Using Reciprocals of Prime Numbers to label Proteins

Allen, Jont B.
`jontalle@ieee.org`

Sabharwal, Nippun
`nippuns2@illinois.edu`

Wednesday 31 December, 2025

Abstract

This paper explores a novel theoretical framework for representing protein sequences using the reciprocals of prime numbers. By examining the mathematical properties of prime number reciprocals, we investigate the potential for a unique mapping between protein sequences and their decimal expansions. The research delves into the inherent characteristics of recurring decimals and their potential applications in protein sequence representation and analysis.

1 Properties of reciprocals of primes

The reciprocals of prime numbers are always recurring decimals. For example, $\frac{1}{2} = 0.5\cdots = 0.5\overline{0}$, $\frac{1}{5} = 0.20\cdots = 0.2\overline{0}$ and $\frac{1}{7} = 0.\overline{142857}$. The $\overline{\textit{overbar}}$ is shorthand for a repeating sequence. We assume terminating decimals (*i.e., rational numbers*) to have repeating 0's at the end, to have a *cyclic-period* of 1.

2 Periodicity and Characteristics of Prime Number Reciprocals

The period length (length of the repeating part) varies with each prime number, with no obvious pattern. Both Figure and Table illustrate this relation:

Note symmetry here: a = 142857, b = 769230. Thus $((10^{11}) * b/a) = 538461, 538461.5385\cdots$ repeats after 10^6.

Thus there is a "pole-like" factor in these prime numbers since 1/a,1/b 7.000007000007000e-06 and $1.300001300001300 \times 10^6$, both of which have a period of 6.[1]

Thm: Since 7 and 17 both have the same period, their ratio must also have a period of 6. Thus 17/7 must also be periodic, with a period of 6. Obviously this leads to a theorem, that needs to be identified in literature. How about Yan,Yan and Yan (1991), who cite George Gamow (1954).

3 Theoretical Relationship to Modular Arithmetic

Determining the reciprocal prime's length can be difficult when it is more than 6 digit. The period length (or repeating block) of $1/p$ for a prime $p \neq 2, 5$ is directly related to the order of 10 modulo p, defined as the smallest positive integer e such that:

$$10^e \equiv 1 \pmod{p}$$

From *Fermat's Little Theorem*, we know $10^p \equiv p \pmod{p}$ for a prime p not dividing 10, so the period length e must always be a divisor of $p - 1$. In summary

---

[1] $625 = 5^4$ and $625 - 1 = 2^4 \times 3 \times 13$. How and why is this relevant?

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
π_n	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59
$\partial_k \pi_k \equiv \pi_{k+1} - \pi_k$	1	2	2	4	2	4	2	4	6	2	6	4	2	4	6	6	2
$\partial_k^2 \pi_k \equiv \pi_{k+1} - \pi_{k-1}$	-	1	2	0	0	0	0	2	-2	4	0	2	2	-4	2	-4	4
Period mod_{10}	-	1	1	6	2												

Table 1: Table of the first 17 primes π_n and their periods. Note we define the period as 1 since $1/1 = 1.((0))_1$, $10/2$=5.$((0))_1$, and $1/3 = 0.((3))_1$. The `Wolfram language` is the ideal tool for determining periods longer than 6 digits.

