ,  $CC \rightarrow \text{Pro}$ ,  $AC \rightarrow \text{Thr}$ ,  $GC \rightarrow \text{Ala}$ ,  $CG \rightarrow \text{Arg}$ ,  $GG \rightarrow \text{Gly}$  are made uniform; the other nine are fixed. A subscripted arrow marks an entry displaced from the amino acid named. Seven amino acids now occupy a single codon each: Cys (UGU), Phe (UUU), His (CAC), Ile (AUC), Lys (AAG), Met (AUG), Ser (AGC) — Proposition 6.5. Earlier drafts underlined in each block the one codon already carrying the target and offered it as motivating the assignment; the underlining is removed, that codon being forced by the construction and carrying no information (Remark 6.9). The seven blocks are seven of the eight four-fold family boxes of the target table, which is why  $\tau_1$  admits no intrinsic definition (Remark 6.14).

| 1st | 3rd | U   |                    | C   |                    | A   |                    | G   |                    | 3rd |
|-----|-----|-----|--------------------|-----|--------------------|-----|--------------------|-----|--------------------|-----|
|     |     | cod | aa                 | cod | aa                 | cod | aa                 | cod | aa                 |     |
| U   | U   | UUU | F                  | UCU | $S_{\leftarrow E}$ | UAU | $Y_{\leftarrow E}$ | UGU | C                  | U   |
|     | C   | UUC | $F_{\leftarrow G}$ | UCC | $S_{\leftarrow N}$ | UAC | Y                  | UGC | $C_{\leftarrow V}$ | C   |
|     | A   | UUA | L                  | UCA | $S_{\leftarrow T}$ | UAA | !                  | UGA | !                  | A   |
|     | G   | UUG | $L_{\leftarrow W}$ | UCG | $S_{\leftarrow D}$ | UAG | !                  | UGG | W                  | G   |
| C   | U   | CUU | L                  | CCU | P                  | CAU | $H_{\leftarrow T}$ | CGU | R                  | U   |
|     | C   | CUC | L                  | CCC | P                  | CAC | H                  | CGC | R                  | C   |
|     | A   | CUA | L                  | CCA | P                  | CAA | Q                  | CGA | R                  | A   |
|     | G   | CUG | L                  | CCG | P                  | CAG | $Q_{\leftarrow Y}$ | CGG | R                  | G   |
| A   | U   | AUU | $I_{\leftarrow G}$ | ACU | T                  | AAU | N                  | AGU | $S_{\leftarrow V}$ | U   |
|     | C   | AUC | I                  | ACC | T                  | AAC | $N_{\leftarrow Q}$ | AGC | S                  | C   |
|     | A   | AUA | $I_{\leftarrow R}$ | ACA | T                  | AAA | $K_{\leftarrow P}$ | AGA | R                  | A   |
|     | G   | AUG | M                  | ACG | T                  | AAG | K                  | AGG | $R_{\leftarrow L}$ | G   |
| G   | U   | GUU | V                  | GCU | A                  | GAU | D                  | GGU | G                  | U   |
|     | C   | GUC | V                  | GCC | A                  | GAC | $D_{\leftarrow T}$ | GGC | G                  | C   |
|     | A   | GUA | V                  | GCA | A                  | GAA | $E_{\leftarrow N}$ | GGA | G                  | A   |
|     | G   | GUG | V                  | GCG | A                  | GAG | E                  | GGG | G                  | G   |

**Table 6: After  $\tau_2$**  (Definition 6.6):  $\tau_2(\tau_1(\varphi)) = \varphi_{\text{std}}$  on all sixty-four codons (Theorem 6.7, verified at V10). Six blocks — UU, UA, CA, AA, AG, GA — are made uniform on each half by the purine/pyrimidine dichotomy at the third base; UC is filled with Ser, which is present nowhere in that block after  $\tau_1$ ; and UG and AU are exceptions, retaining Cys/stop/Trp and the three-to-one Ile/Met split respectively. The shape of the first clause is intrinsic; the choice of which amino acid fills each half is not, and neither is  $\tau_1$  (Remark 6.8). The claim that the transformation takes *exactly* two steps is withdrawn for want of a move set (Remark 6.10).

| Codons/aa | Chrysene Code $\varphi$                          | Standard genetic code                       |
|-----------|--------------------------------------------------|---------------------------------------------|
| 1         | Ala                                              | Met, Trp                                    |
| 2         | Cys, Asp, Phe, His, Ile, Pro, Ser                | Cys, Asp, Glu, Phe, His, Lys, Asn, Gln, Tyr |
| 3         | Lys, Met                                         | Ile                                         |
| 4         | Glu, Gly, Leu, Asn, Gln, Arg, Thr, Val, Trp, Tyr | Ala, Gly, Pro, Thr, Val                     |
| 6         | (none)                                           | Leu, Ser, Arg                               |
| Profile   | {1×1, 7×2, 2×3, 10×4}                            | {2×1, 9×2, 1×3, 5×4, 3×6}                   |
| Stops     | 3 (UAA, UAG, UGA)                                | 3 (UAA, UAG, UGA)                           |

