ACS BIO & MED CHEM Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

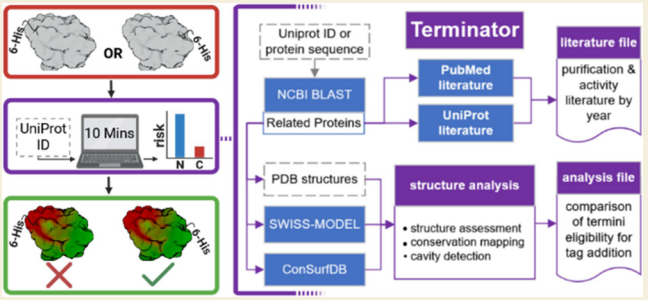
pubs.acs.org/biomedchemau
Article

Terminator: A Software Package for Fast and Local Optimization of His-Tag Placement for Protein Affinity Purification

Rokas Gerulskis and Shelley D. Minteer*

Cite This: ACS Bio Med Chem Au 2025, 5, 55−65
Read Online

ABSTRACT: Although the use of affinity tags can greatly improve purification of expressed enzymes, the placement of affinity tags can significantly impact the expression, solubility, and function of recombinant proteins. To facilitate the optimal design of 6xHis-tagged constructs for protein purification, we developed Terminator, a Python-based software package, which takes a UniProt ID or existing protein sequence as input, identifies related sequences, maps sequence conservation retrieved from ConSurf onto protein 3D structures retrieved from the PDB and SWISS-MODEL, and analyzes proximity to cavities and functional sites to recommend the N- or C-terminus for placement of 6xHis fusion tags <15 residues in length. The package also outputs a document with available purification and activity literature for the target and closely related proteins organized by year. Comparative analysis of Terminator predictions against published experimental tag behavior for 6xHis fusion tags <15 residues in length demonstrates an 86−100% accuracy in predicting the relative risk of ill effects between termini and a 92−93% accuracy in predicting the absolute risk of modifying individual termini. This reliability of Terminator's analysis suggests that proximity to surface cavities, not burial of wild-type termini, is the most reliable predictor of ill effects arising from short 6xHis fusion tags. This tool aims to expedite construct design and enhance the successful production of well-behaved proteins for studies in enzymology and biocatalysis with minimal need for computational resources, programming knowledge, or familiarity with protein-tag interference mechanisms.



KEYWORDS: *6xHis tag, affinity purification, protein design, bioinformatics tool, ConSurf, enzymology, biocatalysis*

INTRODUCTION

The last three decades have seen a significant acceleration in the research and industrial application of biocatalytic methods.[1] This growth is due, in part, to method developments significantly improving the ease of obtaining large amounts of target proteins for catalytic research. Early protein research was hampered by the need to purify target proteins from their original, poorly characterized host organisms. Developments in gene regulatory architectures and genetically tuned host organisms then refocused the protein procurement workflow to growing protein in foreign organisms (recombinant expression), chiefly in workhorse strains of *Escherichia coli* and other species.[2,3] Although protein yields saw significant improvement, procedures were still bottlenecked by the need to purify target genes from the original host organisms and package them into DNA architectures for recombinant expression. This limitation was addressed in the past decade with significant advances in commercial gene synthesis and vector construction services, which reliably and inexpensively convert digital DNA sequences into codon-optimized, physical plasmids ready for expression in host organisms.[4,5] All these advances have narrowed the labor of protein scientists to selecting a target sequence from a protein database such as UniProt,[6] ordering the selected gene from a commercial gene synthesizer, and performing subsequent expression and purification procedures with the aid of available literature.

In parallel to these developments, affinity tags quickly became an indispensable tool for the purification of recombinantly expressed proteins by enabling specific affinity between target proteins and purification matrices.[7] Affinity tags range in size and mechanism, from small 6xHis tags binding nickel-containing polymers to 400-residue maltose-binding protein, offering additional benefits to solubility.[8,9] An affinity tag is added to a protein by extension of the start or end of its genetic sequence and consequently its protein sequence, typically through site-directed mutagenesis, targeted positioning in a DNA vector, or digital modification prior to gene synthesis.

Received: July 5, 2024
Revised: December 16, 2024
Accepted: December 17, 2024
Published: January 3, 2025

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

Carefully selecting which protein terminus should be modified with an affinity tag is critical to the successful expression and purification of a target protein, as protein sequence modification with affinity tags can have radical effects on solubility,[10−12] stability,[13] and catalytic efficiency through structural interference near the site of modification. A long linker sequence is sometimes added between the protein and affinity tag in an attempt to minimize the risk of structural issues (the combination of an affinity tag with linker or cleavage sequences, etc., is known as a *fusion* tag), but this may simply introduce longer-range or even interassembly tag interference.[14] Simply replicating tag architecture established in the literature for a target can be a reliable option for obtaining functional proteins; however, experimental verification of tag noninterference is uncommon, while the rationale for choosing one terminus versus the other is exceptionally rare.

If purification protocols for a target protein are not established, researchers typically evaluate the suitability of a protein terminus for tag addition by loading the protein's crystal structure, when available, into molecular visualization software. One can then visually assess how buried the terminus is within the surrounding structure and its proximity to the enzyme's active site, as informed by published experimental data. However, this approach may be hindered by the researcher's familiarity with the visualization software, including difficulty in identifying catalytic residues or interpreting terminal coordinate assignment due to nonstandard sequence numbering and knowing to convert the default asymmetrical unit stored in a PDB file to the biologically relevant quaternary assembly. Furthermore, conclusive identification of an enzyme active site can take several years of experimental deliberation.[15,16] More rigorous prediction of tag behavior via molecular modeling (MM) and molecular dynamic simulation requires familiarity with computational tools and a thorough understanding of complex modeling outputs, a time investment unavailable to many experimentalists, particularly in biocatalysis fields only tangential to enzymology.[17]

Research in enzymology exploring natural protein behavior requires rigorous investigation of tag−protein interaction and often the removal of tags postpurification to better represent wild-type behavior. In contrast, biocatalytic research areas such as brewing, textile manufacture, and bioelectrochemistry need to devote minimal time to natural protein behavior, devoting attention to downstream optimizations, such as protein operational stability, reaction vessel architecture, immobilization strategy, or interaction in catalytic cascades. In short, many protein-employing research areas merely need to quickly obtain a functioning protein to focus on other optimizations in more complex systems.

This work introduces Terminator, a Python package that automates the collection of literature and protein structure and residue conservation data to rapidly predict structural and catalytic issues arising from the addition of a 6x-His tag to either protein terminus, completing in a few minutes on a personal computer starting with only a UniProt ID or partial protein sequence. Comparative analysis of Terminator's predictions with published data demonstrates an 86−100% success rate in predicting relative suitability between termini when short (<14 residues) tags are employed, with a 92−93% accuracy in predicting the absolute suitability of individual termini for the addition of 6x-His tags.

METHODS

Structure Selection

Terminator begins with an input of either a UniProt ID or, for convenience, a partial protein sequence of an existing protein. In either case, it then performs a BLAST multiple-sequence alignment[18] against the UniProtKB database to find synonymous UniProt entries, well-characterized proteins with >80% sequence identity to the target protein. This operation is especially useful in cases where a manual search for protein structures fails despite structures merely being categorized under separate UniProt entries with identical or very closely related protein sequences. The protein sequence is then retrieved from the input UniProt entry, or the highest-scoring entry is retrieved using the input sequence. This sequence is then modified by subtracting any residues belonging to a transit, localization, or propeptide sequence as cataloged in the UniProt entry, as these are typically removed in vivo in the process of protein maturation. The SWISS-MODEL database[19] is subsequently queried for entries using each synonymous ID until a model is retrieved. The SWISS-MODEL server not only quantifies the relative quality of available structures by discriminating crystal artifacts from biological contacts and can even offer model crystal structures based on similarly folding sequences belonging to unrelated proteins. Consequently, the SWISS-MODEL query serves a similar function as synonymous entry retrieval, as it can retrieve a crystal structure when no structure is available across *any* synonymous entries.