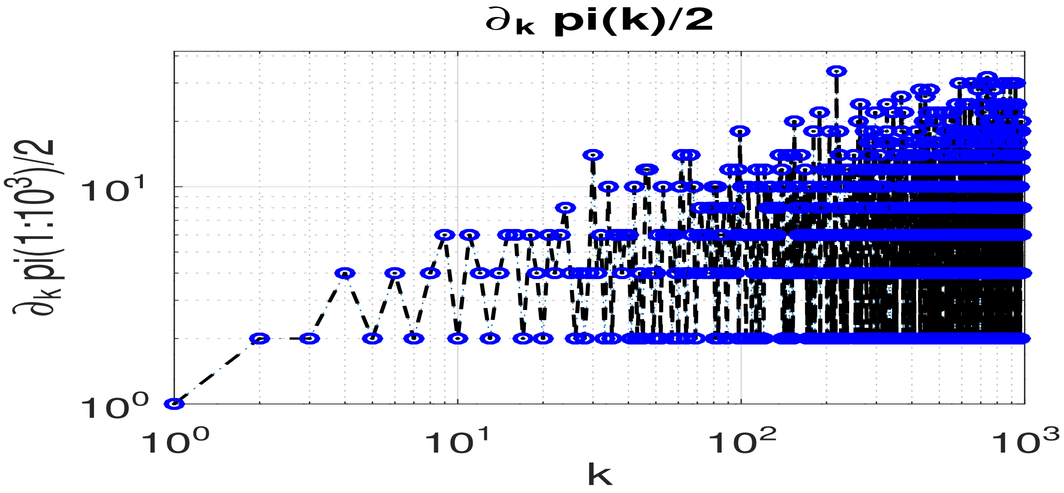


Figure 1: The differences of the first four primes $\partial_k(k)$ are 1, 2, 2, 4. From this chart we can see the frequent of all the differences, as they chaotically jump by $k = 2, 6, 8$, or more.

- 1/7 has period of 6.
- 1/17 has period of 16.
- 1/19 has period of 18.

This seems to explain why different primes can have widely varying period lengths. This is not a general rule across all primes, as explored next.

4 Diversity of Recurring Decimal Expansions

Observe that the digits of the recurring part vary. Note that both 1/7 and 1/13 have a repeating period length of 6, but with different digits. Here $10^5/7 = 142857.1428571429$ and $10^6/13 = 769230.7692307692$, both of which have a cyclic period of 6. Furthermore $10^{17}/19 = 5263157894736842$ is beyond the limits of IEEE-64 bit integer arithmetic. The Wolfram language is needed to explore this larger cyclic period greater than IEEE limits.. Also This observaton needs generalization followed by the number-theory explaination.

Since there are infinitely many primes, we may find reciprocal primes with arbitrary period length, and every distinct combination of digits. For example we can find a period length of 6 with any combination of the 9 digits, say 012345 or 56789 or 142857, and so on. Looking for patterns is a lifetime's work, and already been explored by many famous mathematicians. These questions have been around, but still unresolved, from the beginning of time.

Table 2: For prime 37, to obtain the sequence, we need to expand $10^4/37 - 270.270270\cdots$. The process fails with the example 1e4/37-270 = 0.270270270270260 due to the IEEE-754 standard limitations. To avoid these problems one would need to work in Python or perhaps better, the Wolfram language.

Prime (p)	Period Length	Reciprocal (1/p)
7	6	$((0.142857))_6$
13	6	$((0.769230, 769230, \cdots))_6$
37	3	$((0.270, 270, 270, 270, 270, \cdots))_3$

Note that primes exist with any **combination** of digits, but not for every **permutation**. This is because as you take larger prime numbers, the reciprocal becomes smaller, This may fail for very large primes For instance, all possible reciprocals of primes can only have the numbers 0, 1, 2, 5 and 5 in the tenths place. The expansion of [2] $13/17 \times 1e17 = 7.647, 058, 823, 529, 411, \cdots \times 10^{16}$ is interesting since both 13 and 17 must have the same period of 6. Assuming this generalizes, it potentially leads to another theorem.

Prime (p)	Reciprocal (1/p)
2	0.5
3	0.3
5	0.2
7	$((0.142857))_6$
11	0.09
13	$((0.076923))_6$
17	0.058,823,529,411,7647,1

Table 3: Only numbers 2, 3, 5 and 7 have non-zero digits in the tenths place. Bast on an experiment using an Octave code, this seems to disagree with the period for 1/17 ($100/17 = 5.882, 352, 941, 176, 471,$), which is not periodic. However $1/16 = 2^{-4}$. This leads to an interesting conclusion, since 17/7 must have a period of 6.

5 Other Phenomena in Decimal Expansions

The decimal expansions showcase both variety and regularity:

- **Cyclic Numbers:** A cyclic number is an integer for which cyclic permutations of the digits are successive integer multiples of the number. A prime p for which $1/p$ attains a period length of $p - 1$ may yield a cyclic number, such as 1/7.
- **Midy's Theorem:** If the period length of $1/p$ is even, then splitting the repeating block into two equal parts yields halves that sum to a string of 9s (e.g., 142+857=999 for $p = 7$).

