FEATURE ARTICLE

The T3 Method ("T3 It!"): Transcribe → Translate → Transform

BRETT MALAS

Abstract

Traditional transcription-translation exercises are instructionally incomplete by failing to link prescriptive genetic information with protein structure and function. The T3 Method solves this problem by adding a conceptually powerful yet easily learned third step where students use simple protein folding codes to transform *their translations into corresponding protein structural models. This brings structural sense to sequence and makes the information-to-proteins connection that is so profoundly important to understand in biology more directly evident, experiential, and intrinsically meaningful. The T3 Method has further utility, proving versatile and adaptive to a wide range of academic levels and learning contexts, with possibilities for differentiated instruction, application, and extension.*

Key Words: *Central Dogma; protein synthesis; transcription; translation; protein folding; protein structure and function; genetics; mutation.*

The Problem

Translations should enlighten, yet here instruction falls short for arguably the most fundamental, life-defining process: genetically prescribed protein synthesis. Traditional transcription-translation exercises are commonly used to teach biology's "Central Dogma." Unfortunately, these exercises are little different from having students translate Greek into Latin while knowing neither and being told their results are meaningful. Without a way for students to make sense of their translations, the results—e.g., "Met-Ala-Lys . . . "—are effectively meaningless. This leaves instructional goals unrealized and learning incomplete.

Translations should enlighten, yet here instruction falls short for arguably the most fundamental, life-defining process: genetically prescribed protein synthesis.

Educationally, this is no small brush-aside. Cognitive science is definitive that "Knowledge must have meaning for learning to occur" (Willingham, 2009). But like 3D-images encoded in 2D-autostereograms, we mean for students to see the 3D-structure specified by 1D-sequence, but they only see the latter and it means very little to them. The greater conceptual point is literally lost in translation.

This is no disparagement of teachers but reflects the complexities of protein folding. Protein structure prediction accuracy has improved substantially from 50 percent in the 1970s to over 80 percent today (e.g., Chen et al., 2006; Heffernan et al., 2015). However, the immense computational resources needed to achieve this, including sophisticated computer modeling with machine-learning artificial neural networks, is telling. Student-friendly modeling technology is helpful, but still comparable to "learning" a foreign language via Google Translate®. Lacking even a rudimentary command of protein folding language for themselves, student understanding remains limited, teacher dependent—because we or a computer say so—and therefore still on loan.

But what if biology teachers could equip their students with the tools to "transform" sequence into structure? Students would then have the means to independently verify and experience firsthand this important conceptual understanding on their own, versus the far less gratifying experience of simply believing, on another's word, that their work is significant even if they can't see it for themselves.

The Solution

After fifteen years of searching, I devised a solution: use simple protein folding codes to make structural sense of sequence. Globular protein structure is largely specified by the linear order of polar and nonpolar amino acids (Dill et al., 1995). This discovery simplifies the twenty-letter amino acid alphabet to two, and enables polypeptide sequence recoding. The binary patterns that emerge correlate with specific

The American Biology Teacher, Vol. 79, No 4, pages. 257–271, ISSN 0002-7685, electronic ISSN 1938-4211. DOI: https://doi.org/10.1525/abt.2017.79.4.257.

secondary structures and have reliable predictive value that is used to successfully design novel proteins *de novo* (e.g., Kamtekar et al., 1993). Instructionally, they're the Rosetta Stone equivalent needed to decode primary sequence and bridge the gap between the world of proteins and biological information (Figure 1).

○ The T3 Method ("T3 It!"): Transcribe → Translate → Transform

The T3 Method brings structural sense to sequence by leveraging simple protein folding codes to link genetic information with the protein structure-function it prescribes. T3 extends and conceptually completes traditional two-step transcription-translation problems by taking them to their logical end with the addition of a third: "Transform" (Figure 2). In this step, students: (1) convert amino acid sequences from translation into binary form; (2) identify binary patterns that specify secondary structures; and (3) draw representative protein picture models. Drawing is important for improving and assessing student understanding of protein structure (Harle & Towns, 2013).

The T3 Method is adaptive to a wide range of academic levels. It is easily differentiated and scaffolded to student ability. T3 is suited for use in introductory high school biology to college level courses like general biology and genetics. I've successfully used the method with regular to honors level high school biology students. In the process, I've developed and refined the method across four levels of difficulty: basic, standard, intermediate, and advanced. These broadly align with early secondary to college level instruction but not inflexibly, so that teachers may use and adapt the method to their own instructional contexts.

Standard Level: Two-Structure Option (α-helices and β-strands)

The "Big" or "Elite 7" amino acids VILFMYW (mnemonic: "President 'W.FLIMVY'") are most associated with hydrophobic cores and

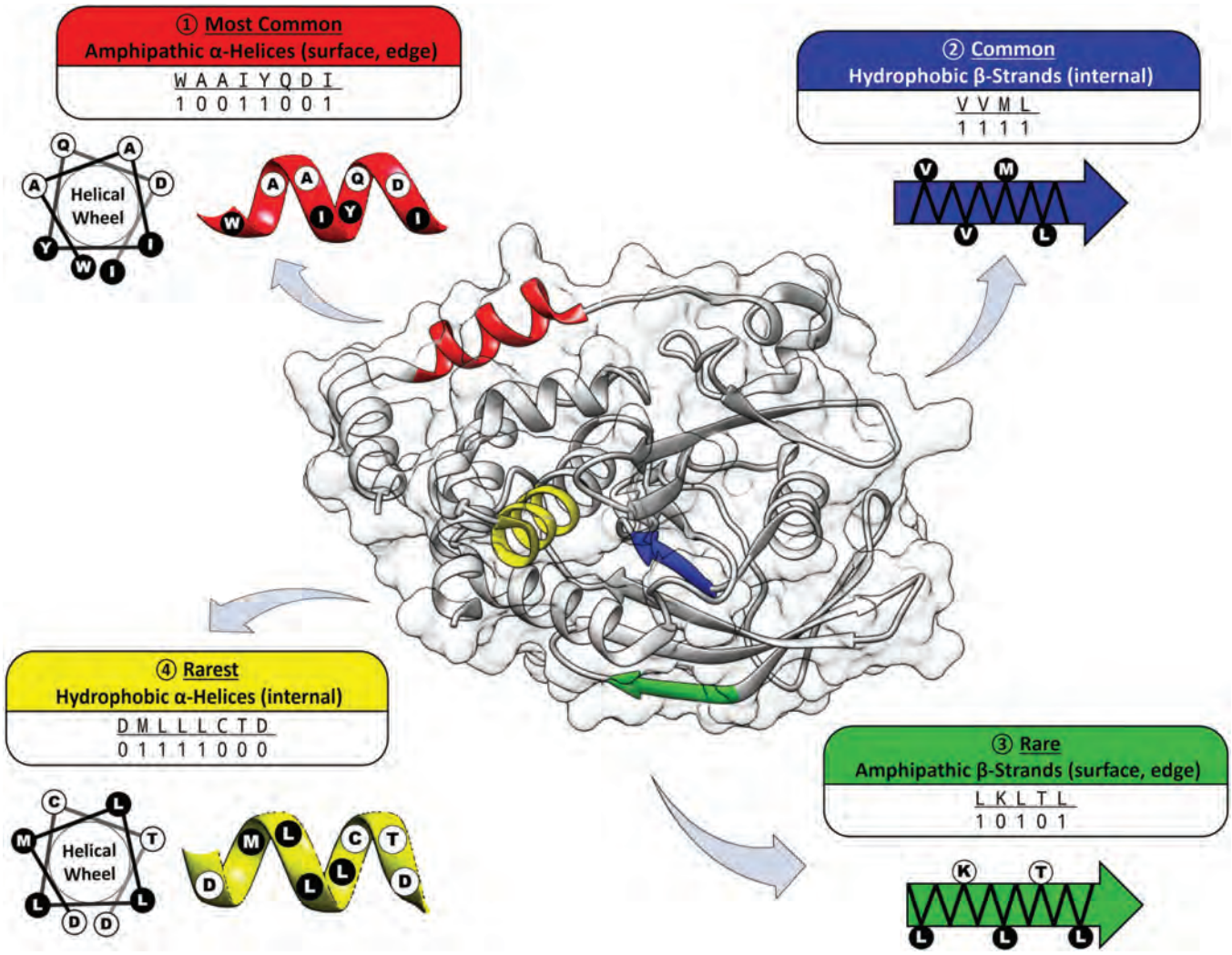


Figure 1. "Transforming" sequence into structure. Protein structure is largely encoded by the linear binary sequence of polar and nonpolar amino acids in polypeptides (Trifonov, 2008). Polar amino acids are represented by "0s" and white circles; nonpolar by "1s" and black circles. Secondary structures 1–3 are strongly associated with the binary sequence "folding codes" shown. The image of protein-tyrosine phosphatase 1B (PDB ID: 3A5J; Iwamoto et al., Forthcoming) was created with UCSF Chimera (Pettersen et al., 2004).