Literature Retrieval

The tangential utility of the Terminator package is the literature retrieval tool, which provides an Excel document with a significant number of structure, purification, and activity literature, with links, retrieved from UniProt and a Boolean PubMed[20] query, organized by year. The PubMed query functions by submitting every name for the target protein and species of origin collected from synonymous UniProt entries, with the narrowing terms "expression", "purification", "mechanism", and "activity". The user is advised to search for additional literature, nevertheless, due to occasional inconsistencies in enzyme naming in the literature versus names assigned in UniProt and because both databases may exclude literature in biochemistry-adjacent fields such as bioelectrochemistry. However, test runs for enzymes with which we are rigorously familiar did retrieve additional papers that we had missed in traditional searches. Even if the user judges that crystal structure data are insufficient or their analysis by Terminator is unnecessary, the literature retrieval function is an invaluable time-saving tool for collecting data to write up a purification or assay protocol.

ConSurf

The ConSurf server is instrumental in the determination of protein functional regions from sequence-evolutionary data alone.[21] The tool functions by clustering related protein sequences into a phylogenetic tree and calculating residue-specific conservation scores as a function of evolutionary rate. In short, ConSurf estimates structurally and catalytically important residues in a protein sequence without information from wet-lab experiments, providing insights that are rarely determined experimentally due to the exponential complexity of sequence-space-dependent experiments. One interesting example is W274 in oxalate decarboxylase (OxDC) from *Bacillus subtilis*, which has the ninth-highest conservation score of OxDC's 385 residues. This residue was shown to participate in interchain electron−hole hopping to protect the protein from oxidative damage in 2021.[22] The importance of this residue to OxDC function could have been demonstrated by ConSurf possibly as early as 2001, 20 years before it was demonstrated experimentally.[23] The specific way in which Terminator employs ConSurf outputs is discussed further in the following section and in detail in the Supporting Information (SI).

Terminus Quality Metric Calculation

A detailed discussion of how Terminator analyzes a crystal structure and compares protein termini is provided in the Supporting Information; however, a simplified explanation is provided below.

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

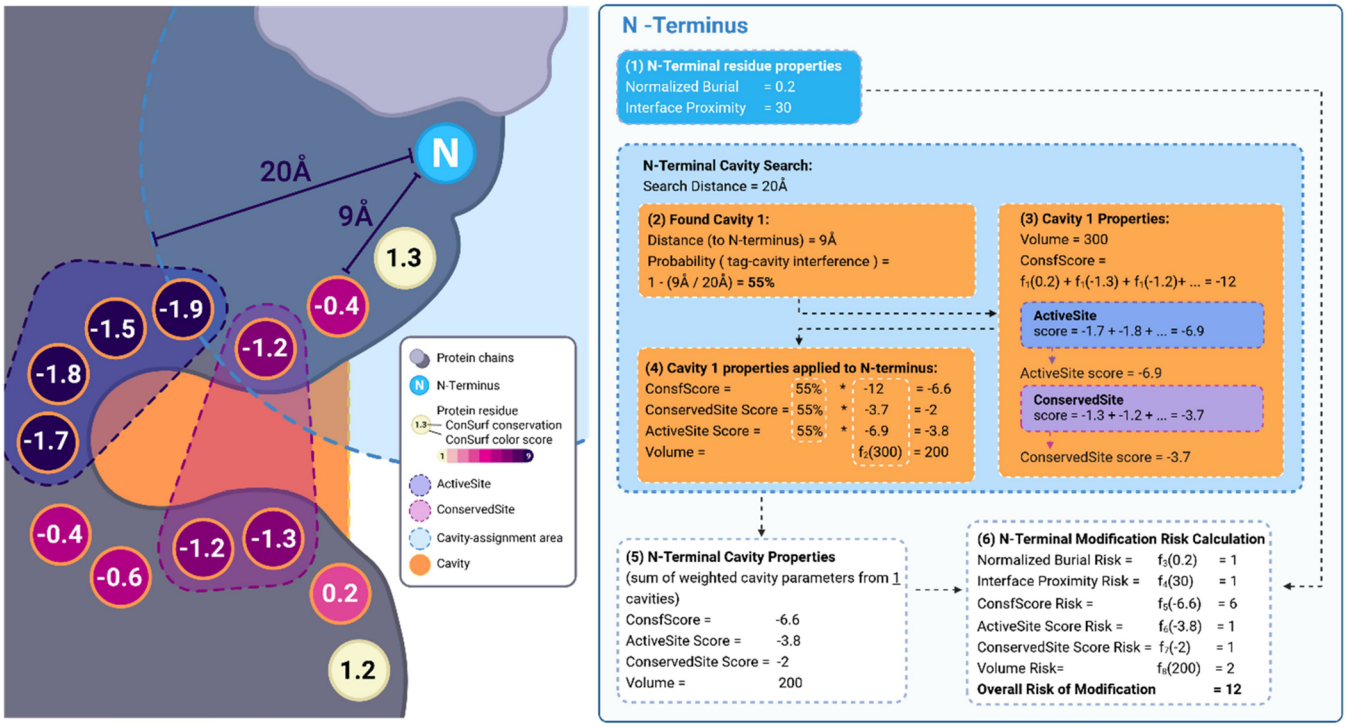


Figure 1. Simplified representation of Terminator's structure−analysis algorithm on a hypothetical two-dimensional protein. Light and dark gray regions represent separate protein chains. Beige to purple circles represent simplified protein residues colored by their ConSurf-derived color value and numbered according to their ConSurf-derived conservation value. The N-terminus is represented by the cyan "N" circle, with its cavity−assignment radius as a large light-blue region. An ActiveSite and ConservedSite are represented as navy and purple regions, respectively. A cavity is represented as the orange region with its assigned residues outlined in orange. To the right, a flowchart illustrating how each object recognized by Terminator (ActiveSite, ConservedSite, cavity, and terminus) defines its own properties and aggregates the properties of subordinate objects, leading to an evaluation of the status of a terminus.

Figure 1 shows a simple representation of Terminator analyzing a terminus on a hypothetical, 2-dimensional protein.

First, Terminator searches the crystal structure for collections of proximal residues (Figure 1, beige/purple/navy circles) having the same ConSurf-provided color score, implying a similar conservation value among protein homologs and, consequently, a similar level of functional importance. These collections are defined as a new object called a "Site," which comes in two forms: an "ActiveSite," composed of residues with the highest possible color score of 9, and a "ConservedSite," whose residues each have a color score of 8. These Site objects are assigned a conservation "site score" equal to the sum of the ConSurf-provided conservation values of their residues. This score will be employed again later in quantifying the functional importance of protein cavities (Figure 1, box 3). Commencing the analysis of termini (Figure 1, box 1), the residue-level properties of the N-terminus are determined: its burial value normalized against other residues in the structure, representing the likelihood of a tag added to this terminus disturbing the tertiary structure, and its interface proximity, representing the likelihood that a tag may disturb the quaternary assembly of the protein (represented in Figure 1 by proximity to a foreign chain, light gray). Next, Terminator floods an extended area surrounding the terminus with pseudoatoms (not pictured; see *Example 1* in Discussion section III) to determine the location and volume of each nearby cavity. When a cavity is found (Figure 1, box 2), a "cavity" object is created and assigned a tag−cavity interference probability based on the cavity−terminus distance and the radius of the cavity−assignment area (cyan dashed region). A volume is also assigned to the cavity based on the number of pseudoatoms assigned to it during its initial detection. Then (Figure 1, box 3), each cavity is further assigned a broad conservation score "ConsfScore" as a modified sum of the ConSurf-derived conservation values of the residues assigned to the cavity (orange outlined circles). Similarly, the cavity is also assigned ConservedSites and ActiveSites based on the cavity and the Site sharing residues, and the "site scores" of these Sites are transferred directly to the cavity's "ConservedSite score" and "ActiveSite score". Thus, the cavity is assigned four properties to represent its functional importance: a volume, a general "ConsfScore" encompassing all of its residues, an "ActiveSite score" from the site score of its assigned ActiveSites, and a "ConservedSite score" from the site score of its assigned ConservedSites.

These cavity properties are transferred to the proximal terminus so that its overall suitability for tag addition may be evaluated (box 4). The ConsfScore, ConservedSite score, and ActiveSite score of the cavity are normalized by the cavity−tag interference probability, while the volume is modified using a pseudoatom searching algorithm ("$f_1()$"). The terminus is then assigned homonymous properties using the sum of these modified values from each assigned cavity (box 5). When the eligibility of a protein's N- and C-termini for tag addition is compared, these 6 terminus-assigned properties (in boxes 1 and 5) are compared directly between the termini, as they collectively represent the functional importance of regions near the protein terminus, which might be disturbed if the termini were modified with a His-tag. When a risk of modification is calculated for the terminus (box 6), these values are instead normalized with separate integer-division functions ("$f_{3-8}()$") and summed to a "risk" score for each terminus, intended to suggest to the user the likelihood that a 6xHis fusion tag added to this terminus may cause structural interference leading to a change in catalytic activity (an activity altering terminus, AT) or not cause structural interference (a safe terminus, ST). Readers are directed to the Supporting Information for a guide on how to download and operate Terminator (github.com/MinteerLab).