6 Protein Representation Hypothesis

Our claim is that every protein can be related to the reciprocal of a prime, with this relation being unique. Proteins are long chains of amino acids that are sequences of molecules. Note that each

---

[2] The location of the decimal point needs further analysis.

amino acid is a discrete unit, commonly encoded with symbols or numerical indices. Proteins have finite lengths (50, 100, or 700 amino acids). The linear combination of 20 possible amino acids leads to an astronomical number of possible sequences. Minor sequence changes can significantly alter protein functionality.

Table 4: Combination of digits in reciprocals Prime (p) and Reciprocal (1/p)

prime p	period of $1/p$
2	0.5
3	0.3
5	0.2
7	0.142857
11	0.09
13	0.0769230
17	0.05882352941176471

Thus, every element of the sequence can be represented as a number, creating a sequence with:

- A finite length that may or may not be repeated
- Any combination of amino acids/molecules in that sequence

In this way we can create a one-to-one mapping of each protein to a prime number

Table 5: Only numbers 2, 3, 5 and 7 have non-zero digits in the tenths place. Using OCTAVE, the period of 17 is 16 ($10^k/17 = 5.882,352,941,176,471$) for a large range of integers k. Thus 1/17 has a period of 16. These numbers seem consistent with *octave* integer calculations, such as OCTAVE: 10000000/13/769230-1, which has a period of 16, since $100/13 = 10^9/17/13$ ANS $=$ 4524886.877828054. Thus 100/17 has a period 100/13 =7.69230,769230,7693 of 6. This says that 1/13 has a period 769230, 6 digits long. For the prime 17, 1/17 has a period of 16. We may verify this using octave. If we let $R = 10^k/17 = 5.882,352,941,176,471$, R is independent of integer k over a range of $0 \leq k < 10^9$, meaning $1/17 = 10^k$/1,000,000,000,000,000 `=1,000,000,000,000,000=$10^{16}$`, which has a period of 16, as predicted from theory. Subtracting 1, gives 0 to 16 decimal places, as predicted by 16=17-1. Using the Wolfram language one may explore larger ranges for the primes greater than 17. Periods of primes are also available at `https://en.wikipedia.org/wiki/Reciprocals_of_primes`, which has `10/17=0.588,235,294,117,647`, which has a small error, since *the last digit is 1 is missing.*

However on further checking, I also realize that the order of protein structures matter, i.e. $ABCD \neq ABDC$.

As established in the last section we cannot generate every permutation of primes, just every combination. Thus suppose the protein A corresponds to number 0, B to 1, C to 2, D to 3, E to 4

Then how will we map a protein that has a sequence EAFC if we do not have any reciprocal that has 4 in the tenths place i.e. starts 0.4... and continues from there on?

7 Encoding Method

Since there are 20 amino acids, we propose a two-letter encoding for amino acids:

- Alanine (A): 00
- Cysteine (C): 01
- Aspartic acid (D): 02
- ...
- Tryptophan (S): 19
- Tyrosine (T): 20

Example:

- Protein sequence: “ACGLST”
- Encoded digit string: “00:02:07:12:19:20”

8 Reciprocals of the products of primes.

We can expand this analysis to the products of reciprocals of prime. Here assume x, y are each a prime. Then

$$\frac{1}{x*y} = \frac{a}{x} + \frac{b}{y}$$

for some constants x and y
for example, $\frac{1}{21} = \frac{a}{7} + \frac{b}{3}$ for $(a, b) = (a, \frac{-7a+1}{2})$

Thus,

$$\frac{1}{xy} = \frac{a}{x} + \frac{b}{y} = \frac{ay+bx}{xy}$$

That means the **numerators** must be equal since:

$$ay + bx = 1$$

But this equation **depends on both** x and y.

Solution i) $a \neq 0, \quad y = \frac{1-bx}{a}, \quad -x + bx^2 \neq 0$

Solution ii) $a = 0, \quad b \neq 0, \quad x = \frac{1}{b}, \quad y \neq 0$

9 Question: For the reciprocal of any cycling sequence of digits, does one always obtain a prime?.

This is bound to fail since the sequence must repeat forever, which is not possible with Matlab/octave. The reciprocal of every prime is a repeating sequence, but the reverse is NOT.