**Table 7: Degeneracy comparison** (Proposition 5.16).  $\varphi$  has no six-fold class and a maximum of four, reflecting the eight-and-twelve class structure; the two surjections are therefore not equivalent, by [Sch26e, Prop. 3.15(ii)], which is the premise of Section 6 and not a defect of the construction. Codon gains from  $\varphi$  to  $\varphi_{\text{std}}$ , by Proposition 6.11: Ser  $2 \rightarrow 6$  (+4); Ala  $1 \rightarrow 4$  (+3); Leu  $4 \rightarrow 6$ , Pro  $2 \rightarrow 4$ , Arg  $4 \rightarrow 6$  (each +2); Ile +1; seven unchanged; Lys −1; Tyr, Gln, Asn, Met, Glu each −2; Trp  $4 \rightarrow 1$  (−3). Earlier drafts named Ser, Leu and Arg as the three largest gains and attached a prebiotic-abundance rider; both are withdrawn (Remarks 6.12, 6.13).

| #  | aa | Codon | D | S | pHBD | Sz    | HB    | Hydr   | Soft  | Rigid | Total  |
|----|----|-------|---|---|------|-------|-------|--------|-------|-------|--------|
| 1  | F  | UUU   | 0 | 2 | 2    | 0.850 | 0.000 | 0.622  | 0.200 | 0.000 | 1.672  |
| 2  | D  | GAU   | 1 | 1 | 2    | 0.150 | 0.222 | -0.622 | 0.000 | 0.053 | -0.197 |
| 3  | C  | UGU   | 2 | 1 | 3    | 0.100 | 0.194 | 0.333  | 1.000 | 0.133 | 1.761  |
| 4  | A  | GCU   | 3 | 2 | 2    | 0.100 | 0.000 | 0.160  | 0.000 | 0.200 | 0.460  |
| 5  | I  | AUC   | 2 | 2 | 2    | 0.750 | 0.000 | 0.600  | 0.000 | 0.200 | 1.550  |
| 6  | H  | CAC   | 3 | 1 | 2    | 0.325 | 0.222 | -0.284 | 0.150 | 0.300 | 0.713  |
| 7  | S  | AGC   | 4 | 1 | 3    | 0.075 | 0.278 | -0.036 | 0.000 | 0.213 | 0.531  |
| 8  | P  | CCC   | 5 | 2 | 2    | 0.250 | 0.000 | 0.000  | 0.000 | 1.000 | 1.250  |
| 9  | G  | GGA   | 1 | 0 | 3    | 0.000 | 0.000 | -0.071 | 0.000 | 0.033 | -0.038 |
| 10 | W  | UGG   | 1 | 0 | 4    | 0.000 | 0.111 | -0.160 | 0.075 | 0.133 | 0.159  |
| 11 | L  | CUU   | 2 | 2 | 2    | 0.800 | 0.000 | 0.507  | 0.000 | 0.200 | 1.507  |
| 12 | T  | ACG   | 2 | 1 | 3    | 0.200 | 0.278 | -0.093 | 0.000 | 0.107 | 0.491  |
| 13 | E  | GAG   | 2 | 0 | 3    | 0.000 | 0.333 | -0.467 | 0.000 | 0.107 | -0.027 |
| 14 | N  | AAU   | 3 | 1 | 2    | 0.225 | 0.222 | -0.311 | 0.000 | 0.160 | 0.296  |
| 15 | M  | AUG   | 2 | 1 | 3    | 0.350 | 0.083 | 0.253  | 0.225 | 0.160 | 1.072  |
| 16 | V  | GUA   | 3 | 1 | 2    | 0.175 | 0.000 | 0.373  | 0.000 | 0.300 | 0.848  |
| 17 | Y  | UAC   | 3 | 1 | 2    | 0.450 | 0.222 | -0.116 | 0.060 | 0.360 | 0.977  |
| 18 | K  | AAG   | 4 | 0 | 3    | 0.000 | 0.333 | -0.173 | 0.000 | 0.080 | 0.240  |
| 19 | R  | CGC   | 3 | 1 | 3    | 0.475 | 0.639 | -0.400 | 0.000 | 0.160 | 0.874  |
| 20 | Q  | CAA   | 4 | 0 | 2    | 0.000 | 0.222 | -0.156 | 0.000 | 0.213 | 0.280  |

**Table 8: Filter scores for the rank-one coarsening** (unique *maximiser* over the 128, Theorem 5.12). Definitions are at Definition 4.5; all five scores are shown and the Total is their sum.  $|D|$  is the Hamming weight of the registry class, read from Table 1; earlier drafts used  $|D| = 2$  at class 2 and  $|D| = 3$  at class 20, contradicting the registry sequences 00100 and 11101 of that table, and the two rows are corrected here to  $-0.197$  and  $+0.280$ . The two errors cancel exactly in every one of the 128 totals, so the maximiser and the margin are unaffected. The Codon column is a datum of the reading and not of the assembly (Remark 4.11); the score is not an energy and carries no physical dimension (Remark 7.8).