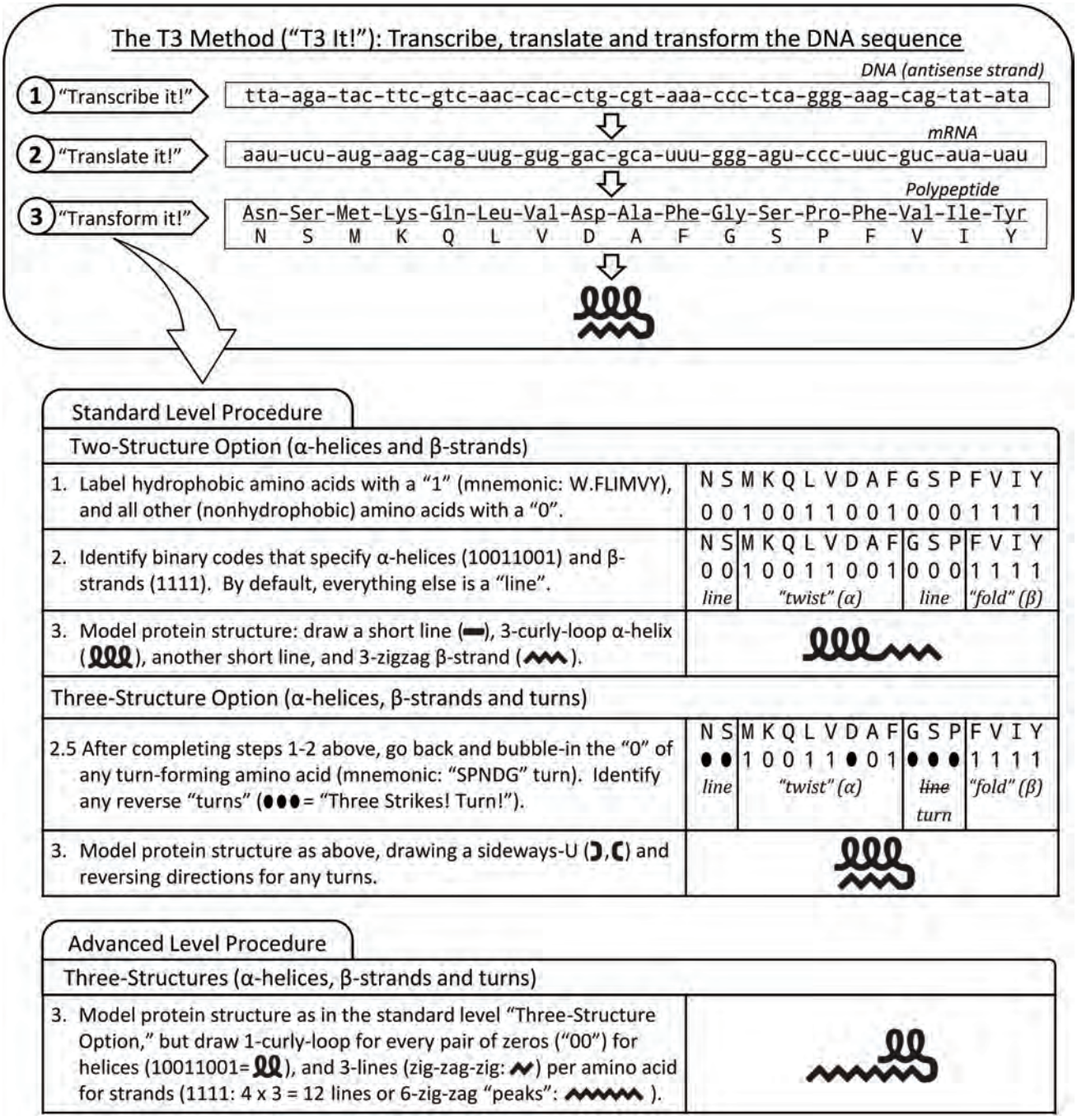


Figure 2. The T3 Method: sample problem with standard and advanced level solutions.

regular secondary structures (helices and strands). They form the hydrophobic alphabet used for binary conversion in Hydrophobic Cluster Analysis (HCA) (Callebaut et al., 1997) and are here adopted for the same (Figure 3). Because of their primary, "presidential" importance in protein structure, they're always number "1." All other amino acids are "0" (e.g., MQLV = 1011 in binary code).

Most secondary structures are α-helices, and most globular protein helices are on the solvent-surface and amphipathic (two-sided), with inward-/outward-facing hydrophobic/hydrophilic sides (Figure 1.1). One complete helical turn occurs every 3.6 amino acids, creating an unmistakable hydrophobic-nonhydrophobic binary pattern that alternates every 3-4 residues: 10011001. Dubbed the "Alternating Twist" or "Double-'o-Sandwich" folding code , students draw the α-helix symbol (3-curly loops:) whenever this code is encountered (Figure 4).

In *de novo* design, the repeating "10011001 . . ." and "1010 . . ." are common sequence motifs used to construct amphipathic helices and strands, respectively (e.g., Bradley et al., 2006). In nature, however, most globular strands are not amphipathic (Figure 1.3) but hydrophobic and internally buried (Figure 1.2), often as part of a protein's stabilizing hydrophobic core (West & Hecht, 1995; Broome & Hecht, 2000). The hydrophobic cluster "1111" is a common, representative example (Eudes et al., 2007). Students model this "Four-Fold" or "Exclusive Club" folding code by drawing the β-strand symbol (3-zigzags:) whenever it is encountered.

Amino Acid Alphabets for Binary Code Conversion		
Hydrophobic Alphabet ("The Big/Elite 7")[a]	1 = VILFMYW	Mnemonic: "President 'W.FLIMVY'"
Nonhydrophobic Alphabet ("The Rest")	0 = Everything else (ARCQTEKHPGDNS)	Mnemonic: N/A
"Turn" Alphabet ("The Deal Breakers")	● = PGDNS	Mnemonic: "'SPNDG' Turn"
[a]T3 follows the hydrophobic alphabet used in Hydrophobic Cluster Analysis (HCA).		

Figure 3. Amino acid alphabets for binary code conversion. Hydrophobic classification depends on the parameters used to measure hydrophobicity. Simm et al. (2016) identified 98 different "hydrophobicity scales." T3 adopts the hydrophobic alphabet "VILFMYW" used in Hydrophobic Cluster Analysis (HCA). Callebaut et al. (1997) explain the rationale for this classification.

Standard Level Reference Guide (exact match codes and symbols)			
Structure	"Twists" (α-helices)	"Folds" (β-strands)	"Turns" (γ-turns)
Folding Code[a]	"Alternating twist/double-'o-sandwich" 10011001	"Four-fold/exclusive club" 1111	"Three strikes! Turn!" ●●●
Modeling Symbol	Draw: 3-curly loops	Draw: 3-zigzags	Draw: sideways-U Ɔ,C
Orientation	Up/nonspecific	Up/nonspecific	Left/right Ɔ,C
Code Duplication	10011001**1**0011001 (iterative)[b]	11111111 = 1111 + 1111	●●●●●● = ●●● + ●●● Ɔ + C = Ↄ
[a]"Folding codes" correspond to "hydrophobic clusters" identified as "canonical" in Hydrophobic Cluster Analysis (HCA) for their high statistical correlation with α-helices (10011001) and β-strands (1111) (Eudes et al., 2007). [b]α-helix doubling doesn't work (i.e. destroys alternating pattern "10011001 + 10011001" = "1001100**1**10011001"). Students must be taught the pattern is iterative with the middle "1" shared (10011001**1**0011001 = 10011001 + 10011001).			

Figure 4. Standard level reference guide.

Using just these two simple folding codes (10011001 and 1111), a variety of structures can be modeled. In the process, students directly link sequence and structure in a simple yet visually powerful and confirming way that is both immediate and reinforcing with each new problem, e.g.:

- β-strand: cagtataacaag → gucauauuguuc → VILF → 1111 → ∿∿
- α-helix: 10011001 → ⌇⌇⌇
- α-β: 100110011111 → ⌇⌇⌇∿∿
- β-α: 111110011001 → ∿∿⌇⌇⌇
- β-α-β: 1111100110011111 → ∿∿⌇⌇⌇∿∿

and so-on. Simple, yet effective.

Codes can also be duplicated in series (Figure 4). For β-strands this is straightforward: e.g., β-β: 11111111 = 1111 + 1111 = ∿∿ + ∿∿ = ∿∿∿∿. For α-helices, students must be taught that the code is iterative with the middle "1" shared: e.g., α-α: 10011001**1**0011001 = ⌇⌇⌇ + ⌇⌇⌇ = ⌇⌇⌇⌇⌇⌇.

Standard Level: Three-Structure Option (α-helices, β-strands and turns)

The range of possibilities is greatly increased by incorporating reverse turns. Opposite the "Elite 7" are the five "Deal Breakers," PGDNS (mnemonic: "'SPNDG' Turn"; Figure 3). These amino acids are least associated with regular secondary structures and, in fact, routinely break up helices and strands by forming turns and loops (Eudes et al., 2007). Students bubble-in SPNDG zeros (" 0 " → " ● ") to form a ternary code. Four residue β-turns are most common, but three-residue γ-turns were chosen to avoid confusion with the "Four-Fold" β-strand code. Students then learn the "Three Strikes! Turn!" Rule (Figure 4): Whenever three consecutive bubbled-"0s" occur ("●●●"), reverse direction by drawing a sideways-U (Ɔ, C).

At most then, the standard level has only three modeling symbols with exact-match folding codes for students to learn: 10011001 = ⌇⌇⌇, 1111 = ∿∿, and ●●● = Ɔ, C. By default, anything else is modeled with a line (*l*): e.g., 111 = —. For an example of "two-" and "three-structure option" solutions for the same problem, see Figure 2.