Selection of Literature to Test Package Efficacy

An initial literature search for studies that explicitly discuss His-tag-related enzyme performance was conducted with a Boolean search on PubMed, with additional literature collected by parsing a JSON export

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

from the Braunschweig Enzyme Database (BRENDA).[24] This resulted in 43 papers explicitly discussing His-tag performance. Of these, 14 were omitted due to a lack of available structures or a lack of clarity regarding the length of fusion tags employed in experiments. This left 29 papers discussing 31 proteins (Table 1), 16 of which investigated His-tags at both termini independently (p1, p4−5, p7−12, p16−17, p24−26, and p29−30), while the remaining 15 proteins were experimentally investigated with a His-tag at only one terminus relative to the wild-type protein (p2−3, p6, p13−15, p18−23, p27−28, and p31). Although AlphaFold structures are available for the majority of UniProt entries, these could not be employed for the structure-lacking publications because AlphaFold does not yet identify the oligomeric state of the protein or determine the structure of the quaternary assembly, which is critical to an accurate evaluation of tag−structure interference.

RESULTS AND DISCUSSION

I. On the Applicability of Available Experimental Data

Because Terminator predicts 6xHis fusion tag effects by searching for cavities within a 20 Å radius of a protein terminus as it appears in the crystal structure, there are two caveats to consider when comparing Terminator's prediction of fusion tag behavior to published kinetic data of 6xHis fusion-tagged proteins: the issue of terminus coordinates assignment in the available structural data employed by Terminator and the issue of fusion tag lengths experimentally employed in the published literature.

In the first case, it is not uncommon for several residues at a terminus in a crystal structure to be too labile to demonstrate a well-defined position resulting in a lack of assigned coordinates in the final structure. The position of a fusion tag added to this terminus may then be displaced from the apparent location of the terminus on the crystal structure (the first residue with successfully assigned coordinates) by up to 3.8 Å per unassigned residue. Terminator may therefore erroneously predict AT status when the apparent terminus is closer to a cavity than the actual wild-type terminus (false positive) or erroneously predict ST status when the apparent terminus is erroneously further away (false negative). Terminator identifies the number of residues with unassigned coordinates at both termini to guide the user in selecting only high-quality structures for in-depth analysis by automatically parsing structural data against the associated UniProt entry. Although any protein that failed to demonstrate a structure with sufficient terminal coordinate assignment was largely omitted from analysis in this work, Terminator users must carefully consider this point when interpreting Terminator's predictions regarding their protein of interest.

The other caveat to consider when comparing Terminator's predictions against published kinetic data for tagged proteins is the tag length in the published data. A minimal 6x-His tag, as offered C-terminally in many pET plasmids, has a significantly shorter interaction range with protein surface elements than would a tag containing spacers or cleavage sequences, such as the N-terminal tags of pET15b (20 residues) or pET30a (44 residues). Because Terminator assigns cavities only within 20 Å of a protein terminus, it often fails to recognize every avenue of interference afforded to longer fusion tags. This can be expected to increase false negative rates when the AT status was experimentally determined using long fusion tags. A high false positive rate might be expected if longer tags are denied interference with proximal cavities due to the formation of rigid secondary structures, high entropic costs for local interference, or formation of stable interactions with distal, benign cavities in a catalytically favorable orientation. In the following section, we introduce a set of proteins with published data concerning the effect of 6x-His tags on protein activity, process these proteins using Terminator, and analyze the relationship between Terminator accuracy and the length of fusion tags employed in these publications.

II. Evaluating Terminator's Accuracy against Published Data

Thirty proteins from 28 publications were analyzed with Terminator to test its capacity to assign AT or ST status and capacity to accurately predict which terminus would exhibit a less severe effect on enzyme activity when modified with a 6x-His tag (Table 1). All proteins were analyzed with only their UniProt ID as input, targeting the structure retrieved from SWISS-MODEL, with exception to p3, whose SWISS-MODEL-retrieved structure contained an N-terminal propeptide sequence as identified in its UniProt entry. These 28 papers provide experimental evidence for 21 ATs (terminus < WT or terminus > WT in the "relative experimental performance of tagged proteins" column, where N and C refer to the N- or C-terminus of the preceding protein, p4N, p16−21N, p29−30N, p13−17C, p22−23N, p27−28N, p31N, p29−30C) and 5 STs (terminus = WT, p2−4C, p5−6N), and 14 selections of a safer terminus (N < C or C < N, proteins p4−5, p7−p12, p17, p24−26, p29−30). p1 is provided as an example of a lack of terminal coordinate assignment demonstrating erroneous predictions to emphasize the warnings discussed in Discussion Section I but is omitted from statistical analysis. Evaluation of Terminator's predictions is discussed below.

Interpretation of Predicted Risk Values. A risk cutoff of 8 was chosen as the upper limit for termini to be assigned ST status, employing termini experimentally characterized with fusion tags less than 15 residues in length in published experiments (Table 1) as this risk cutoff value minimizes the false positive rate to 20% (1/5 termini; the incorrect prediction of p4C is discussed in detail in Discussion Section III, *example* 3) while maintaining a false negative rate of 0% (0/9 termini) (Figure 2). The selection of 14 residues as the upper limit for affinity tag length is elaborated in the following section. False positive refers to incorrectly predicting AT status for an experimentally determined ST, and false negative refers to incorrectly predicting ST status for an experimentally determined AT.

Effect of Affinity Tag Length on Prediction Accuracy. When including all tested proteins regardless of the length of the fusion tag employed in published experimental data, Terminator demonstrates only a 64% (9/14 proteins) accuracy in selecting the safer terminus on modification (Figure 3A, protein-specific result provided in the legend) and a 72% (18/25 termini) accuracy in categorizing individual termini as AT or ST (Figure 3B, terminus-specific results in the legend). As discussed in Discussion Section I above, the accuracy of Terminator's predictions depends on the length of the fusion tag employed in experiments. Accordingly, excluding proteins and termini that were tested with fusion tags longer than 14 residues in published work increases accuracy in selecting the safer terminus to 86% (6/7 proteins, p17 contradicting experimental data) and increases the accuracy in assigning individual AT/ST status to 93% (13/14 termini, p4C contradicting experimental data, as discussed in greater detail in Discussion Section III, *example* 3). p17 can be categorized as

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

Table 1. Complete Set of Proteins Analyzed by Terminator in This Work: The Number of Terminal Residues Without Coordinates in Their Best Crystal Structure, the Predictions Made by Terminator Regarding the Effect of Adding an Affinity Tag to Each Terminus, and the Experimental Performance of Each Protein With and Without Affinity Tags in Published Work[a]