| Stop | Class | Size | aa under $\varphi$ | Class score |
|------|-------|------|--------------------|-------------|
| UGA  | 15    | 4    | Met                | 1.072       |
| UAG  | 4     | 2    | Ala                | 0.460       |
| UAA  | 18    | 4    | Lys                | 0.240       |

**Table 9: The three classes containing a stop** (Proposition 7.1). Each stop lies in a distinct class; class scores from Table 8, ordered  $O_{15} > O_4 > O_{18}$  (Proposition 7.4). UAG is the flip-image of GCU, its class having size two (Proposition 7.2); and class 4 is the one class fixed by its own adjacency, which is what makes it the decisive single measurement of Target 8.17. Earlier drafts added a classification column partitioning the three as two mitochondrial and one universal. It is withdrawn: the score gaps cluster the three one and two rather than two and one (Remark 7.6), and the three biological claims attached to it are incorrect (Remark 7.7). No inference to the behaviour of any translation system is drawn.

## B Concordance, Domain Items, Defect Ledger, and Verification

### B.1 Scope and method

This appendix records, for every prior statement of the corpus that this paper touches, whether the statement is *preserved* (used unchanged), *extended* (used with a further hypothesis or conclusion, without alteration of its content), *restated* (reformulated here in a manner the reviewer must check is equivalent), *corrected* (found false or type-incorrect as written, with the repair given), *not applicable* (outside the scope of the objects considered here), or *noted* (recorded without change of status). The classification is made against the criterion  $C$  of [Sch26b, Def. 2]: a statement of this paper is concordant when it neither contradicts a prior statement of the formalism nor silently weakens a structure a prior statement relies upon.

One entry requires comment before the tables, because it is the entry a reviewer will look for first. [Sch26e, Prop. 3.15(iii)] states that no identification of the twenty classes with any twenty-element set, combinatorial or biological, is asserted *in that paper*. This paper asserts one. That is not a contradiction: clause (iii) is a statement about what its own paper asserts, verified there by inspection of its statements, and is not a prohibition on any later paper. What would contradict the predecessor is an identification inconsistent with its clause (ii), which proves that the class surjection and the genetic code are not the same map; Section 5 is consistent with that clause and indeed uses it, the difference in fibre-size profile being the premise of Section 6 rather than an obstacle to it. The entry is recorded as *preserved and used*, not as *extended*.

### B.2 Concordance with the twenty-registry-classes paper

| Prior statement                                        | Disposition                                                                                                                                       |
|--------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Thm. 3.3, the frame classification                     | Preserved; its class set is the object classified here                                                                                            |
| Cor. 3.7, eight and twelve                             | Preserved; used at Propositions 8.3 and 5.16                                                                                                      |
| Prop. 3.15(i), the multiset coincidence                | Preserved with its polarity: imported at Proposition 9.1 and refuted, not used (Remark 3.3)                                                       |
| Prop. 3.15(ii), the two surjections differ             | Preserved and <i>used</i> : it is the premise of Section 6 and the content of Proposition 3.1(iv)                                                 |
| Prop. 3.15(iii), no identification asserted there      | Preserved; see the comment in B.1. This paper asserts an identification; that paper's clause is about that paper                                  |
| Rem. 3.16, the external datum                          | Preserved and extended: the datum's role changes from refutation to construction, declared at Remark 2.4; the profile is recomputed and not cited |
| Thm. 4.9, transport and descent                        | Preserved; instantiated at Proposition 2.15                                                                                                       |
| Lem. 4.5, isomorphisms of structured alphabets         | Preserved; supplies the invariance across the eight letter assignments in Theorem 5.4                                                             |
| Prop. 4.14, eight of twenty-four                       | Preserved; instantiated at Proposition 2.15(ii)                                                                                                   |
| Cor. 5.4 and Thm. 5.8, four maps and the residual pair | Preserved; the residual pair is the exchange $G \leftrightarrow U$ , $A \leftrightarrow C$ at Proposition 2.15(iv)                                |
| Prop. 5.5, $K$ on registry sequences                   | Noted; used nowhere here, referred to at Target 7.10                                                                                              |
| Hyp. 2.23, $H_{\text{sec}}$                            | Instantiated by $H_{\text{base}}$ (Hypothesis 2.11); Remark 2.13 records that the partition is the weakest of the four equivalent data            |
| Hyp. 2.29, $H_{\text{fit}}$                            | Instantiated: the class score of Definition 4.9 is the required function on classes (Proposition 4.10)                                            |
| Hyp. 2.32, $H_{\text{var}}$                            | Instantiated by the selection of Theorem 5.12; used in no proof, per Remark 2.27                                                                  |
| Prop. 6.1, the invariance trap                         | Preserved; it is why the filters must be label-dependent, Remark 2.27                                                                             |
| Prop. 6.11, the degeneracy locus is null               | Preserved; used to demote the uniqueness of a maximiser from evidence to genericity, ledger item 11                                               |
| Prop. 7.1(viii), uniqueness is not selection           | Preserved; restated at Remark 5.14                                                                                                                |
| Target 7.5, provenance of the weight                   | Engaged, not discharged (Remark 10.2)                                                                                                             |
| Target 7.7, a zero for the secondary alphabet          | Not discharged; supplied from outside the formalism (Remark 2.16)                                                                                 |
| App. A, Tab. 2, the twenty classes                     | Preserved; Table 1 adds a column, which is new material of this paper                                                                             |