Advanced Level

The advanced level uses the same alphabets for binary conversion (Figure 3), but with pattern-based folding codes and improved modeling accuracy (Figure 5). Codes have form and length requirements:

- "Ones" Rule: All folding codes must begin and end with "1" except turns (●●●) and optional amphipathic β-strands. The latter must still begin with "1" but may end with "1" or "0" (e.g., 1010, 10101).

- "5-4-3" Minimum-Length Rule: Folding codes must be at least 5 amino acids long for α-helices (10011, 11001); 4 for β-strands (1111, 1010); and 3 for turns (●●●).

These minimum-length codes are building blocks for larger patterns. For β-strands, 1111 expands to 11111, 111111, etc. (or 1010 to 10101, 101010, etc., for amphipathic β-strands). For α-helices, 10011 and 11001 are the two halves of 10011001. These rearrange to give 110010011 or expand indefinitely (e.g., "100110010011 . . ."). Turns expand to include both three- and four-residue γ- and β-turns, respectively (●●●, ●●●●).

Modeling is not constrained by fixed symbols but is informed by sequence (Figure 5). Students must correctly "read" and interpret binary sequences to accurately model:

- α-helix "right-handedness": Upward-facing when drawn from left to right, and downward-facing when drawn from right to left.
- α-helix turn number: Helical turns occur at regular, predictable intervals every 3.6 amino acids (3.6, 7.2, 10.8, 14.4, 18, . . . , etc.) and are approximated by drawing one loop for every pair of zeros that occurs in a folding code: e.g., 11001 = , 10011001 = , etc.
- β-strand length: One amino acid occurs every "zig-zag-zig" () of a β-strand's accordion-like backbone, so multiplying the number of amino acids in a folding code by three gives the correct number of lines to draw: e.g., "1111" = (4 amino acids × 3 = 12 lines or 6 zig-zag "peaks");
- Antiparallel β-sheets: Inverted L-turns (,) connect and offset neighboring β-strands to model antiparallel orientation ().

For examples of standard vs. advanced level solutions for the same problem, see Figures 2 and 6.

Linking Sequence and Structure

No claim is made that the T3 Method enables students to derive the complete three-dimensional conformation of complex proteins from primary sequence alone. Even our most sophisticated *de novo* and *ab initio* prediction methods cannot achieve this. The goal is conceptual: link sequence and structure. T3 accomplishes this. Students immediately see from their drawings that

Advanced Level Reference Guide (pattern-based codes and symbols)

Structure	"Twists" (α-helices)	"Folds" (β-strands)	"Turns" (γ/β-turns)
Folding Code	"Alternating twist/double-'o-sandwich" "10011001..."	"Four-or-more-fold"; "Every-other-one" "1111..."; "1010..." (optional)[a]	"Three-or-four strikes! Turn!" ●●●, ●●●●
Modeling Symbol	Draw: 1-loop ( ) per zero pair[b] 10011, 11001 = 10011001, 110010011 = 100110010011, 110010011001 = 100110010011001, etc. =	Draw: 3-lines (zig-zag-zig) per aa[c] 1111, 1010 = 11111, 10101 = 111111, 101010 = 1111111, etc. =	Draw: sideways-U/inverted-L , (β-sheets)[d]
Orientation	Up/down (right-handed helix)[e]	Up (but offset for β-sheets)[d]	Left/right
"Ones" Rule (α/β)	All codes must begin/end with "1"	Codes must begin/end with "1," except optional amphipathic strands, which can end with "1" or "0" (1010, 10101)	N/A
"5-4-3" Rule	Minimum length of α-helices: 5 10011, 11001	Minimum length of β-strands: 4 1111, 1010	Minimum length of turns: 3 ●●●

Miscellaneous Modeling

Default Line	By default, any undefined binary sequence that does not code for a structure is modeled by a line, approximately the same length as the undefined sequence: 0111100011110000
Parallel β-Sheets	Loop "up," "down," "over & around," and "under & around" symbols can be used to model parallel β-sheets. = =
Other	Symbols can be added as needed or desired: • S3— —S4: "Loop over and around between structures 3 and 4." • S1—147aa—S2: "Add 147aa between structures 1 and 2 in your model." 147aa

[a]Optional amphipathic β-strands (Figure 1.3) with alternating "1s" and "0s" ("Every-other-one": "1010...").
[b]All of the following are considered zero-pairs: "0 0," "● 0," "0 ●," "● ●".
[c]3-lines (zig-zag-zig) or 1½ zig-zag "peaks" per amino acid (e.g. 1111 = 4 x 3 = 12 lines or 6 "peaks":).
[d]Inverted-L turns connect and offset adjacent β-strands to model antiparallel β-sheets: ⇨ ⇨
[e]Up when drawn L to R; down when drawn R to L to model right-handedness.

Figure 5. Advanced level reference guide.

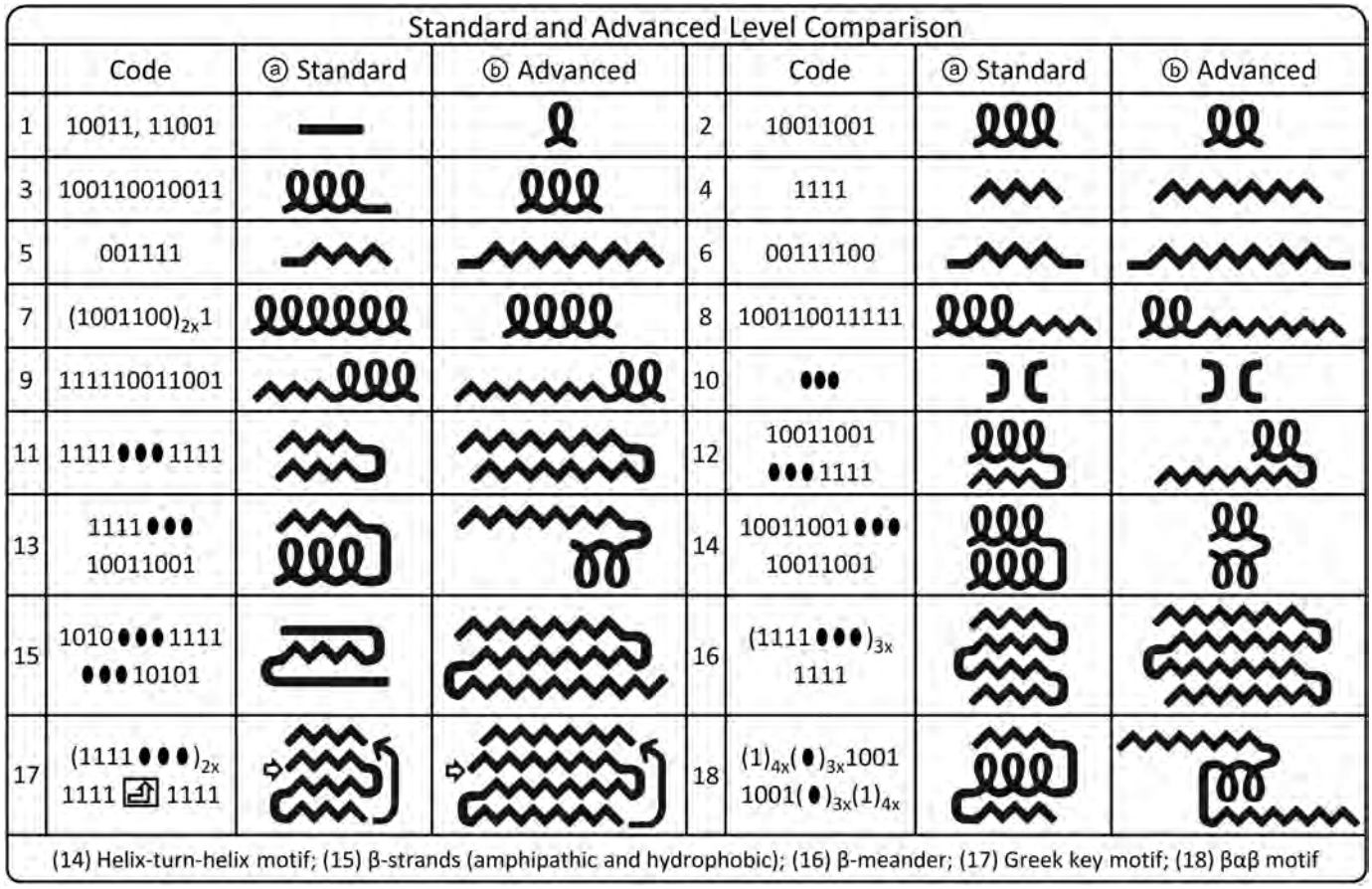


Figure 6. Standard (a) and advanced level (b) comparison chart.

different nucleotide and polypeptide sequences are not the meaningless, random jumble of letters they appear to be, but carry information and specify different structural outcomes.

Correcting Misconceptions

The T3 Method has further utility. T3 not only links sequence and structure, it gives students a deeper, more accurate understanding of how the former specifies the latter. Traditionally the easier, more obvious aspects of protein structure are taught to students (e.g., +/– attraction). This is understandable but misleading. Electrically charged amino acids are too small in number and effect to play a determinative role (Dill, 1990). Similarly, the amide-carbonyl hydrogen bonds that characterize secondary structure are too generic to explain secondary structure formation. Any amino acid can form these bonds, but only specific polar-nonpolar binary patterns contain enough information to specify secondary structure generally.