		number of terminal residues without coordinates in best crystal structure		terminator outputs								published experimental data			
				N-terminus			C-terminus			selection					
abbreviation	UniProt ID	N-terminus	C-terminus	risk of structure alteration	AT/ST status	correct prediction	risk of structure alteration	AT/ST status	correct prediction	preferred terminus	Correct prediction	relative experimental performance of tagged proteins	length of N-terminal tag used	length of C-terminal tag used	reference
p1	Q96EY8	24	11	3	ST	No	15	AT	No	N	No	$N < C = WT$	6	6	25
p2	Q9BHM6	1	0	2	ST	[b]	0	ST	Yes	C	[b]	$C = WT$	[c]	8	26
p3	P81453	1	0	5	ST	[b]	5	ST	Yes	C	[b]	$C = WT$	[c]	6	27
p4	P22256	1	1	12	AT	Yes	9	AT	No	C	Yes	$N < C = WT$	6	6	28
p5	P02787	2	0	0	ST	Yes	26	AT	Yes	N	Yes	$C < N = WT$	14	11	29,30
p6	P07320	1	0	2	ST	Yes	2	ST	[b]	C	[b]	$N = WT$	11	[c]	31
p7	G9F1Y9	3	3	7	ST	[b]	20	AT	[b]	N	Yes	$C < N$	20	9	11
p8	Q47PU3	10	0	11	AT	[b]	2	ST	[b]	C	Yes	$N < C$	10	8	32
p9	B1VK30	2	0	2	ST	[b]	14	AT	[b]	N	Yes	$C < N$	10	9	33
p10	D0VWQ0	0	0	7	ST	[b]	38	AT	[b]	N	Yes	$C < N$	6	6	34,35
p11	G8LK72	1	3	10	AT	[b]	11	AT	[b]	N	Yes	$C < N$	20	6	36
p12	H7C697	3	0	9	AT	[b]	86	AT	[b]	N	Yes	$C < N$	6	6	37
p13	P00722	3	0	15	AT	[b]	9	AT	Yes	C	[b]	$C < WT$	[c]	8	38
p14	Q6QGP7	0	0	3	ST	[b]	9	AT	Yes	N	[b]	$C < WT$	[c]	6	39
p15	Q8VNN2	0	0	33	AT	[b]	11	AT	Yes	C	[b]	$C < WT$	[c]	10	40
p16	P53704	1	0	51	AT	Yes	67	AT	Yes	N	Ambiguous	$C < WT > N$	9	8	41
p17	P04537	0	2	14	AT	Yes	12	AT	Yes	C	Ambiguous	$C < WT < N$	14	10	42
p18	Q93VR3	12	2	15	AT	Yes	27	AT	[b]	N	[b]	$N < WT$	21	[c]	43,44
p19	Q9HV14	0	1	13	AT	Yes	9	AT	[b]	C	[b]	$N < WT$	22	[c]	14
p20	Q5HH30	0	3	9	AT	Yes	9	AT	[b]	C	[b]	$N < WT$	24	[c]	14
p21	Q9KL03	2	0	14	AT	Yes	10	AT	[b]	C	[b]	$N < WT$	24	[c]	14
p22	A0A089Q240	1	0	0	ST	No	2	ST	[b]	N	[b]	$N < WT$	21	[c]	45
p23	A0A0D8BQX7	2	0	0	ST	No	4	ST	[b]	N	[b]	$N < WT$	21	[c]	46
p24	A0A892IHP6	1	1	2	ST	[b]	19	AT	[b]	N	No	$N < C$	21	9	12
p25	G0SGU4	0	1	2	ST	[b]	11	AT	[b]	N	No	$N < C$	32	6	47
p26	O46414	5	4	9	AT	[b]	21	AT	[b]	N	No	$N < C$	50	13	10
p27	P40013	7	220	2	ST	No	2	ST	[b]	Equal	[b]	$WT < N$	19	[c]	48
p28	P59807	0	0	2	ST	No	0	ST	[b]	C	[b]	$N < WT$	19	[c]	49
p29	Q55080	0	1	12	AT	Yes	0	ST	No	C	No	$C < N < WT$	20	16	50
p30	Q816 V1	0	0	9	AT	Yes	28	AT	Yes	N	Yes	$C < N < WT$	20	22	51
p31	Q84EK3	5	9	0	ST	No	2	ST	[b]	N	[b]	$N < WT$	20	[c]	52−55[d]

[a]A large risk value (>8) for a terminus is predictive of AT status, while values below 9 are predictive of ST. [b]Experimental data is insufficient to verify Terminator's predictions for this category. AT/ST verification requires comparing the performance of a protein tagged at the relevant terminus to wild-type, while verification of terminus preference requires comparison of proteins tagged at each terminus. Ambiguous: both tags alter activity differently depending on the terminus. [c]Terminus was not modified in publication. [d]Publication discusses three identically tagged proteins identified by Terminator as sibling entries (>99.4% identity), therefore sharing a single analysis. p1 is included as an example of poor terminal coordinate assignment producing incorrect predictions but is omitted from further statistical analysis. A file with complete Terminator outputs for this set of proteins is available on our GitHub (github.com/MinteerLab/Terminator).

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

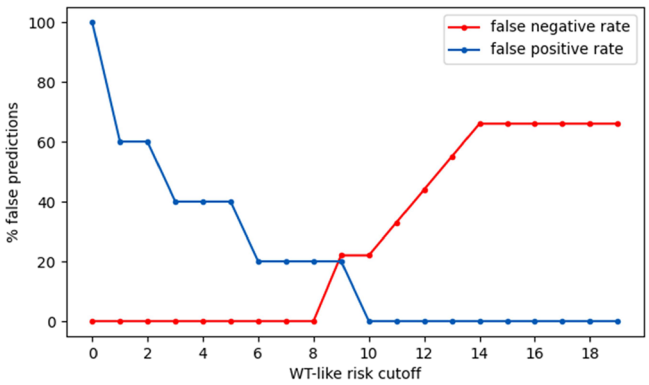


Figure 2. Comparison of false positive and false negative rates at a variety of risk cutoffs for 14 proteins, which were experimentally tested with His-tags shorter than 15 residues in published works. Risk cutoff referring to the highest value of risk for a terminus with which it is still assigned ST status. False positive refers to predicting AT for an experimentally demonstrated ST, while false negative refers to predicting ST for an experimentally demonstrated AT. At the selected cutoff value of 8, true positives: [p4N, p5C, p13C, p14C, p15C, p16N, p16C, p17N, p17C]; no false positives; true negatives: [p3C, p5N, p2C, p6N]; false negatives: [p4C].

a different structural class from the other proteins, as discussed in greater detail in Discussion Section III, *example 4*. If users refrain from analyzing this category of proteins with Terminator, this data set demonstrates a 100% accuracy in selecting the safer terminus (6/6 proteins) and a 92% (11/12 termini) accuracy in assigning individual AT/ST status to termini for proteins with fusion tags shorter than 15 residues. Although it may appear that this refers to only 6 proteins and both of their termini, the selection and AT/ST assignment data sets share only one protein (p5), with the latter set representing 13 termini from 10 proteins (Figure 3, legend).

III. Selection of Detailed Examples

Several detailed examples of Terminator analysis are provided in this section to help the reader understand the calculation of structural properties by Terminator and the correlation between Terminator outputs and the manual analysis of crystal structures.

Example 1: A Well-Studied Case. Tropinone reductase II (TrII, p9, UniProt B1VK30) from *Solanum dulcamara* is an insightful protein to study from the perspective of terminal modification, as its terminal tag effects were investigated both experimentally and with MM.[33] The MM (and the SWISS-MODEL structure analysis performed by Terminator) was conducted on a close ortholog TrII from *Datura stramonium* with a sequence identity of 92%. The experimental results demonstrate that an N-terminal modification results in catalytic activity close to wild-type, while MM demonstrates that a C-terminal His-tag fills the active site-adjacent cavity, preventing substrate entry. Terminator provides the same cavity interference prediction, as well as the convenient automatic detection of the catalytic significance of this cavity without the need to manually parse literature as may be necessary in the case of manual structure analysis, all in less than 10 min of automatic calculation with minimal computational resources and user input beyond providing a UniProt ID. Weighing the relative magnitude of each property assigned to each terminus, Terminator correctly predicts a significantly higher risk for C-terminal modification (risk values of 0 and 13 for the N- and C-termini, respectively, Table 1).

Figure 4 demonstrates a recolored output of the label_cavities script, which employs PyMOL[56] to visualize the cavities detected on this protein by Terminator. Cavities are detected by flood-filling the protein surface with pseudoatoms (as described quantitatively in the Supporting Information). The orange spheres in this figure represent the pseudoatoms assigned to a single wide cavity detected on this protein. The conservation properties for this cavity are derived from pseudoatom adjacent residues, and these properties are then weighed by the (short) distance between the C-terminus and this cavity and transferred to the C-terminus. These terminus-assigned values are used to compare the suitability of the two termini for tag addition and calculate their modification risks to determine the AT/ST status. Because no cavities were detected near the N-terminus, it is far from any chain