### B.3 Concordance with the encoded-stack paper

| Prior statement                                            | Disposition                                                                                                        |
|------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Def. 3.3, the marker                                       | Preserved; its molecular reading and the perimeter-orientation criterion are not reopened                          |
| Def. 3.10 and Rem. 3.11, the vessel                        | Preserved; the six-site object is named <i>cluster</i> , and “reaction vessel” for it is withdrawn (Remark 2.8)    |
| Defs. 3.12–3.15, the two binaries and registry             | Preserved                                                                                                          |
| Prop. 4.4, what depends on the frame                       | Preserved and <i>used to correct</i> : it is why <i>E1</i> is restated with the projection reference (Remark 2.17) |
| Cor. 6.9, the designated zero is aligned                   | Preserved; used in the same remark                                                                                 |
| Props. 4.7, 4.10, the structural group                     | Preserved                                                                                                          |
| Prop. 3.19 and Hyp. 2.11                                   | Preserved; supplies that all sixty-four codons occur                                                               |
| Rem. 3.20, $N = 3$ gives sixty-four words                  | Preserved; cited for the length, ledger item 29                                                                    |
| Prop. A.7, four frame-setters, four words                  | Preserved; used at Proposition 8.10                                                                                |
| Rem. 4.19 and Rem. A.15, the twenty multisets              | Noted; used in the refutation at Remark 3.3, the multisets being the image of the order-blind counterfactual       |
| Thm. 4.17, Prop. 4.18, sequence dependence                 | Not applicable; <i>A7</i> is used nowhere (Proposition 2.5, Remark 2.6)                                            |
| Rem. 3.7 and Prop. 10.1(ii), the marker is not a chirality | Preserved; it is what confines the L-configuration clause of <i>E2</i> to an [E] remark (Remark 8.12)              |
| Rem. 9.10, constraint functionals are not energies         | Preserved; the same discipline is applied to the composite score (Remark 7.8)                                      |
| Rem. 2.20, three senses of registry                        | Preserved; no fourth sense is added (Remark 2.8)                                                                   |
| Rem. B.3, catalyst and substrate assignment                | Preserved; checked against <i>A1</i> at domain item B.5(c)                                                         |
| Target 10.5, the transformation                            | Not discharged; an instance is exhibited and proved admissible (Remark 2.16)                                       |
| Target 10.4, marker binary and chirality                   | Touched negatively only; no statement here bears on it                                                             |

### B.4 Concordance with the twelve-axiom paper and the editorial papers

| Prior statement                                     | Disposition                                                                                                                                      |
|-----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| A1, A2, A6, A8, A12                                 | Preserved; no axiom added (Remark 2.2)                                                                                                           |
| A8, the cipher pairing $\delta_{dc}$                | Preserved; the relation $\delta_{dc}(12) = 11$ is used to disambiguate the locants of <i>E2</i> and is recorded, not used in proof (Remark 8.14) |
| Thm. 61, non-vacuity                                | Preserved, not re-established                                                                                                                    |
| Prop. 39, acyclicity of the ladder                  | Preserved; the imports are confined to five axioms                                                                                               |
| Thm. 47, emergent chirality                         | Not touched; see Remark 8.12                                                                                                                     |
| Thm. 41, the $GF(4)$ target                         | Not applicable; this paper concerns $\Sigma_0$ and $\Sigma'_0$                                                                                   |
| Defs. 1, 3, 5, 6, status vocabulary                 | Preserved                                                                                                                                        |
| [Sch26b, §3.1] the governing restraint              | Preserved; Section 3 is written in its form                                                                                                      |
| [Sch26b, §4.3, §4.5] review bar and scope           | Preserved; this paper is within the class requiring orthogonal-constraint review                                                                 |
| [Sch26b, Rem. 1] differentiated correction profiles | Noted; reproduced here on a third manuscript, see B.8                                                                                            |
| [Sch26b, §9.2] no synthesis reported                | Preserved; not removed here (Remark 8.16)                                                                                                        |
| [Sch26a, Def. 2.7] status labels                    | Preserved                                                                                                                                        |
| [Sch26a, Rem. B.2] type mismatch                    | Preserved; applied to the composite score                                                                                                        |

### B.5 Chemistry and domain items

Four items are placed before the chemistry review. All are of type [E] or are statements about a published table; none is used in a proof.