Example: *works:* 1e6/7 = 142857.142857,... $((142857))_6$ while 1/123.123123123123123 fails OCTAVE: 1/125125125127127129, ANS = 7.9919999999872128E-18 . Exact: octave: 1/8 = 0.125000000000000 fails, since 1/0.125125125125125; ans = 7.992000000000008

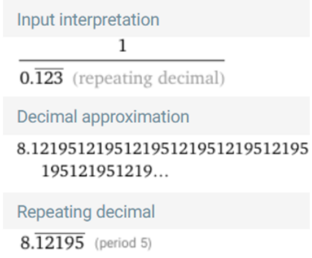
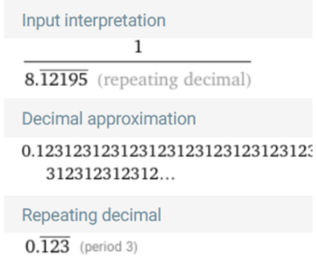
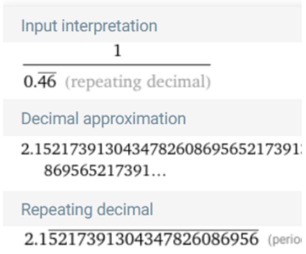
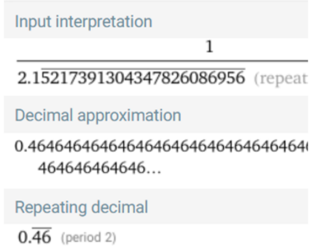
We may need a recursive algebra program such as Wolfram language, that repeats the sequence, and then observe when (assuming) the number becomes an integer, and finally verify that the integer is prime, by factoring it.

Many issues: 1) its may not true. 2) easy to show with a simple example. 3) factoring large number is very difficult as it is computationally time consuming, and for large enough numbers, likely not even possible. Thus the method fails to converge give a simple example.

Second try: We can cook up a repeating sequence and take its reciprocal, and we should get a prime number. However we have not tried this.

But upon testing it, we do not necessarily get prime numbers, we may get recurring numbers as outputs, but seldom get integer prime numbers. What is needed is a computer program that does the algebra using strings rather than numbers.

Summary: The reciprocal of every prime is a repeating decimal (reptend), but the reciprocal of every periodic decimal is not a prime. The size of repeating decimals is much larger than the space of reciprocal primes.

Repeating Sequence	**Reciprocal**	**Reciprocal of reciprocal (original number)**
0.123	 Input interpretation $\frac{1}{0.\overline{123}}$ (repeating decimal) Decimal approximation 8.12195121951219512195121951219512195 195121951219... Repeating decimal $8.\overline{12195}$ (period 5)	 Input interpretation $\frac{1}{8.\overline{12195}}$ (repeating decimal) Decimal approximation 0.12312312312312312312312312312312312 312312312312... Repeating decimal $0.\overline{123}$ (period 3)
0.46	 Input interpretation $\frac{1}{0.\overline{46}}$ (repeating decimal) Decimal approximation 2.152173913043478260869565217391 869565217391... Repeating decimal $2.1\overline{5217391304347826086956}$ (perio	 Input interpretation $\frac{1}{2.1\overline{5217391304347826086956}}$ (repeat Decimal approximation 0.4646464646464646464646464646464646 464646464646... Repeating decimal $0.\overline{46}$ (period 2)

10 Potential Theoretical Applications (Speculative)

Representing a protein as a reciprocal of a prime number can be useful, since it unlocks the properties of prime numbers for these proteins. This taxonomy or nomenclature opens up the possibility of performing operations with the reciprocals of primes. Below are a few notable prime number properties and number-theoretic constructs that could, at least theoretically, be leveraged if each protein were uniquely represented by (or associated with) a prime number.

While much of this remains speculative and would require concrete mathematical and bioinformatic frameworks, these properties hint at the depth of potential connections between prime-based representations and biological sequences.

10.1 Cyclic Groups and Modular Arithmetic

If a protein is tied to a prime p, the order of 10 modulo p (i.e., the smallest integer k such that $10^k \equiv 1 \pmod{p}$) becomes relevant. Each prime p generates a multiplicative group $\{1, 2, \ldots, p-1\}$ under multiplication modulo p. This group is cyclic, meaning there exists a generator g such that every element can be expressed as $g^k mod p$ for some integer k.