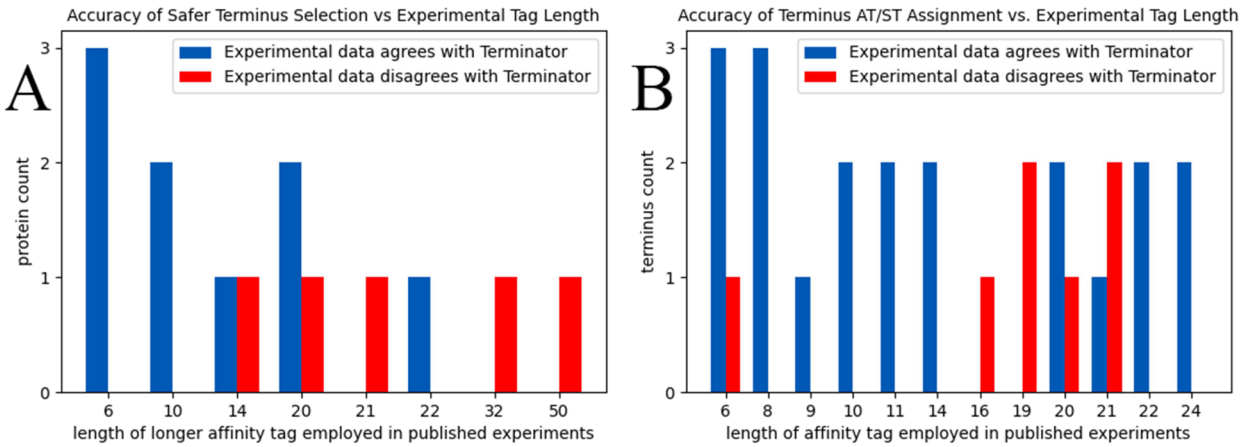


Figure 3. (A) Comparison of Terminator's selection of the safer terminus versus experimentally determined preference, organized by the length of the longer affinity tag (between the N- and C-terminus) employed in experiments. (B) Comparison of Terminator's predicted AT/ST status for individual termini versus experimentally determined status. Data points refer to (listed as length: protein) for agreeing with experimental data in A [6: p4, p10, p12; 10: p8, p9; 14: p5; 20: p7, p11; 22: p30]. For disagreeing with experimental data in A [14: p17; 20: p29; 21: p24; 32: p25; 50: p26]. Termini agreeing with experimental data in B [6: p3C, p4N, p14C; 8: p2C, p13C, p16C; 9: p16N; 10: p15C, p17C; 11: p5C, p6N; 14: p5N, p17N; 20: p29N, p30N; 21: p18N; 22: p19N, p30C; 24: p20N, p21N]. Disagreeing with experimental data in B [6: p4C; 16: p29C; 19: p27N, p28N; 20: p31N; 21: p22N, p23N].

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

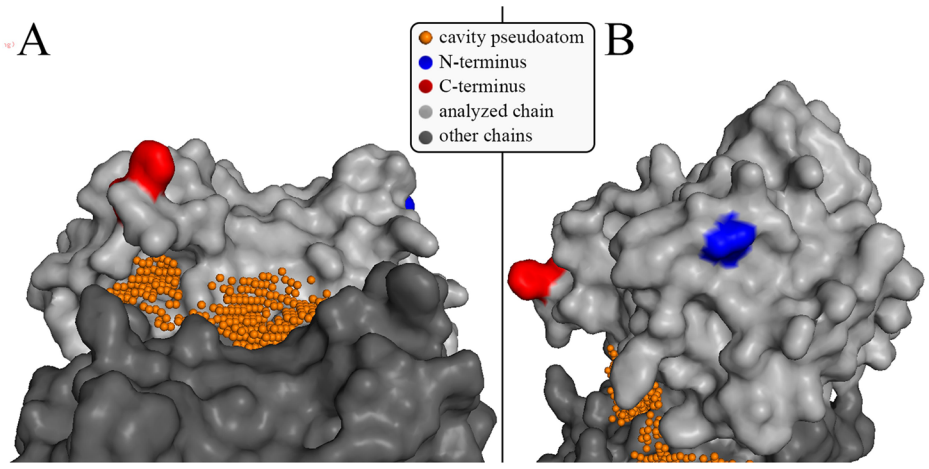


Figure 4. Crystal structure of trII from *Datura stramonium* (p9, UniProt B1VK30) (A) focused on C-terminus and identified cavity and (B) tilted down 90° and rotated left 90° to focus on the N-terminus. The structure is colored by chain (shades of gray), N-terminus (blue), and C-terminus (red). Pseudoatoms generated by Terminator to detect cavities (orange spheres), representing a single, structure spanning cleft, with the protein's active site located in the dense collection immediately below the C-terminus.

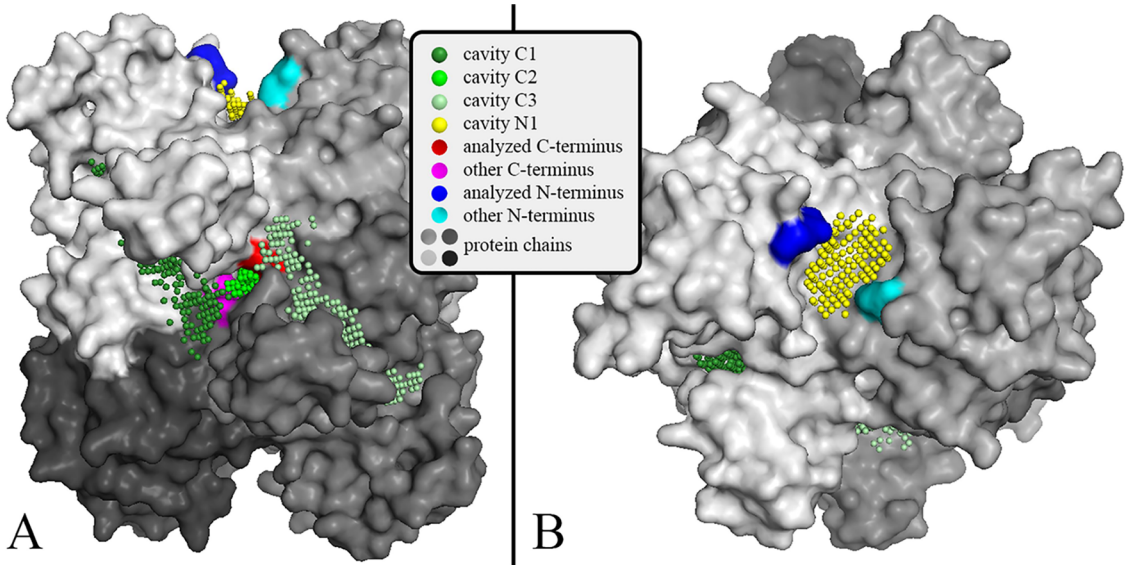


Figure 5. Crystal structure of 3-hydroxybutyrate dehydrogenase from *Alcaligenes faecalis* (p10, UniProt D0VWQ0) (A) focused on the analyzed C-terminus (red) and other adjacent C-terminus (magenta) and (B) tilted down 90° to focus on analyzed N-terminus (blue) and other adjacent N-terminus (cyan). Cavity pseudoatoms (small spheres) for the C-terminus assigned cavities C1 (dark green), C2 (lime green), and C3 (pale green). N-terminus assigned cavity N1 (yellow spheres). Protein chains are shown in shades of gray. Although only a small opening is visualized here, cavity C2 is rather large at the interior of the quaternary structure (as suggested by a large volume in Table 2). With so many avenues for structural interference at the C-terminus, it may be difficult to imagine what orientation of C-terminal 6xHis tags produces the nonzero experimental activity reported for this protein.

interfaces in the quaternary structure, and its terminal residue is relatively unburied; this terminus is assigned a modification risk of 0, correctly predicting that a 6xHis tag added to this terminus would have no negative effect on catalytic activity.

Example 2: On the Weighing of Cavities. 3-Hydroxybutyrate dehydrogenase from *Alcaligenes faecalis* (p10, UniProt D0VWQ0) is reported as having a k_{cat}/K_m value 1000× higher with an N-terminal 6xHis tag compared to a C-terminal 6xHis tag.[34] A cursory glance at a PDB structure for the protein (Figure 5) reveals that the C-terminus lies at the center of several cavities and at the junction of all four protein chains in the quaternary structure (although a new user of protein visualization software may neglect to generate the quaternary assembly required to make these observations). Terminator detects a single, small cavity only a few angstroms from the N-terminus (Table 1, cavity N1; Figure 5B, yellow spheres). The C-terminus in contrast is adjacent to two large,

Table 2. Cavity Data Calculated by Terminator for 3-Hydroxybutyrate Dehydrogenase from *A. faecalis*[a]

cavity	ConsfScore	ActiveSite score	ConservedSite score	volume	Interference probability
C1	-7.7	-15.5	0	501	0.45
C2	-2.5	-7.6	0	1008	0.75
C3	-7.9	-13.8	0	496	0.66
N1	-0.84	0	-5.1	128	0.92

[a]Rows are colored according to pseudoatom coloration in Figure 5.