- (a) **The amino/keto partition.** Remark 2.12 identifies  $\{A, C\} \mid \{G, U\}$  as the amino/keto dichotomy of the exocyclic substituent at the Watson–Crick face, IUPAC *M* and *K*. The review is invited to confirm the chemical description and the correspondence with the IUPAC classes. The formalism uses the partition alone and no chemical property of it.
- (b) **The nucleobase determination.** Remark 2.17 states that the base is fixed by the ordered pair of markers read against the projection reference. The review is invited to confirm that this is the reading intended by the cited patents, the earlier form — determination by relative orientation alone — being inconsistent with [Sch26d, Prop. 4.4].
- (c) **Species and locants.** The assignment  $\mathbf{c} = \{2, 5, 8, 11\} = A_{\mathbb{C}}^{\text{dual}}$ ,  $\mathbf{s} = \{3, 6, 9, 12\} = A_{\mathbb{C}}^{\text{pri}}$  agrees with A1 and with [Sch26d, Def. 2.3]; confirmed at the domain pass. The C11 named in Remark 8.14 is the substrate’s own unsubstituted ring position, and the review is invited to confirm the pathway as restated.
- (d) **Two withdrawals carried to the domain pass.** The prebiotic-abundance rider (Remark 6.13) and the mitochondrial classification of the stops (Remark 7.7) were raised at Pass 0, are withdrawn in the text, and are recorded here as ledger items 22 and 25 with *ver none*: the repairs are deletions and no replacement claim is made about any organism.

## B.6 Self-identified defect ledger

Thirty-four items, in the crit/rev/ver form of [Sch26b, Def. 3]. Items 1–31 were raised at Pass 0, against the plan of this paper and against its table file, before any section was drafted, and are recorded as such and not retrospectively presented as design decisions. Items 32–34 arose during drafting. The full Pass 0 text of items 1–31 is retained in the review record; the entries below are their disposed form.

| #     | crit( <i>R</i> )                                                                                                                                                | rev( <i>R</i> ) / ver( <i>R</i> )                                                                                                     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Nine tables described as compiled from the predecessor                                                                                                          | Fully accepted; they are new results of this paper. Ver explicit: B.1                                                                 |
| 2–5   | The $N = 3$ coincidence presented as support; stated as new; glossed as a count of chemical environments; offered as a checkable condition on $H_{\text{dual}}$ | Fully accepted; demoted, imported, glossed withdrawn, condition withdrawn. Ver explicit: Section 9, Remark 3.3                        |
| 6–8   | §5 claimed derivation; the division of selective work unstated; no non-implication guarded the filters                                                          | Fully accepted; §5 reframed as selection, both figures reported, non-implication added. Ver explicit: Remark 5.2, Proposition 3.1(ii) |
| 9–10  | “Five combined filters” overstates; robustness unreported                                                                                                       | Fully accepted; three discriminate, and both jitter figures are published. Ver explicit: Proposition 5.13                             |
| 11    | Uniqueness of the maximiser offered as discriminating                                                                                                           | Fully accepted; relocated to existence and quantified at 70.6% and 25.9%, with 128 the median. Ver explicit: Proposition 3.1(vi)      |
| 12–13 | “Unique minimiser”; fourteen assignments said locked                                                                                                            | Fully accepted; maximiser, and five locked. Ver explicit: Proposition 5.9                                                             |
| 14    | Two $ D $ values contradict the paper’s own Table 1                                                                                                             | Fully accepted; rows corrected to $-0.197$ and $+0.280$ ; the errors cancel in all 128 totals. Ver explicit: Table 8                  |
| 15–17 | §6 claimed derivation; $\tau_1$ to be defined intrinsically; no non-implication guarded it                                                                      | Fully accepted; transformation equivalence, exhibition, non-implication added. Ver explicit: Remark 6.8                               |
| 18–19 | Step 2 understated by two blocks; “exactly two steps” unjustified                                                                                               | Fully accepted; six blocks plus three exceptions; “exactly” withdrawn. Ver explicit: Definition 6.6, Remark 6.10                      |
| 20    | Anchor uniqueness offered as motivating the block assignment                                                                                                    | Fully accepted; demoted to a remark, the anchor being forced. Ver explicit: Remark 6.9                                                |
| 21–22 | The three largest gains misnamed; prebiotic rider                                                                                                               | Fully accepted; corrected to Ser and Ala; rider withdrawn. Ver explicit item 21: Proposition 6.11; ver none item 22                   |
| 23–25 | M5 labelled a theorem; the 2+1 partition; the mitochondrial classification                                                                                      | Fully accepted; relabelled, partition withdrawn, classification withdrawn. Ver explicit items 23–24; ver none item 25                 |
| 26    | The stop designation unregistered                                                                                                                               | Fully accepted; $H_{\text{stop}}$ registered. Ver explicit: Hypothesis 2.24                                                           |