Instruction that presents the different "levels" of protein structure as a hierarchy also misleads by reinforcing the sequential, step-wise view of protein folding posited by the older framework or hierarchical model: primary → secondary → intermediate form (s) → tertiary (Ptitsyn, 1973). However, empirical support has grown for the hydrophobic collapse model, where secondary and tertiary structure form concurrently instead of sequentially (Dill et al., 1995; Ahluwalia et al., 2012). Current understanding tells us that: (1) the hydrophobic effect is the dominant driving force of globular protein folding; (2) hydrophobic interactions cause protein collapse around an internal hydrophobic core; (3) the hydrophobic core is essential to the stability and maintenance of overall 3D protein structure; (4) secondary and tertiary structure form concurrently as a result of hydrophobic collapse; (5) secondary and tertiary structure is largely encoded in binary form and specified by polar-nonpolar amino acid sequence patterning; and (6) other factors like salt bridges, side chain H-bonding, and disulfide bridges are important to further stabilize, refine, and "lock-in" tertiary structure, but are not themselves the dominant, determinative cause.

Improving Students' Understanding of Protein Structure and Folding

T3 can be used to teach students this more accurate and up-to-date understanding of protein folding: Students (1) play the "Protein Folding Game" to acquire a basic understanding of protein formation through hydrophobic collapse (Figure 7); (2) use T3 to design "minimal protein models" to reinforce the importance of hydrophobic cores in maintaining protein structure and stability (Figure 8); and (3) design three-dimensional protein models to understand the important supporting roles of other factors in stabilizing, refining, and locking-in protein structure (Figure 9). In doing so, students learn that hydrophobic interactions drive the folding and collapse of proteins into a specific three-dimensional shape that is specified primarily by amino acid sequence information encoded in binary form (nonpolar "1s" and polar "0s").

Figure 7. The Protein Folding Game, simulating globular protein folding and hydrophobic collapse. Students fold the "polypeptide" (pipe cleaner) to maximize the number of polar amino acids (white beads) exposed on the surface and nonpolar amino acids (black beads) buried internally to form a hydrophobic core.

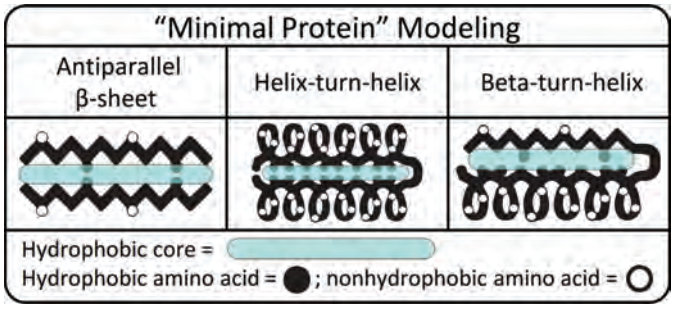


Figure 8. Using "minimal modeling" to teach general protein structure. A globular protein consists of two basic parts or regions: a hydrophilic surface and an internal, stabilizing hydrophobic core. T3 is used to model this with DNA sequences that specify these minimum requirements.

Raising Instruction on Genetic Mutations to an Entirely New Level

After students acquire this improved understanding of proteins, the logical continuation is then to ask how changing the information encoded in amino acid sequences affects protein structure and function. Here, the T3 Method not only supports and extends further inquiry, it raises instruction on genetic mutations to an entirely new level (Figure 10).

Proteins are remarkably tolerant to amino acid substitutions (Bradley et al., 2007), due to hydrophobicity and steric factors (Ladunga & Smith, 1997). Amino acids of comparable size and hydrophobicity can usually substitute for one another without affecting protein structure and function ("1s" for "1s" & "0s" for "0s"). This fact has been exploited by researchers to design proteins "from scratch" (e.g., the binary code *de novo* design strategy of Bradley et al., 2007). Tolerance also varies by location (Bowie et al., 1990). Loops and coils are the least conserved regions in proteins and most prone to amino acid substitutions of any hydrophobicity. Next are substitutions on the protein surface, which are tolerated more than internal ones. For example, amphipathic helices can generally tolerate substitutions as long as they don't involve too many hydrophobic substitutions. Hydrophobic cores, by contrast, are the most highly conserved regions in proteins and most sensitive to substitutions. Hydrophobic substitutions are generally tolerated ("1" for "1"), whereas nonhydrophobic substitutions ("0s" for "1s") are usually highly disruptive to protein structure and function. This reinforces the importance of hydrophobic interactions and hydrophobic cores in protein folding, and in maintaining and stabilizing overall three-dimensional structure.

At its simplest, the general substitution "rule" of polar-for-polar ("0s" for "0s") and nonpolar-for-nonpolar ("1s" for "1s"), provides students a far more accurate and nuanced understanding of genetic mutations beyond what is traditionally taught. It gives students a powerful tool for predicting whether or not substitutions by different amino acids are likely to affect protein structure and function that the T3 Method can be used to model (Figure 10). For example, T3 can be used as an instructional tool to: (1) model the structural effects of mutations from sequence; (2) predict whether a nonsynonymous amino acid substitution will affect protein structure and function; (3) explain the high tolerance of proteins to amino acid substitutions; (4) reinforce the stabilizing importance of hydrophobic interactions and hydrophobic cores in maintaining globular protein structure; (5) analyze and elucidate classic point mutation examples like sickle cell disease; (6) integrate, extend, and connect to other learning activities, like the *American Biology Teacher* "Pencil Transferase" models by Chowning et al. (2012); and more (see Figure 10).

○ Expanded Options

Intermediate Level

The T3 Method also provides teachers with expanded options for delivering instruction. For example, standard and advanced level elements can be combined to create a transitional intermediate level if needed (Figure 11). Teachers might want to use exact-match folding codes from the standard level, but with more detailed, accurate modeling from the advanced (e.g., 1111 =). Conversely, teachers might want to use general folding code patterns from the advanced level, but model them with simple, easy-to-learn symbols from the standard (e.g., 11001 = , 10011001 = , 100110010011 =). The intermediate level provides these options for those who need them.

Basic Level: "T3 Original"

Topics like amino acid hydrophobicity fall outside the scope of some introductory biology courses. The original T3 Method provides a work-around for teachers in this situation (Figure 12). This version is based on amino acid secondary structure propensities and eliminates the need for binary conversion altogether.

Students learn that amino acids have structural "seating preferences" with some preferring "twists" (, mnemonic: "'KHARMEL-Q' Twist"), "folds" (, "WYFI-CTV") or "turns" (, , "'SPNDG' Turn"). The "transform" step is a two-part procedure: (1) Identify clustered groups of amino acids in translated sequences that correlate with specific secondary structures. (2) Draw representative protein picture models using standard level modeling symbols.

This viable, straightforward version of T3 still solidly links sequence and structure with the added advantage of shorter starting sequences. But it has less explanatory power and is more idealized in that secondary structure predictions based solely on amino acid propensities are only 50–60 percent accurate (Chen et al., 2006). Therefore, teachers should view "T3 Original" as an entry point but not a final destination, and strive to work in hydrophobicity if at all