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

rotationally symmetrical clefts across independent chains (cavities C1 and C3), as well as lying just at the opening of a very large cavity at the interior of the assembly (cavity C2).

Although score assignment is described in more detail in the Supporting Information, in short, the properties of assigned cavities are transferred to the relevant terminus, weighted by a tag−cavity interference probability calculated as a function of the linear distance between the terminus and the nearest cavity-assigned surface atom. Accordingly, because Terminator analyzed the red C-terminus pictured in Figure 5, the more distant cavity C1 (dark green) is assigned a lower interference probability compared to C2 (lime green) and C3 (pale green). A terminus is assigned a homonymous value for each cavity-derived property (e.g., ConsfScore), as the interference probability-weighed sum from each of its assigned cavities. These terminus-assigned properties are compared directly between termini for selecting the terminus with a lower AT character and are modified with independent normalization functions to calculate an overall risk of modification used to assign an AT/ST character to the terminus. In accordance with this C-terminus experimentally demonstrating catastrophic effects on activity when modified with a 6xHis tag, Terminator calculates an unusually high modification risk value of 38, four times higher than the minimum value corresponding to an AT (risk = 9).

Example 3: A More Subtle Case. 4-Aminobutyrate aminotransferase (GabT, p4, UniProt P22256) from *Escherichia coli* catalyzes the deamination of 4-aminobutyrate to succinic semialdehyde and subsequent amination of α-ketoglutarate to L-glutamate.[57] This enzyme was shown to have near-wild-type activity with a C-terminal 6xHis tag and close to half of this activity with an N-terminal 6xHis tag.[28] The behavior of this protein is difficult to predict based on manual structural analysis because both termini are nearly equidistant from the active site cavity (Figure 6). Terminator correctly predicts that the N-terminal His-tag produces lower activity than the C-terminal (Table 1) based on a marginally smaller distance (3 Å) between the N-terminus and the active site cavity compared to the C-terminus and the same cavity, as well as a higher quaternary interface proximity. However, due to its proximity to the active site cavity, Terminator calculates a risk value of 9 for this C-terminus, the only experimentally demonstrated ST predicted to be an AT by Terminator (the only false-positive result).

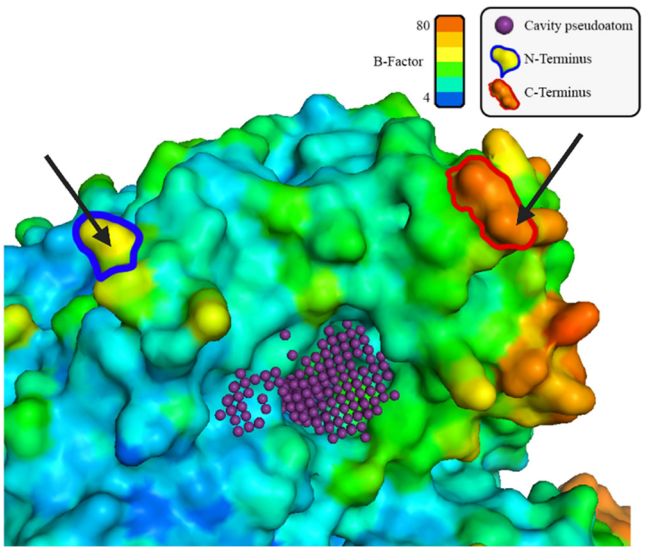


Figure 6. GabT from *E. coli* (p4, UniProt P22256) colored by the B-factor. N- and C-termini outlined in blue and red, respectively, and pointed to with arrows. Cavity pseudoatoms generated by Terminator indicating the location of the active site (purple spheres). Notice the generally lower B-factor (blue to cyan surfaces) surrounding the N-terminus, compared to the higher B-factors (green to orange) surrounding the C-terminus.

It is unclear from structural analysis what may be responsible for the significant difference in impact on catalytic activity between tags at the two termini. One hypothesis is that the shallow cavity area closer to the N-terminus is instrumental in stabilizing a cavity-interfering conformer of an N-terminal tag. Another option is that the quaternary interface proximity of the N-terminus is significant enough to damage the quaternary assembly, leading to catalytic issues. Finally, the issue could lie in the higher B-factor of the C-terminus (Figure 6), such that a tag attached there would simply have a significantly higher entropic cost of approaching or depositing near the cavity. Higher B-factors correspond to higher variance in atom position, with very high B-factors often denoting labile regions, such as hinges and motile lid structures.[58] B-factor is not considered in the current version of Terminator but is an important factor to incorporate in future versions. This protein serves as a warning to parse the Terminator output more rigorously on the rare occasion when termini receive very similar risk counts, especially when the "cavity" row reveals that the script has identified the same cavity for both termini with only a small difference in interference probability.

Example 4: An Unusually Complex Case. Recombinase uvsY from bacteriophage T4 (p17, UniProt P04537) was demonstrated experimentally to have higher DNA amplification activity with a 14 residue N-terminal 6xHis tag than either C-terminally tagged (12 residues) or untagged protein.[42] The analysis of p17 with Terminator is problematic for three reasons. First, because Terminator predicts structural interference and assumes that this will lead to negative effects on activity (the only other activity-improving tag being p27N) when Terminator recognizes more significant avenues for interference at the N-terminus of p17, it incorrectly selects a preference for the C-terminus, whose structural interference incidentally leads to a loss rather than an increase in activity. Second, the analysis of p17 is further complicated by the asymmetrical quaternary structure of this protein, which can produce radically different results based on *which* N-terminus in the quaternary structure is analyzed (Figure 7). While Terminator selects for analysis whichever chain has the most terminal residue assigned coordinates, it does not currently analyze multiple sequence-equivalent chains for different results. Finally, the third reason p17 may not be appropriate to analyze with Terminator is that this protein's active functional form involves a heteromeric assembly with DNA and a single-stranded DNA-binding protein gp32, replaced in the course of activity with recombinase uvsX.[59,60] This is a complex, apparently cavity-independent mechanism outside the scope of Terminator's capacity to analyze, particularly with only the homoheptameric p17 structure available at the time of this work's publication. Thus, it may be argued that p17 does not fairly represent Terminator's capacity to predict the behavior of more common and simpler enzyme structures, and users are warned to be cautious when employing Terminator

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

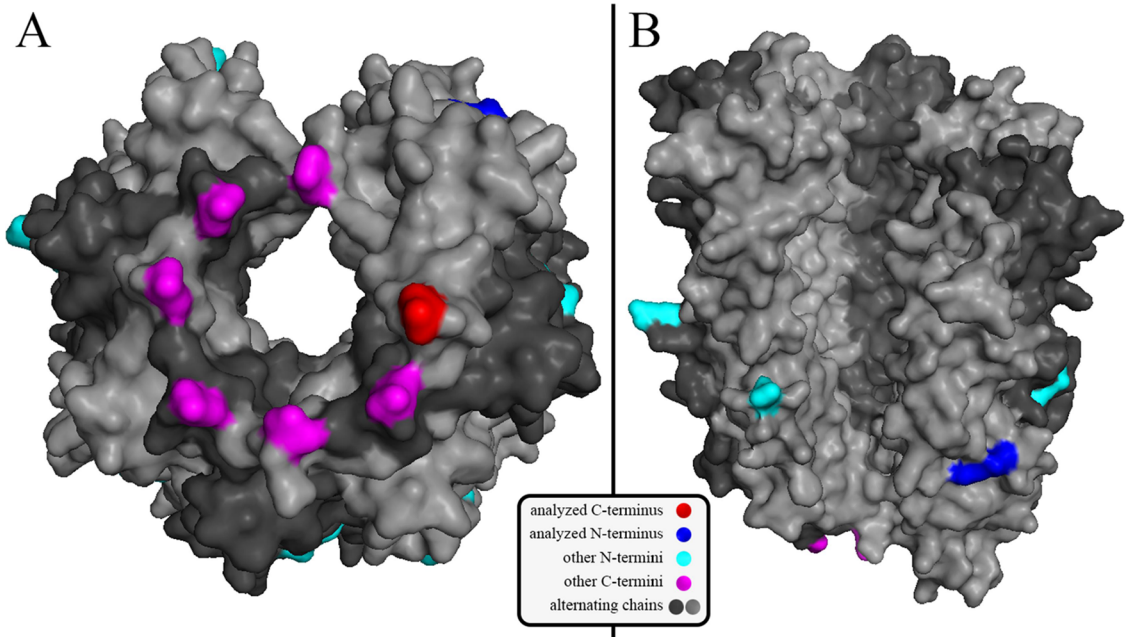


Figure 7. Structure of uvsY from bacteriophage T4 (p17, UniProt P04537, PDB entry 4ZWQ) (A) focused on C-termini and (B) tilted down 90° (C-termini now on bottom of structure) to focus on N-termini. N- and C-termini selected for analysis by Terminator in blue and red, respectively. Other N- and C-termini in cyan and magenta, respectively. Alternating protein chains in shades of gray. Notice the significantly different local environment of the "other" termini relative to those analyzed, being further away from the lengthwise opening of the quaternary structure. The hypothesized catalytic mechanism involves gp32-bound DNA wrapping around the top end of the barrel structure in panel (B), opposite of the bottom region dense with C-termini.

to analyze heteromeric proteins or those that otherwise do not employ traditional active site cavities for activity.