| #  | crit( $R$ )                                                                                                                                                                                                | rev( $R$ ) / ver( $R$ )                                                                                                 |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 27 | “Wobble pairing” misnames $\{A, C\} \{G, U\}$                                                                                                                                                              | Fully accepted; amino/keto, IUPAC $M K$ , throughout. Ver explicit: Remark 2.12                                         |
| 28 | $E1$ asserted a four-way determination from a two-valued datum                                                                                                                                             | Fully accepted; restated with the projection reference and registered as $H_{\text{base}}$ . Ver explicit: Remark 2.17  |
| 29 | B3 and the Table 2 caption gave incompatible accounts of the three vessels                                                                                                                                 | Fully accepted; all three are base-producing, and the wall-and-spacer justification is withdrawn. Ver explicit: Table 2 |
| 30 | $E2$ placed the attack ambiguously at C11 or C12                                                                                                                                                           | Fully accepted; clarified, with the cipher-pair relation recorded and unused. Ver explicit: Remark 8.14                 |
| 31 | $E2$ ’s L-configuration clause asserts assembly-level chirality                                                                                                                                            | Fully accepted; confined to an [E] remark and cross-referenced. Ver none; domain stage                                  |
| 32 | Raised at drafting: Table 3 described the $W S$ obstruction as serine’s split box and named two unmatched amino acids; no amino acid is unmatched in every maximum matching, so the phrase names no object | Fully accepted; replaced by the Hall violators. Ver explicit: Proposition 5.5, Remark 5.6                               |
| 33 | Raised at drafting: $\tau_2$ was to be intrinsic; its shape is, its winner-selection is not, and one block is filled from outside                                                                          | Fully accepted; $\tau_2$ given by exhibition and Proposition 3.1(iii) broadened to cover it. Ver explicit: Remark 6.8   |
| 34 | Raised at drafting: the size ranks of Met, Ile and Leu rest on an arbitrary tie-break entering a 0.17% margin                                                                                              | Fully accepted; disclosed, and shown immaterial over all six orderings. Ver explicit: Remark 4.13                       |

## B.7 Verification record

Nineteen checks. Each was recomputed from the definitions — the marker alphabet, the frame group, the registry sequence, the filter formulas and the standard codon table — by a supplementary script that imports nothing from the tables of this paper, so that agreement is evidence and not a restatement, in the manner of [Sch26e, Rem. B.6].

- V1** Maximum matchings 20, 19, 18 for  $M|K$ ,  $R|Y$ ,  $W|S$ , over all twenty-four letter assignments; constant on each family of eight.
- V2** Degeneracy profiles of both tables; codon gains and losses.
- V3** Block contents of Table 4; four-distinctness in fifteen of sixteen blocks; the seven Step-1 blocks and their unique anchors.
- V4**  $|D|$  recomputed for all twenty classes; two disagreements with the circulated table; the shifts cancel in all 128 totals.
- V5** Permanent of  $\mathcal{A}$  equals 128; five classes locked.
- V6** Table 8 reproduced to three decimals in all twenty rows from Definition 4.5.
- V7** Rank-one unique at 14.4194 against 14.3950; agrees with Table 1 at all twenty classes.
- V8** Unconstrained optimum over all  $20!$  bijections: 21.0794 against 18.5639; twenty disagreements; adjacency violated at sixteen classes, including class 4.
- V9** Leave-one-out and weight jitter, as Proposition 5.13.
- V10** Table 6 agrees with the standard code on all sixty-four codons.
- V11** Twelve block-fitness variants; none returns the  $\tau_1$  blocks or targets; zero of seven on the necessary condition.
- V12** Linear-programme infeasibility over filter weights, non-negative and signed; the second-base collapse of Sz and HB under block aggregation.
- V13** Random codes of the standard profile: matching exists in 3528 of 5000.
- V14** All 13 824 position-wise base relabellings: matching exists in 3584; among those admitting one, the number of matchings has median 128.
- V15** All six orderings of the Met/Ile/Leu size ranks: same maximiser, same margin.
- V16** Hall violators for  $W|S$  and  $R|Y$ ; no amino acid unmatched in every maximum matching.
- V17** After  $\tau_1$ , exactly seven amino acids on one codon each.
- V18** Five candidate rules for the  $\tau_2$  half-winners; best ten of thirteen; UC unreachable by any rule.
- V19**  $2^{2N-2} + 2^{N-1}$  against  $\binom{N+3}{3}$  for  $N = 1, \dots, 20$ .

## B.8 Meta-review

**Remark B.1** (Disposition). Of the thirty-four entries of B.6, thirty-four are disposed fully accepted; none is partially resolved, unresolved, or no longer applicable. Four carry *ver none* (items 22, 25, 31, and the second half of 21) and are referred to the domain stage; in each the repair is a deletion, so what verification would confirm is an absence.

**Remark B.2** (Three observations). First, of the thirty-four items, twenty-five are over-claims in the manuscript's own direction — a hypothesis asserted to do more than it does, a scope asserted wider than the argument, a datum cited rather than recomputed, a coincidence offered as evidence. That is the asymmetry [Sch26b, Rem. 1] reports and [Sch26e, Rem. B.5] reproduces, now recorded on a third manuscript, and it is the reason the review stages are assigned disjoint questions.

Second, the two items with the largest consequence, 6 and 16, were raised before drafting began and are not defects of proof: each is a claim about what a construction establishes rather than about whether a step is valid. Neither would have been produced by checking the mathematics, all of which is correct as computed; both required asking what the computation was a computation *of*. This is the differentiation of correction type by review discipline that [Sch26b, Rem. 1] reports, appearing here at the pre-drafting pass rather than at concordance.