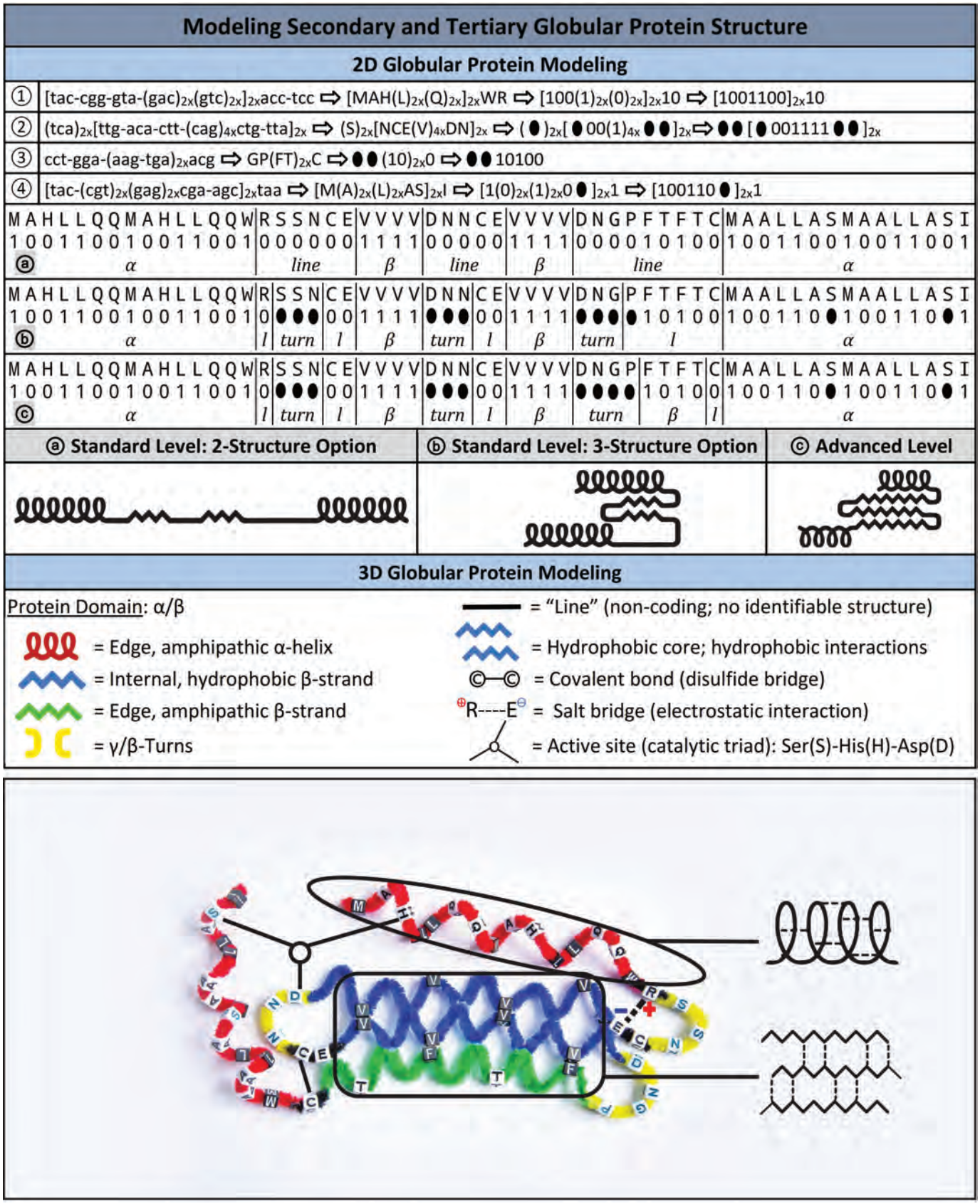


Figure 9. Modeling secondary and tertiary structure. Groups transcribe, translate, and transform four different DNA sequences, and assemble them into a 3D model to study the importance of tertiary refinements in stabilizing and locking in protein structure. Sequence length (shorter vs. longer) and T3 level can be used to differentiate instruction: (a) standard level 2-structure option; (b) standard level 3-structure option; and (c) advanced level.

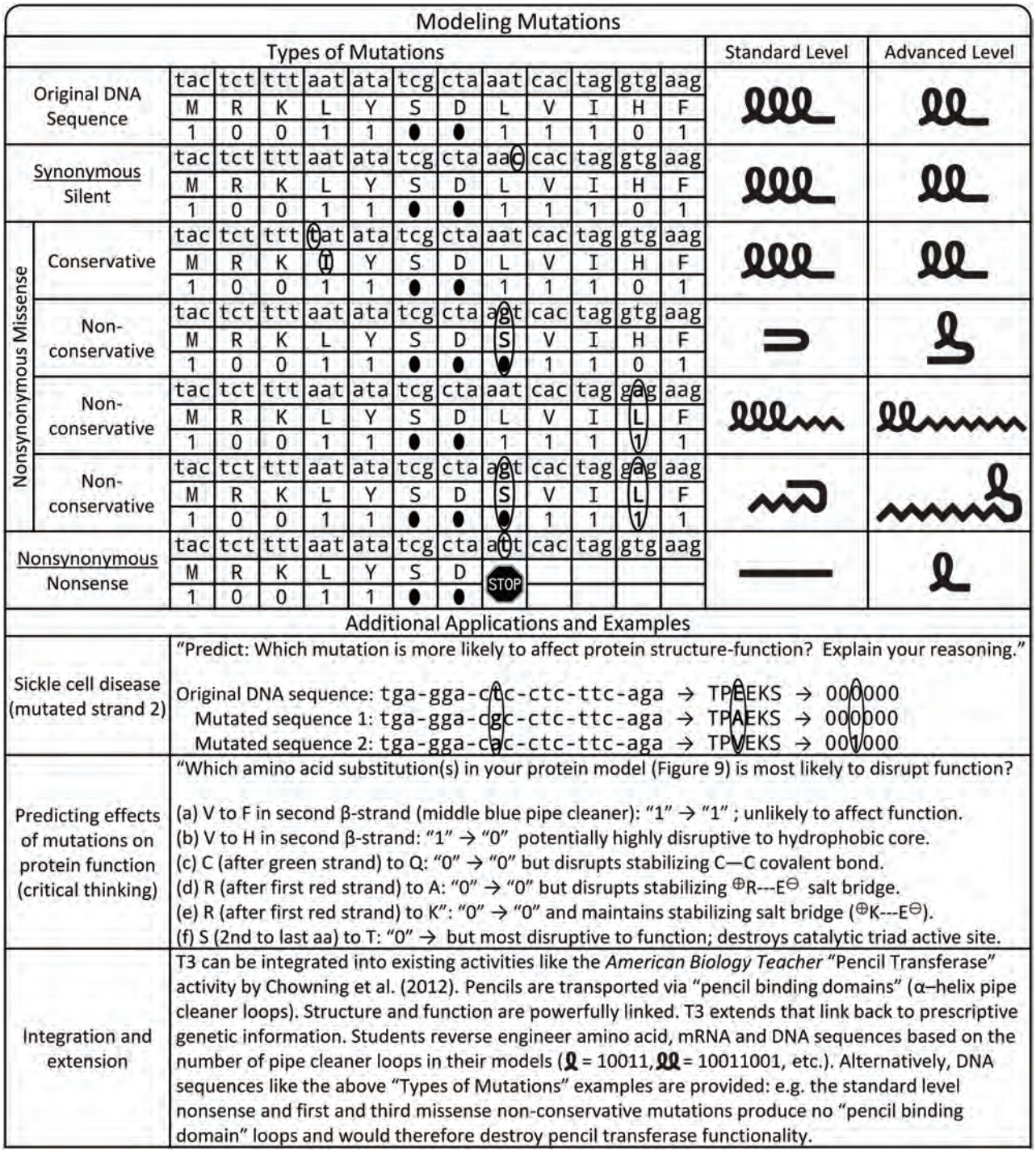


Figure 10. Modeling mutations.

possible. Transition to the standard level is then achieved by using amino acids from their respective propensity groups to design binary sequences for a given structure.

○ Practical Considerations

Implementation

T3 is implemented with minimal prep work by giving students: (1) DNA sequences that reduce to folding codes (standard-advanced) or clustered amino acids (basic) for one or more secondary structures; (2) single-letter amino acid abbreviations; and (3) procedures for step 3 ("transform") of the T3 Method. (See Figures 2–5 for standard-advanced; Figure 12 for basic.). Of the three, DNA sequence design is potentially the most time-consuming. The "T3 Design Tool" was created for this reason with busy teachers in mind to help expedite implementation. It is a spreadsheet-based prototype for a future phone app that allows teachers to enter the desired outcome they want for students. The program generates multiple "T3 It!" starting sequences for each solution input by the user. Teachers can then

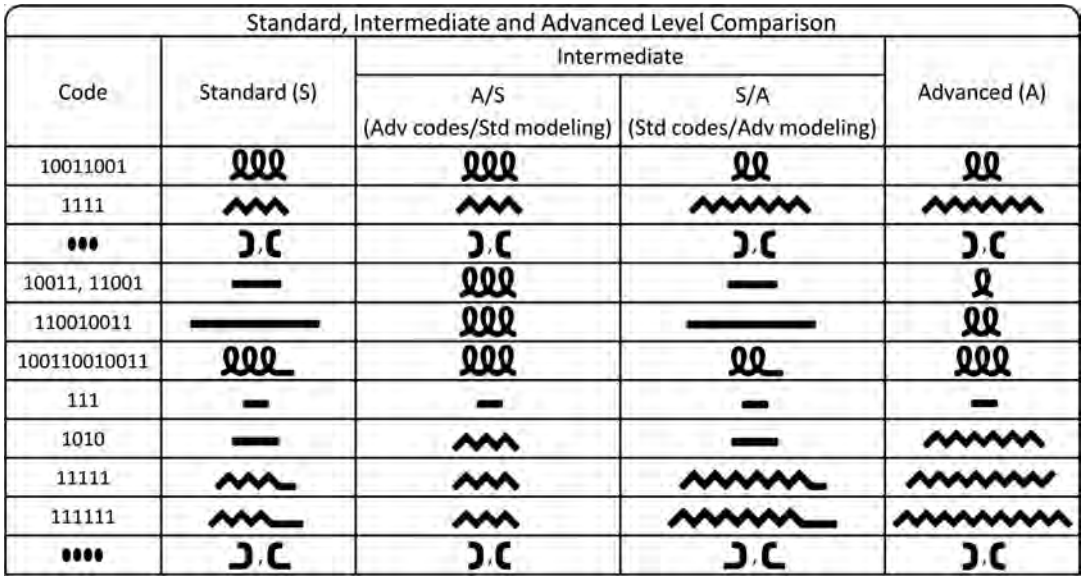


Figure 11. Standard, intermediate, and advanced level comparison chart.