Limitations

Besides the limited range of Terminator's search algorithm and requirement for thorough coordinate assignment of terminal residues in available crystal structures as described in Discussion Section I, Terminator has several other limitations that users must carefully consider before employing the program to analyze a protein of interest. First, the calculation of property inheritance from cavities: In the case of small, deep cavities appearing in many proteins, any transient interaction between the tag and the surface of an active site cavity can be expected to be detrimental to substrate−active site interaction. However, when Terminator detects a very large cleft in proximity to a terminus, with the active site tunnel far outside the reach of said terminus, the terminus nevertheless inherits the properties of the entire cleft and tunnel weighed only by the cleft−terminus distance, even when the highly conserved tunnel is far outside the reach of a fusion tag. This can be addressed in a future update by placing an upper limit on cavity volume, thus subdividing a cavity into independent, interacting objects. Users are thus advised to more carefully consider Terminator-calculated properties and supplement analysis with manual examination of the crystal structure when very large cavity volumes (>700) are presented in the cavities row of the output file. Second, Terminator struggles when evaluating rare crystal structures that encode multiple chains in separate quaternary assemblies presented as a single asymmetrical unit, which may be verified by simply glancing at the relevant structure in visualization software. Third, in certain proteins, active site cavities may alternate between open and closed states[16]; hence, Terminator may fail to accurately predict tag behavior in proteins if no open-state structures have been determined. Fourth, as discussed in Example 4 above, because Terminator predicts *structural interference*, it is not able to distinguish more common detrimental structural interactions from very rare, beneficial interactions. Furthermore, because Terminator relies mostly on cavity detection, it is not likely to accurately analyze cavity-independent mechanisms, such as those employing certain heteromeric complexes, particularly in enzymes that catalyze reactions with DNA or polypeptide chains. Expanding further on this example, because Terminator only performs conservation analysis on a single protein, it is not currently able to calculate conservation-based properties for heterocomplexes of different proteins even when such crystal structures are available.

CONCLUSIONS

This work introduces a fast and accessible tool integrating protein sequence conservation analysis with structural data to predict the most suitable terminus for placement of short (<15 residues) 6xHis fusion tags. When comparing Terminator predictions to published experimental data for proteins modified with 6xHis tag sequences totaling less than 14 residues, Terminator demonstrates an 86−100% accuracy in predicting the safer terminus for tag addition and a 92−93% accuracy in predicting the risk of activity-modifying structural interactions, resulting from the addition of fusion tags to specific termini. Users are advised to avoid analyzing crystal structures with a significant number of terminal residues lacking coordinate assignment (>3 residues) and to supplement Terminator with manual analysis if detected cavities demonstrate volumes above 700 Å^3 or when termini are assigned the same cavity with small variation in interaction probability. The minimal need for time investment, domain knowledge, and computational resources is the main advantage of Terminator over manual crystal structure interpretation and

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65

molecular modeling of tag effects. Critically, Terminator can be retroactively employed to analyze enzymes in published biocatalytic systems to easily improve the performance of these systems by employing proteins with better-behaved tags. The speed and ease of employing Terminator will expedite construct design and enhance the successful production of well-behaved proteins. Terminator eliminates the labor and time required to purify and compare the relative performance of a protein tagged at either terminus, thus accelerating the publication of valuable research in the fields of biocatalysis and enzymology.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsbiomedchemau.4c00055.

User guide covering installation, implementation, and proper use of Terminator and expanded quantitative description of the algorithms involved in terminus selection and evaluation (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Shelley D. Minteer − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; Kummer Institute Center for Resource Sustainability, Missouri University of Science and Technology, Rolla, Missouri 65409, United States;* orcid.org/0000-0002-5788-2249; Email: shelley.minteer@mst.edu

Author

Rokas Gerulskis − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsbiomedchemau.4c00055

Author Contributions

CRediT: **Rokas Gerulskis** conceptualization, data curation, formal analysis, validation, writing - original draft; **Shelley D. Minteer** funding acquisition, project administration, supervision, writing - review & editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research (ONR) for funding this research (Grant No.: N000142114008).

■ REFERENCES

(1) Bornscheuer, U. T.; et al. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185−194.

(2) Tabor, S.; Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. 1985. *Biotechnology* **1992**, *24*, 280−284.

(3) Dubendorf, J. W.; Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **1991**, *219*, 45−59.

(4) Pyser, J. B.; Chakrabarty, S.; Romero, E. O.; Narayan, A. R. H. *State-of-the-Art Biocatalysis. ACS Cent. Sci.* **2021**, *7*, 1105−1116.

(5) Hughes, R. A.; Ellington, A. D. Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* **2017**, *9* (1), No. a023812.

(6) The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523−D531.

(7) Terpe, K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **2003**, *60*, 523−533.

(8) Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **2005**, *41*, 98−105.

(9) Kapust, R. B.; Waugh, D. S. *Escherichia coli* maltose-binding protein is effective in promoting solubility. *Protein Sci.* **1999**, *8*, 1668−1674.

(10) Zhang, H.; et al. Soluble expression and purification of recombinant bovine ferritin H-chain. *Protein Expr. Purif.* **2023**, *211*, No. 106340.

(11) Mordaka, P. M.; Hall, S. J.; Minton, N.; Stephens, G. Recombinant expression and characterisation of the oxygensensitive 2-enoate reductase from *Clostridium sporogenes*. *Microbiol. (United Kingdom)* **2018**, *164*, 122−132.

(12) Alpdağtaş, S.; Çelik, A.; Ertan, F.; Binay, B. DMSO tolerant NAD(P)H recycler enzyme from a pathogenic bacterium, *Burkholderia dolosa* PC543: effect of N-/C-terminal His Tag extension on protein solubility and activity. *Eng. Life Sci.* **2018**, *18*, 893−903.

(13) Halliwell, C. M.; Morgan, G.; Ou, C. P.; Cass, A. E. G. Introduction of a (poly)histidine tag in L-lactate dehydrogenase produces a mixture of active and inactive molecules. *Anal. Biochem.* **2001**, *295*, 257−261.

(14) Majorek, K. A.; Kuhn, M. L.; Chruszcz, M.; Anderson, W. F.; Minor, W. Double trouble - Buffer selection and his-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.* **2014**, *23*, 1359−1368.

(15) Tanner, A.; Bornemann, S. *Bacillus subtilis* yvrK is an acid-induced oxalate decarboxylase. *J. Bacteriol.* **2000**, *182*, 5271−5273.

(16) Just, V. J.; et al. A Closed Conformation of *Bacillus subtilis* Oxalate Decarboxylase OxdC Provides Evidence for the True Identity of the Active Site. *J. Biol. Chem.* **2004**, *279*, 19867−19874.

(17) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129−1143.

(18) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(19) Waterhouse, A.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.

(20) Sayers, E. W.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20−D26.

(21) Celniker, G.; et al. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.* **2013**, *53*, 199−206.

(22) Pastore, A. J.; et al. Oxalate decarboxylase uses electron hole hopping for catalysis. *J. Biol. Chem.* **2021**, *297*, No. 100857.

(23) Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: An algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **2001**, *307*, 447−463.

(24) Chang, A.; et al. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **2021**, *49*, D498−D508.

(25) Fan, C.; Bobik, T. A. Functional characterization and mutation analysis of human ATP:Cob(I)alamin adenosyltransferase. *Biochemistry* **2008**, *47*, 2806−2813.