Third, no criticism at any stage required a result to be withdrawn. The repairs narrow claims, supply witnesses, correct two table rows, rename a partition and remove three biological assertions; every theorem stated at Pass 0 survives in the text, with the single exception of the two-step minimality claim, which was not a theorem but an adjective.

**Remark B.3** (The limitation of the verification recorded here). Targeted verification for this manuscript re-runs the mathematical stage rather than employing an independent reviewing system. By [Sch26b, §9.4], independence of pass is not independence of reviewing system, and every entry marked *ver explicit* in B.6 inherits that limitation and should be discounted accordingly. The concordance and domain stages are not affected, being assigned distinct questions in the sense of [Sch26b, §4.3]. The nineteen numerical verifications of B.7 are of a different kind: each is an exhaustive enumeration or a closed computation over a finite set, reproducible from the statement of the object enumerated without reference to any reviewing system.

What the review does not certify is stated with the same plainness. Nine Targets are open, three of them load-bearing by Remark 10.1.  $H_{\text{dual}}$  in all its clauses,  $H_{\text{crys}}$ , and  $H_{\text{stop}}$  are registered and undischarged. No synthesis is reported. And the selection of Section 5 rests on a margin of 0.17%, which Proposition 5.13 shows to survive a twenty per cent perturbation of the weights and to fail in nearly three trials in ten at fifty per cent.

## C Established Mathematical and Domain Literature

As in the literature appendices of [Sch26c, Sch26b, Sch26d, Sch26e], none of the subjects below is claimed to be novel in isolation. The purpose is to identify the established machinery each part of this paper draws upon, and to keep it visibly distinct from the specific organisation proposed here. The mathematics used is elementary; that is worth recording, since the paper's claims rest on finite computations any reader can repeat rather than on machinery that must be taken on trust.

### C.1 Group actions and orbit counting

The orbit-counting lemma underlying the class count imported from [Sch26e] is classical and is stated in [Bur11, §119]; a modern treatment together with the orbit–stabiliser theorem is [Rot95, Ch. 3]. The attribution to Burnside is conventional and inaccurate, the result being due to Cauchy and Frobenius; the name is retained because the corpus will be searched under it.

### C.2 Bipartite matching, Hall's theorem, and permanents

Theorem 5.4 and Proposition 5.5 rest on Hall's marriage theorem [Hal35], in the deficiency form: a bipartite graph has a matching saturating one side unless some subset of that side has a smaller neighbourhood, and the maximum matching falls short by the largest such deficiency. The standard modern treatment, including the deficiency version and the König–Ore formula, is [LP86, Ch. 1–2]. The count of Proposition 5.8 is the permanent of a zero-one matrix; the standard reference is [Min78]. The assignment problem solved in the witness of Proposition 3.1(ii) — maximising a total over all bijections of two twenty-element sets — is [Kuh55]. None of these is used beyond its textbook form: the matrices here are  $20 \times 20$  with at most four entries per row, and every computation reported is exhaustive rather than algorithmic.

### C.3 Elementary enumeration

The multiset count of Section 9 is the standard stars-and-bars formula [Sta12, Ch. 1]. The invariance of the multiset of fibre sizes of a surjection under bijections of source and target, on which Proposition 3.1(iv) turns, is immediate from the definition and is recorded only because the proposition rests on it.

### C.4 Torsors

The notion of a torsor under a finite abelian group, inherited here from [Sch26d, Prop. 8.7] and [Sch26e, Prop. 5.3] and used in Proposition 2.15, is standard; see [Bou70, Ch. I, §5] and, for the finite case, [Ser79, Ch. I]. The property used is that a

torsor has no canonical base point, so that a statement about its elements must be invariant under the acting group or must name the choice that breaks the symmetry.

### C.5 The genetic code and nucleotide nomenclature

The code itself is cited to [Cri68]; its degeneracy profile was recomputed from the standard codon table at the domain pass and not read from the citation, in the manner of [Sch26e, Rem. 3.16]. The IUPAC one-letter ambiguity classes used in Remark 2.12 —  $W$  and  $S$ ,  $R$  and  $Y$ ,  $M$  and  $K$  — are [NCIUB85]; the identification of  $M|K$  as amino against keto is that document's own, and the observation that these three exhaust the partitions of a four-letter alphabet into two classes of two is arithmetic. The correlation of four-fold family boxes with the composition of the first two codon positions is [Lag78], cited for context at Remark 6.14 and used in no proof. Variant and organellar codes, relevant only to the withdrawal recorded at Remark 7.7, are surveyed in [OJW92] and [KFL01].

### C.6 Amino-acid property data

The residue volumes underlying the size ranks of Section 4 are [Zam72]; the hydrophathy values are [KD82]. The donor and acceptor counts, the soft-Lewis-base scores and the rigidity scores are tabulated in Section 4 rather than cited, being conventions of this paper; their physical reading is  $H_{\text{crys}}$  and is used in no proof. The prebiotic inventories against which the withdrawn rider of Remark 6.13 would have to be tested are [Mil53] and [CP97], with the consensus chronology of [Tri00]; none is used here, and Target 6.19 states what using them would require.