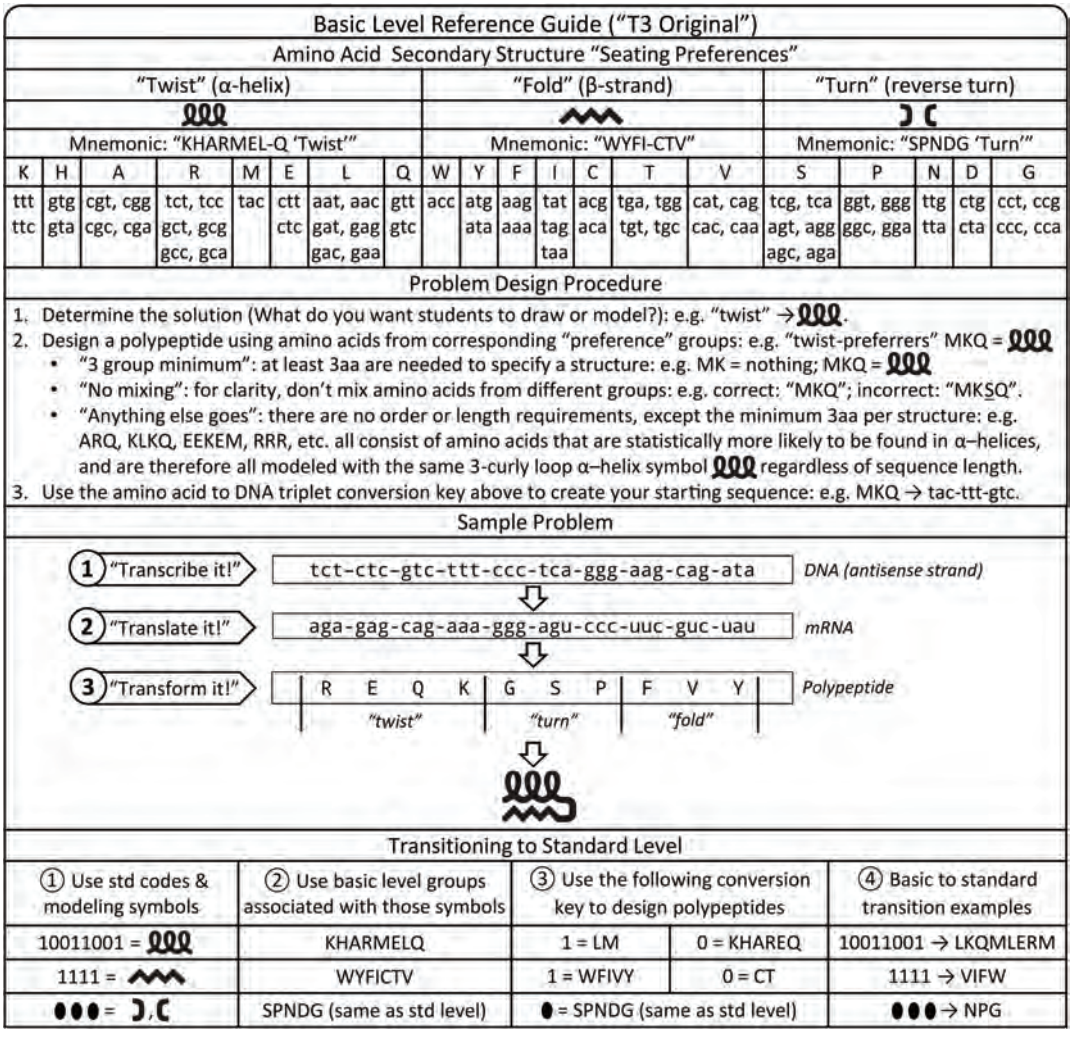
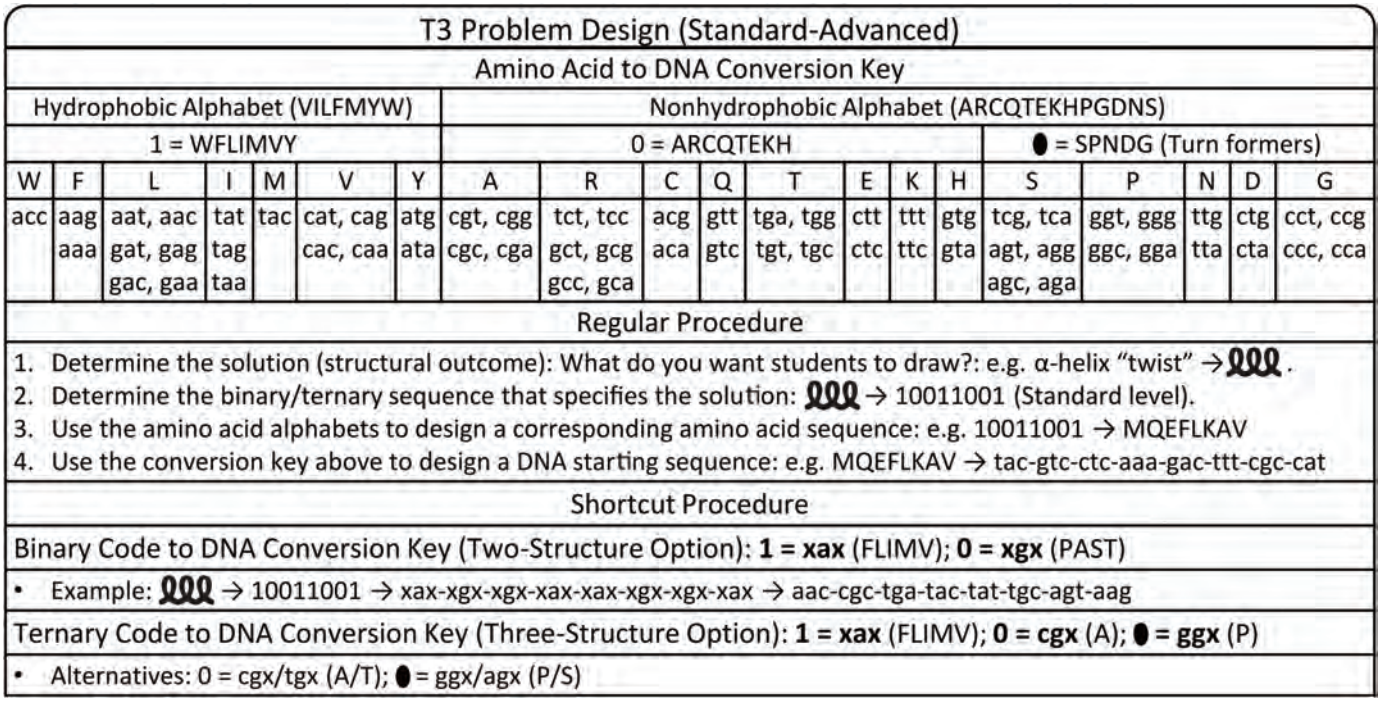


Figure 12. Basic level ("T3 Original") reference guide. Amino acid secondary structure propensity groups follow Williams et al. (1987) and largely agree with other studies. For example, Malkov et al. (2008) produced identical groups except for C and H, which they classified separately.



T3 Problem Design (Standard-Advanced)

Amino Acid to DNA Conversion Key

Hydrophobic Alphabet (VILFMYW)							Nonhydrophobic Alphabet (ARCQTEKHPGDNS)												
1 = WFLIMVY							0 = ARCQTEKH								● = SPNDG (Turn formers)				
W	F	L	I	M	V	Y	A	R	C	Q	T	E	K	H	S	P	N	D	G
acc	aag aaa	aat, aac gat, gag gac, gaa	tat tag taa	tac	cat, cag cac, caa	atg ata	cgt, cgg cgc, cga	tct, tcc gct, gcg gcc, gca	acg aca	gtt gtc	tga, tgg tgt, tgc	ctt ctc	ttt ttc	gtg gta	tcg, tca agt, agg agc, aga	ggt, ggg ggc, gga	ttg tta	ctg cta	cct, ccg ccc, cca

Regular Procedure

1. Determine the solution (structural outcome): What do you want students to draw?: e.g. α-helix "twist" → ➰.
2. Determine the binary/ternary sequence that specifies the solution: ➰ → 10011001 (Standard level).
3. Use the amino acid alphabets to design a corresponding amino acid sequence: e.g. 10011001 → MQEFLKAV
4. Use the conversion key above to design a DNA starting sequence: e.g. MQEFLKAV → tac-gtc-ctc-aaa-gac-ttt-cgc-cat

Shortcut Procedure

Binary Code to DNA Conversion Key (Two-Structure Option): **1 = xax** (FLIMV); **0 = xgx** (PAST)

- Example: ➰ → 10011001 → xax-xgx-xgx-xax-xax-xgx-xgx-xax → aac-cgc-tga-tac-tat-tgc-agt-aag

Ternary Code to DNA Conversion Key (Three-Structure Option): **1 = xax** (FLIMV); **0 = cgx** (A); **● = ggx** (P)

- Alternatives: 0 = cgx/tgx (A/T); ● = ggx/agx (P/S)

Figure 13. T3 Method problem design.

choose one sequence to give all their students to solve, or different sequences to work independently that still yield the same result for quick assessment. To stay informed of progress in development and learn how you can obtain your own T3 Design Tool, email t3designtool@gmail.com for more information.

Teachers can also reverse-engineer their own problems using the procedures in Figures 12 (basic) and 13 (standard-advanced). A faster shortcut procedure also exists for teachers in a time crunch (Figure 13).

The time needed to "transform" is variable and depends on numerous factors (e.g., academic level, learning curve, subsequent practice, sequence length, problem complexity, etc.). Once mastered, though, it generally takes students less time to transform than it does to translate at the basic level, and a comparable amount of time for the standard to advanced levels. Complex problems require additional time for students to critically assess. The total time needed to transcribe, translate, and transform a DNA sequence like the fifty-one nucleotide example in Figure 2 is around five minutes on average, once T3 is mastered.