(26) Smits, S. H. J.; Mueller, A.; Grieshaber, M. K.; Schmitt, L. Coenzyme- and His-tag-induced crystallization of octopine dehydrogenase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2008**, *64*, 836−839.

(27) Marx, C. K.; Hertel, T. C.; Pietzsch, M. Purification and activation of a recombinant histidine-tagged pro-transglutaminase

after soluble expression in *Escherichia coli* and partial characterization of the active enzyme. *Enzyme Microb. Technol.* **2008**, *42*, 568−575.
(28) Meng, L.; et al. Effects of His-tag on Catalytic Activity and Enantioselectivity of Recombinant Transaminases. *Appl. Biochem. Biotechnol.* **2020**, *190*, 880−895.
(29) Mason, A. B.; et al. Expression, purification, and characterization of recombinant nonglycosylated human serum transferrin containing a C-terminal hexahistidine tag. *Protein Expr. Purif.* **2001**, *23*, 142−150.
(30) Mason, A. B.; et al. Differential effect of a His tag at the N- and C-termini: Functional studies with recombinant human serum transferrin. *Biochemistry* **2002**, *41*, 9448−9454.
(31) Lin, Y. W.; Ying, T. L.; Liao, L. F. Molecular modeling and dynamics simulation of a histidine-tagged cytochrome b 5. *J. Mol. Model.* **2011**, *17*, 971−978.
(32) Parshin, P. D.; et al. Effect of His6-tag Position on the Expression and Properties of Phenylacetone Monooxygenase from *Thermobifida fusca*. *Biochem.* **2020**, *85*, 575−582.
(33) Freydank, A. C.; Brandt, W.; Dräger, B. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins Struct. Funct. Genet.* **2008**, *72*, 173−183.
(34) Yeon, Y. J.; Park, H. J.; Park, H. Y.; Yoo, Y. J. Effect of His-tag location on the catalytic activity of 3-hydroxybutyrate dehydrogenase. *Biotechnol. Bioprocess Eng.* **2014**, *19*, 798−802.
(35) Yeon, Y. J.; Park, H. Y.; Yoo, Y. J. Enzymatic reduction of levulinic acid by engineering the substrate specificity of 3-hydroxybutyrate dehydrogenase. *Bioresour. Technol.* **2013**, *134*, 377−380.
(36) Zhang, J.; Cui, T.; Li, X. Screening and identification of an *Enterobacter ludwigii* strain expressing an active β-xylosidase. *Ann. Microbiol.* **2018**, *68*, 261−271.
(37) Li, Y.; et al. Recombinant glutamine synthetase (GS) from *C. glutamicum* existed as both hexamers & dedocamers and C-terminal His-tag enhanced inclusion bodies formation in *E. coli*. *Appl. Biochem. Biotechnol.* **2009**, *159*, 614−622.
(38) Flores, S. S.; et al. His-tag β-galactosidase supramolecular performance. *Biophys. Chem.* **2022**, *281*, No. 106739.
(39) Kutyshenko, V. P.; et al. Effect of C-terminal His-tag and purification routine on the activity and structure of the metalloenzyme, L-alanyl-D-glutamate peptidase of the bacteriophage T5. *Int. J. Biol. Macromol.* **2019**, *124*, 810−818.
(40) Nichols, E. R.; Craig, D. B. Single molecule assays reveal differences between in vitro and in vivo synthesized β-galactosidase. *Protein J.* **2008**, *27*, 376−383.
(41) Olchowy, J.; Kur, K.; Sachadyn, P.; Milewski, S. Construction, purification, and functional characterization of His-tagged *Candida albicans* glucosamine-6-phosphate synthase expressed in *Escherichia coli*. *Protein Expr. Purif.* **2006**, *46*, 309−315.
(42) Juma, K. M.; et al. Modified uvsY by N-terminal hexahistidine tag addition enhances efficiency of recombinase polymerase amplification to detect SARS-CoV-2 DNA. *Mol. Biol. Rep.* **2022**, *49*, 2847−2856.
(43) Wolucka, B. A.; Van Montagu, M. GDP-Mannose 3′,5′-Epimerase Forms GDP-L-gulose, a Putative Intermediate for the de Novo Biosynthesis of Vitamin C in Plants. *J. Biol. Chem.* **2003**, *278*, 47483−47490.
(44) Wolucka, B. A.; et al. Partial purification and identification of GDP-mannose 3″,5″-epimerase of *Arabidopsis thaliana*, a key enzyme of the plant vitamin C pathway. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14843−14848.
(45) De Almeida, J. M.; et al. Tailoring recombinant lipases: Keeping the His-Tag favors esterification reactions, removing it favors hydrolysis reactions. *Sci. Rep.* **2018**, *8* (1), 10000.
(46) Özdemir, F. İ.; Tülek, A.; Erdoğan, D. Identification and Heterologous Production of a Lipase from *Geobacillus kaustophilus* DSM 7263T and Tailoring Its N-Terminal by a His-Tag Epitope. *Protein J.* **2021**, *40*, 436−447.
(47) Esen, H.; Alpdağtaş, S.; Mervan Çakar, M.; Binay, B. Tailoring of recombinant FDH: effect of histidine tag location on solubility and catalytic properties of *Chaetomium thermophilum* formate dehydrogenase (CtFDH). *Prep. Biochem. Biotechnol.* **2019**, *49*, 529−534.
(48) Zhu, Z. C.; et al. Interactions between EB1 and microtubules: Dramatic effect of affinity tags and evidence for cooperative behavior. *J. Biol. Chem.* **2009**, *284*, 32651−32661.
(49) Chen, Z.; Li, Y.; Yuan, Q. Study the effect of His-tag on chondroitinase ABC I based on characterization of enzyme. *Int. J. Biol. Macromol.* **2015**, *78*, 96−101.
(50) Aslantas, Y.; Surmeli, N. B. Effects of N-terminal and C-terminal polyhistidine tag on the stability and function of the thermophilic P450 CYP119. *Bioinorg. Chem. Appl.* **2019**, *2019*, No. 8080697.
(51) Hyun, J.; Abigail, M.; Choo, J. W.; Ryu, J.; Kim, H. K. Effects of N-/C-terminal extra tags on the optimal reaction conditions, activity, and quaternary structure of *Bacillus thuringiensis* glucose 1-dehydrogenase. *J. Microbiol. Biotechnol.* **2016**, *26*, 1708−1716.
(52) Mosbah, H.; Sayari, A.; Bezzine, S.; Gargouri, Y. Expression, purification, and characterization of His-tagged *Staphylococcus xylosus* lipase wild-type and its mutant Asp 290 Ala. *Protein Expr. Purif.* **2006**, *47*, 516−523.
(53) Sayari, A.; Mosbah, H.; Gargouri, Y. Importance of the residue Asp 290 on chain length selectivity and catalytic efficiency of recombinant *Staphylococcus simulans* lipase expressed in *E. coli*. *Mol. Biotechnol.* **2007**, *36*, 14−22.
(54) Horchani, H.; Ouertani, S.; Gargouri, Y.; Sayari, A. The N-terminal His-tag and the recombination process affect the biochemical properties of *Staphylococcus aureus* lipase produced in *Escherichia coli*. *J. Mol. Catal. B Enzym.* **2009**, *61*, 194−201.
(55) Horchani, H.; et al. Heterologous expression and N-terminal His-tagging processes affect the catalytic properties of staphylococcal lipases: A monolayer study. *J. Colloid Interface Sci.* **2010**, *350*, 586−594.
(56) *The PyMOL Molecular Graphics System, Version 3.0*; Schrödinger, LLC., 2015.
(57) Liu, W.; et al. Crystal structures of unbound and aminooxyacetate-bound *Escherichia coli* γ-aminobutyrate aminotransferase. *Biochemistry* **2004**, *43*, 10896−10905.
(58) Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **2019**, *119*, 1626−1665.
(59) Gajewski, S.; et al. Structure and mechanism of the phage T4 recombination mediator protein UvsY. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 3275−3280.
(60) Xu, H.; Beernink, H. T. H.; Morrical, S. W. DNA-binding properties of T4 UvsY recombination mediator protein: Polynucleotide wrapping promotes high-affinity binding to single-stranded DNA. *Nucleic Acids Res.* **2010**, *38*, 4821−4833.

https://doi.org/10.1021/acsbiomedchemau.4c00055
ACS Bio Med Chem Au 2025, 5, 55−65