### C.7 What is drawn from the Chrysene Formalism itself

Axioms A1–A12, the status vocabulary, the non-vacuity theorem and the acyclicity of the axiom ladder are [Sch26c]. The encoded stack, the alphabet  $\Sigma_0$ , the structural group, the registry invariant, the frame and its group, and the maximal-frame-invariant theorem are [Sch26d]. The classification of marker assignments into frame classes, the descent theorem, the structure of the induced action, and the variational criterion whose weight and selection hypothesis this paper instantiates are [Sch26e]. The review architecture, the editorial classification and the non-implication conventions are [Sch26b]. The survival programme, to which nothing here bears, is [Sch26a]. None of these is external literature in the sense of this appendix; they are listed so that the boundary between established machinery and the formalism's own claims is visible in one place.

## References

- [Bou70] Nicolas Bourbaki. *Algebra I: Chapters 1–3*. Hermann / Addison-Wesley, 1970.
- [Bur11] William Burnside. *Theory of Groups of Finite Order*. Cambridge University Press, 2nd edition, 1911.
- [CP97] John R. Cronin and Sandra Pizzarello. Enantiomeric excesses in meteoritic amino acids. *Science*, 275(5302):951–955, 1997.
- [Cri68] Francis H. C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38(3):367–379, 1968.
- [Hal35] Philip Hall. On representatives of subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935.
- [KD82] Jack Kyte and Russell F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, 1982.
- [KFL01] Robin D. Knight, Stephen J. Freeland, and Laura F. Landweber. Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, 2(1):49–58, 2001.
- [Kuh55] Harold W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.
- [Lag78] Ulf Lagerkvist. “Two out of three”: an alternative method for codon reading. *Proceedings of the National Academy of Sciences*, 75(4):1759–1762, 1978.
- [LP86] László Lovász and Michael D. Plummer. *Matching Theory*, volume 121 of *North-Holland Mathematics Studies*. North-Holland, Amsterdam, 1986.
- [Mil53] Stanley L. Miller. A production of amino acids under possible primitive Earth conditions. *Science*, 117(3046):528–529, 1953.
- [Min78] Henryk Minc. *Permanents*, volume 6 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Reading, MA, 1978.
- [NCIUB85] Nomenclature Committee of the International Union of Biochemistry. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *European Journal of Biochemistry*, 150(1):1–5, 1985.
- [OJW92] Syozo Osawa, Thomas H. Jukes, Kimitsuna Watanabe, and Akira Muto. Recent evidence for evolution of the genetic code. *Microbiological Reviews*, 56(1):229–264, 1992.
- [Rot95] Joseph J. Rotman. *An Introduction to the Theory of Groups*, volume 148 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 4th edition, 1995.
- [Sch25a] Charles D. Schaper. Encoded design integrated synthesis of nucleic acids, and phospholipids, and related pharmaceutical products, 2025. US Patent 12,195,423.
- [Sch25b] Charles D. Schaper. Encoded design integrated synthesis of nucleic acids, and phospholipids, and related pharmaceutical products, 2025. US Patent 12,473,247.
- [Sch26a] Charles D. Schaper. Stable Transfer Geometry of the Substituted Chrysene Network:  $H^\infty$  Transformation Metrics, Dissipative Well-Posedness, and Survival Under Persistent Disturbance. *Annals of the Chrysene Formalism*, 1(2):603–645, 2026.
- [Sch26b] Charles D. Schaper. The Chrysene Formalism as Candidate Mathematics and Candidate Biogenesis: A Quantitative and Categorical Survey of Volume 1, Number 1. *Annals of the Chrysene Formalism*, 1(2):476–488, 2026.
- [Sch26c] Charles D. Schaper. The Chrysene Formalism Revisited: A Twelve-Axiom Foundation for Molecular Characterization, Tensorial Noncommutative Geometry, and Operator Response. *Annals of the Chrysene Formalism*, 1(2): 489–560, 2026. see [Sch26e, Rem. B.3].
- [Sch26d] Charles D. Schaper. The Encoded Stack: A Frame-Indexed Base-Four Structure on the  $\mathbb{Z}^3$ -Fibrated Chrysene Network, its Sequence Algebra, and the Set-Up for Survival. *Annals of the Chrysene Formalism*, 1(2):646–675, 2026.
- [Sch26e] Charles D. Schaper. The Twenty Registry Classes of a Chrysene Stack: Frame Orbits, Descent of a Letter Transformation, and a Variational Criterion. *Annals of the Chrysene Formalism*, 1(2):676–704, 2026.
- [Ser79] Jean-Pierre Serre. *Local Fields*, volume 67 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1979.

- [Sta12] Richard P. Stanley. *Enumerative Combinatorics, Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2nd edition, 2012.
- [Tri00] Edward N. Trifonov. Consensus temporal order of amino acids and evolution of the triplet code. *Gene*, 261(1):139–151, 2000.
- [Zam72] Alexander A. Zamyatnin. Protein volume in solution. *Progress in Biophysics and Molecular Biology*, 24:107–123, 1972.