The triple length of DNA and mRNA compared to polypeptides constrains problem length and complexity. Strategies to mitigate this include: (1) abbreviated problems and structure substitution; (2) "T3 Direct": students convert DNA sequence directly into binary code; and (3) group collaboration: longer, more complex problems are divided up and solved within groups (Figure 14).

Troubleshooting

Differentiated instruction, advanced level pattern-based codes, and even expansion beyond "codes-only" at the standard level to include "lines" (i.e., undefined sequences that don't code for a specific structure) provide teachers with flexible options, but can also create multiple-solution problems when only a single solution is desired. This occurs any time a binary sequence can be interpreted in more than one way. To use an analogy, should the hypothetical sequence "Darwinisnowhere" be interpreted as Darwin is "nowhere" or "now-here"? A similar problem exists when folding codes overlap. For example, at the standard level, is 11110011001 interpreted as "1111(β)-0011001(*l*)" 〰—, or as "111(*l*)-10011001(α)" ➰? In nature, the precise dividing line between structures isn't always clear. To capitalize on this teachable moment, retain ambiguity and leave interpretation to students. To eliminate ambiguity, ensure only single-solution problems are used. Figure 15 provides "Quick Fixes" and "Modeling 'Rules'" that teachers can use to ensure this. Depending on instructional context, some teachers may need to use Quick Fixes and "Modeling 'Rules'" extensively. Other teachers may find they never need to use them at all. Flexibility exists either way, but care should be taken to ensure that complexity never obscures conceptual clarity.

○ Looking Forward

The T3 Method is a powerful instructional tool for biology teachers that is versatile and easily adapted to diverse learning contexts. It has remarkable utility and potential for enhancing biology instruction that includes further applications and extensions to other areas of biology. The Appendix contains a list of additional ideas to illustrate the scope and range of possibilities. These can doubtless be expanded upon by the ingenuity of fellow educators.

Finally, I'd be remiss for limiting the T3 Method to students when I have benefited as well. It has improved my own understanding and for the first time enabled me to "read" amino acid sequences for myself, predict structural outcomes, confirm points of agreement, ponder possible reasons for disparities, and more—all without the aid of a computer; and that is a satisfying experience.

Managing Longer Problems

Abbreviation Method

Abbreviate problems and have students solve through binary conversion before expanding. Examples:

- $(cat)_{4x} \rightarrow (gua)_{4x} \rightarrow (V)_{4x} \rightarrow (1)_{4x} \rightarrow 1111 \rightarrow$ 
- $(tac\text{-}ttc\text{-}gtc\text{-}aac)_{2x} \rightarrow (aug\text{-}aag\text{-}cag\text{-}uug)_{2x} \rightarrow (MKQL)_{2x} \rightarrow (1001)_{2x} \rightarrow 10011001 \rightarrow$ 
- $(ccc)_{3x} \rightarrow (ggg)_{3x} \rightarrow (G)_{3x} \rightarrow (●)_{3x} \rightarrow$ ●●● →
- $[tac\text{-}ttc\text{-}gtc\text{-}aac]_{2x}[ccc]_{3x}[cat]_{4x} \rightarrow \rightarrow [MKQL]_{2x}[G]_{3x}[V]_{4x} \rightarrow [1001]_{2x}[●]_{3x}[1]_{4x} \rightarrow 10011001$ ●●● 1111 → 

Substitution Method

"Part 1: 'T3 It!': transcribe, translate and transform the following DNA sequences:

A. tac-ttc-gtc-aac-caa-ctg-cgt-aaa → aug-aag-cag-uug-guu-gac-gca-uuu → MKQLVDAF → 10011001 → 

B. aag-cag-tat-ata → uuc-guc-aua-uau → FVIY → 1111 → 

C. ccc-tca-ggg → ggg-agu-ccc → GSP → ●●● → 

"Part 2: Use your results from Part 1 (A, B & C) to model the following protein: A-C-B-C-B-C-B" 

Group Collaboration ("Divide and Conquer") Method

"Divide and conquer" longer, more complex problems as a group (See Figure 9; Appendix)

"T3 Direct" Method

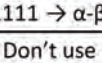
Teach students how to convert DNA directly into binary code using the "Shortcut Procedure" in Figure 13. Then give students problems designed by the "Shortcut Procedure" to solve. With a little time and practice, students can go further and learn how to convert any DNA sequence directly into binary or ternary code (See below).

Binary Code "Shortcut Procedure"	xax =1 (FLIMV); xgx = 0 (PAST)
Binary Code Expanded	1 = xax, acc, at[a/g]; "STOP" = act, at[c/t]; 0 = everything else
Ternary Code	● = [a/g]gx, ccx, ct_ , tc_ or tt_ + [a/g]

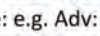
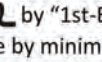
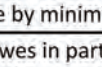
Figure 14. Managing longer problems.

Single-Solution Problems (eliminating ambiguity)

"Quick Fixes"

"Codes-Only" Conversion	Convert sequences to exact-match "codes-only" ("no lines") problems: e.g. Standard level: 10011001111 → 100110**1**1111 → 10011001-1111 → α-β → 
"Zero-Spacers"[a]	Insert "zero-spacers" to clarify problems (Note: Don't use "SPNDG" turn-formers): e.g. Advanced level: 101010011 → 1010**0**10011 → 1010-**0**-10011 → β-l-α →

Modeling "Rules"

"1st-Encountered" Rule	Model codes in the order they occur: e.g. Std/Adv Differentiation: ●●●● = ❩, ❨ (Adv); and ●●●—● = "*turn-line*" ❩, ❨ (Std); *not* ●—●●● = "*line-turn*" ❩, ❨
"Minimize-the-Lines" Rule (T3 version of parsimony)	The "best" model has the fewest lines without violating the other rules. Prioritizes "meaningful modeling" by maximizing structure: e.g. Adv: 101010011001 → 1010-10011001 →  ; 1010, 10101 and 101010 are all possible "1st-Encountered" codes, but only 1010 maximizes structure by minimizing the lines—in this case—to zero.
"Extreme-Structure" Rule (Advanced level option)	Minimize line number *and* length to maximize structure (Trumps "1st-Encountered" Rule): e.g. Adv: 110011001 → "11001(α)-1001(*l*)" →  by "1st-Encountered," but the alternative "1(*l*)-10011001(*a*)" →  , maximizes structure by minimizing both line number and length

[a]The success of hydrophobic cluster analysis (HCA) in predicting structure owes in part to its connectivity distance requirement (Hennetin et al., 2003). HCA avoids overlapping or intertwined hydrophobic clusters ("codes"). It only uses conditional binary patterns for analysis that are separated by a minimum of four nonhydrophobic amino acids ("0000"). For teachers, a practical approach is recommended to keep sequence length manageable: use only the minimum number of "zero-spacers" needed to clarify problems.

Figure 15. Troubleshooting: eliminating ambiguity and ensuring single-solution problems.

○ Acknowledgments

I wish to thank my high school biology students for their help in developing and testing the T3 Method, and colleague Dr. Claudia Geocaris for her helpful suggestions and reviews of manuscript drafts. I also wish to thank Ben Malas, Colleen Nedrow and Dina Portik for their photography and graphic design work. The image of PDB 3AJ5 was created with UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

References

Ahluwalia, U., Katyal, N., & Deep, S. (2012). Models of Protein Folding. *Journal of Proteins & Proteomics, 3*(2), 85–93.

Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., & Sauer, R. T. (1990). Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science, 247*(4948), 1306–1310.

Bradley, L. H., Thumfort, P. P., & Hecht, M. H. (2006). De novo proteins from binary-patterned combinatorial libraries. In R. Guerois & M. López de la Paz (Eds.), *Protein Design* (pp. 53–69). Totowa, NJ: Humana Press.

Bradley, L. H., Wei, Y., Thumfort, P., Wurth, C., & Hecht, M. H. (2007). Protein design by binary patterning of polar and nonpolar amino acids. In K. M. Arndt & K. M. Müller (Eds.), *Protein engineering protocols* (pp. 155–166). Totowa, NJ: Humana Press.

Broome, B., & Hecht, M. (2000). Nature Disfavors Sequences of Alternating Polar and Non-polar Amino Acids: Implications for Amyloidogenesis. *Journal of Molecular Biology, 296*, 961–968.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrisatt, B. & Mornon, J. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): Current status and perspectives. *Cellular and Molecular Life Sciences (CMLS), 53*(8), 621–645.

Chen, H., Gu, F., & Huang, Z. (2006). Improved Chou-Fasman method for protein secondary structure prediction. *BMC bioinformatics, 7*(Suppl 4), S14.

Chowning, J. T., Kovarik, D., & Griswold, J. (2012). Modeling Protein Structure & Function: Pencil Transferase. *American Biology Teacher, 74*(8), 581–582.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry, 29*(31), 7133–7155.

Dill, K., Bromberg, S., Yue, K., Chan, H., Ftebig, K., Yee, D., & Thomas, P. (1995). Principles of protein folding—A perspective from simple exact models. *Protein Science, 4*, 561–602.

Eudes, R., Le Tuan, K., Delettré, J., Mornon, J. P., & Callebaut, I. (2007). A generalized analysis of hydrophobic and loop clusters within globular protein sequences. *BMC structural biology, 7*(1), 2.

Harle, M., & Towns, M. H. (2013). Students' understanding of primary and secondary protein structure: Drawing secondary protein structure reveals student understanding better than simple recognition of structures. *Biochemistry and Molecular Biology Education, 41*(6), 369–376.

Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., . . . & Zhou, Y. (2015). Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Scientific reports, 5*, 1–13.

Hennetin, J., Le Tuan, K., Canard, L., Colloc'h, N., Mornon, J. P., & Callebaut, I. (2003). Non-intertwined binary patterns of hydrophobic/nonhydrophobic amino acids are considerably better markers of regular secondary structures than nonconstrained patterns. *Proteins: Structure, Function, and Bioinformatics, 51*(2), 236–244.

Iwamoto, N., Ito, S., Sumi, D., Kobayashi, M., Kumagai, T. (Forthcoming). Post-translational modification of a non-catalytic Cys121 of PTP1B. Currently available at Protein Data Bank (PDB ID: 3A5J), http://www.rcsb.org/pdb/explore.do?structureId=3a5j.

Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., & Hecht, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science, 262*(5140), 1680–1685.

Ladunga, I., & Smith, R. F. (1997). Amino acid substitutions preserve protein folding by conserving steric and hydrophobicity properties. *Protein engineering, 10*(3), 187–196.

Malkov, S. N., Živković, M. V., Beljanski, M. V., Hall, M. B., & Zarić, S. D. (2008). A reexamination of the propensities of amino acids towards a particular secondary structure: Classification of amino acids based on their chemical structure. *Journal of molecular modeling, 14*(8), 769–775.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of computational chemistry, 25*(13), 1605–1612.

Ptitsyn, O. B. (1973). Stages in the mechanism of self-organization of protein molecules. *Doklady Akademii Nauk SSSR, 210*(5), 1213.

Simm, S., Einloft, J., Mirus, O., & Schleiff, E. (2016). 50 years of amino acid hydrophobicity scales: Revisiting the capacity for peptide classification. *Biological Research, 49*(1), 31.

Trifonov, E. N. (2008). Codes of biosequences. In M. Barbieri (Ed.), *The codes of life: The rules of macroevolution* (pp. 3–14). Dordrecht, NL: Springer.

West, M. W., & Hecht, M. H. (1995). Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Science, 4*(10), 2032–2039.

Williams, R. W., Chang, A., Juretić, D., & Loughran, S. (1987). Secondary structure predictions and medium range interactions. *Biochimica et Biophysica Acta (BBA)—Protein Structure and Molecular Enzymology, 916*(2), 200–204.

Willingham, D. (2009). *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco, CA: Jossey-Bass.

BRETT MALAS is a National Board-Certified Teacher, author, lecturer, and associate with the Center for Teacher Effectiveness, and high school science teacher with degrees in biology and paleontology; e-mail: t3method@gmail.com

APPENDIX

① **Connecting to Other Biology Topics:** T3 naturally connects to other topics in biology like biochemistry, molecular biology, genetics, and evolution. The use of T3 to model genetic mutations, predict structural and functional effects, and study point mutations has already been discussed (Figure 10), but T3 has potential utility with any protein-related topic, including: enzyme catalysis; genetic disorders; molecular evolution; allele, nucleotide, and polypeptide sequence-structure comparisons; cell signaling and communication; protein misfolding diseases (e.g., Alzheimer-related fibrillar β-amyloid proteins); pharmaceutical applications; multiple sequence alignment; evolutionary history (e.g., hydrophobic cores are highly conserved, and the ancestral form of any given sequence is suggested by the binary code; Trifonov, 2008); and more.

② **Instructional Utility:** T3 can be taught at a specific, targeted level that is scaffolded to student ability or adapted to meet course requirements and objectives. Students can be individually assigned the same problem to solve either at the basic level, at the standard level two-structure option, at the standard level three-structure option, at the intermediate level, or at the advanced level to differentiate instruction, all in the same classroom. T3 can be taught progressively. For example, transition students from the basic level to the standard level, or from the standard level to the advanced level: (1) as ability improves; (2) for further inquiry; (3) to reinforce an important concept; (4) to explore or study a particular aspect of protein structure in greater detail; (5) to teach a particular skill or process (e.g., progressing from standard to advanced level modeling to teach the importance of model refinement in science).

③ **Differentiated Instruction and Group Collaboration:** Use T3 to differentiate instruction and model primary to quaternary structure. Solve larger T3 problems via group collaboration; abbreviate and break-up long sequences into smaller ones (Figure 14). Individually assign sequences of varying length and complexity to differentiate instruction within collaborative groups. Further differentiate by having students individually solve at levels scaffolded to their ability with later collaborative group assembly of the final tertiary (Figure 9) or quaternary structure protein product.

④ **3D Modeling:** Extend pencil-paper models with materials commonly used in 3D modeling (e.g., pipe cleaners, beads, etc.). Direct students to model their 3D protein according to specific teacher-designed or student-designed keys that use different colored pipe-cleaners to distinguish different secondary structures (Figure 9), and/or beads of different shape (e.g., spherical/round vs. cubic/square), color, size, and/or composition (e.g., plastic, metallic, wood) to highlight or distinguish features of interest. For example: (1) we have used black and white beads to distinguish nonpolar and polar amino acids (Figures 7, 9); (2) we have used beads of different shapes and/or colors to delineate secondary structure propensities of amino acids grouped by "twist-," "turn-," and "fold-preferers" at the basic level (see this article's header image); (3) we have modified further to experiment with other possibilities (e.g., different colors and/or shapes for amino acids with special characteristics; large wooden cubes for amino acids commonly found in hydrophobic cores) (see article header image); and (4) *a personal fan-favorite*: we have used glow-in-the-dark beads to highlight amino acids with special characteristics (e.g., disulfide bridge-forming cysteines; salt bridge-forming charged amino acids: $K^{\oplus}R^{\oplus}$—$E^{\ominus}D^{\ominus}$).

⑤ **"T3 Reverse":** Give students protein structural models and have them reverse-engineer binary, amino acid, mRNA, and DNA sequences.

⑥ **"Designer Proteins":** *De novo* protein design is an exciting area of application normally inaccessible to students that T3 unlocks. Students apply their T3-acquired knowledge individually or as a group to design their own proteins complete with structure, putative function, binary patterning, and corresponding amino acid, mRNA, and DNA sequences. The final product, then, is a student-designed T3 problem with accompanying solution that can include 3D modeling (Figure 9). Design can be open-ended or narrowed to specific problem-solving tasks (e.g., genetic disorders, pharmaceuticals). Instructionally, this activity *reinforces sequence-to-structure and structure-to-function conceptual understanding in an unprecedented way* through students' newly acquired knowledge of protein folding.

⑦ **"Cracking the [Protein Folding] Code":** T3 is easily adapted to different instructional styles, strategies, and methodologies from traditional to constructivist. For example, instead of giving students folding codes, teachers can take a "Discovery" or constructivist approach and have students derive folding codes from their own observations. Students can then also supply their own folding code names and "rules." Transition to the advanced level can occur by the same (e.g., students infer the repeating "100110010011 . . . " codes for α-helices from its similarity to the standard level "10011001").

⑧ **Structural Motifs and Domains:** Use T3 to expand structural modeling from the basic to higher-level advanced. Model domains (α, β, α + β, α / β) and supersecondary structure (structural motifs): e.g., parallel (Figure 6) and antiparallel β-sheets (Figures 5, 6.11b); helix-turn-helix (Figure 6.14); βαβ motif (Figure 6.18); Greek Key motif (Figure 6.17).

⑨ **Protein Structure Prediction:** Have students "read" and interpret real or designed globular protein sequences. Students convert polypeptides into binary code (or receive sequences already in binary code) and draw representative structural models. Students then compare their models with structural "wiring diagrams" from the Protein Data Bank (PDB) for a given sequence. Problems can range from exact correspondence to increasing disparity. Have students go

deeper with their observations, noting areas of correspondence and theorizing explanations for deviations from empirically established results.

⑩ **"Beyond Protein Folding 101":** Advance and refine students' ability to "read" and interpret globular protein sequences. For example, Eudes et al. (2007) provide a supplementary dictionary of the 461 most frequently occurring hydrophobic clusters and loop clusters in globular proteins. These can be given student-friendly names and used to expand students' "folding code" vocabulary. For example, additional binary sequences that correlate with β-strands include "Single-Zero Intruders" (1011, 1101), "Isolated Pairs" (100111), and "Split Pairs" (1011101). Incorporate amino acid propensities and substitution tolerances to further improve students' ability to "read" and interpret binary sequences: e.g., nonhydrophobic ("0") threonine (T) frequently substitutes for hydrophobic ("1") valine (V) or isoleucine (I) in β-strands.

⑪ **Advanced Study and "Exploring Exceptions":** Apply hydrophobicity and binary coding principles for specialized and advanced study: e.g., study the amphipathic-type binary patterning in leucine zippers of certain eukaryotic transcription factors, the stabilizing hydrophobic core created by dimer formation, and how the resulting leucine zipper structure relates to DNA-binding functionality. Extend application to non-globular, fibrous proteins: e.g., the 1st and 4th position hydrophobic residues at the internal interface of strands in heptad repeats of coiled coil structure in keratin. Study exceptions to binary sequence paradigms: e.g., non-globular, fibrous proteins like the collagen triple helix with its characteristic "(Gly-X-Y)n" sequence motif (X = proline; Y = hydroxyproline).

⑫ **Other ("Project T3"):** Additional ideas are welcomed. Email t3method@gmail.com with subject line "Project T3" if you are interested in contributing.