Potential Application: If a protein X is mapped to prime p, its sequence could be explored through the lens of generators and modular orders. This may reveal "symmetries" or "cycles" in the amino acid arrangement (especially when amino acids are encoded as digits or in other modular schemes).

10.2 Cyclic Numbers and Repeating Decimals

Primes p such that $1/p$ has a repeating decimal of length $p-1$ yield cyclic numbers. For example, $1/7 = 0.\overline{142857}$ is a cyclic number: multiplying 142857 by any number from 1 to 6 results in a cyclic permutation of the digits.

Potential Application: If a protein is represented by a prime whose reciprocal is a cyclic number, its repeating decimal block could be used to generate related sequences—possibly mimicking alternative folding, conformational variants, or functional analogs that preserve certain biochemical motifs.

10.3 Midy's Theorem and Digit Sums

Midy's Theorem states that if the decimal expansion of $1/p$ has an even-length period $2d$, splitting the repeating block into two equal parts of d digits yields numbers whose sum is a string of d digits of $(b-1)$ (where b is the base). For example, with $p=7$ in base 10: $0.\overline{142857}$ splits into 142 and 857, and $142+857=999$.

Potential Application: If a protein's prime-associated decimal expansion has a suitable period, Midy's Theorem might provide a natural checksum or error-detecting scheme. This could enable internal integrity checks within a prime-based protein labeling system.

10.4 Prime Distributions

Prime numbers become less dense as they increase, but they follow well-characterized asymptotic trends (e.g., the Prime Number Theorem). Prime gaps and distributions can thus inform how many proteins can be encoded within a given range.

Potential Application: If proteins are assigned to primes for cataloging, understanding the distribution of primes could help optimize allocation strategies—for example, grouping proteins by functional class, size, or family within specific prime ranges.

10.5 Prime Factorization and Composite Intermediates

Operations like $\alpha \cdot (1/p_1) + \beta \cdot (1/p_2)$ usually yield rational numbers with composite denominators. Decomposing such numbers into prime factors may reveal interesting combinatorial or relational structures.

Potential Application: Suppose two proteins A and B are tied to primes p_1 and p_2. Then $1/p_1 \pm 1/p_2$ yields a new rational number, potentially corresponding to a third "intermediate" entity. Analyzing the prime factorization of the resulting denominator might reveal a conceptual or functional link between A and B—potentially forming a mathematical network of protein interactions.

10.6 Higher-Base Representations

The approach generalizes naturally to other bases. For instance, encoding proteins in base 20 (one symbol per amino acid) might align more directly with biological reality.

Potential Application: In base 20, the repeating sequences of $1/p$ may better match amino acid alphabets. Modular orders and period lengths in base 20 could reflect or even help discover new biological motifs and periodicities that are less visible in base 10.

11 Challenges and Limitations

Despite the theoretical richness, practical challenges remain:

- **Varying Protein Lengths:** Real proteins range from tens to thousands of amino acids, while decimal expansions of $1/p$ may not align cleanly with desired sequence lengths.
- **Encoding More Than 10 Letters:** There are 20 standard amino acids but only 10 decimal digits. Directly mapping digits to amino acids requires compressing or expanding the representation (e.g., using multiple digits or better, letters, per amino acid).
- **Distribution of Period Lengths and Digit Sequences:** Not all primes yield useful or sufficiently long repeating patterns. Even when period lengths match, the resulting digit sequences might lack biological plausibility or coverage.

12 Conclusion

This theoretical framework offers a novel mathematical perspective on protein sequence representation, potentially opening new avenues for computational biology and bioinformatics research.

Citations need to be researched and added to the document future.

I would like to contact the following people who have studied this problem. along with Ellen Zhong (computational Biologist @ Princeton Univ.) and John Jumper (Theoretical Chemist @ DeepMind). For example see:
https://www.youtube.com/watch?v=cx7l9ZGFZkw. Also is the “protein databank”
The following URL *https://www.youtube.com/watch?v=cx7l9ZGFZkw*
is the brief inside story of how David Baker, Demis Hassabis and John Jumper won the 2024 Nobel Prize in Chemistry for advances in computer-assisted protein design and structure prediction.
See: In work with Alex Eskin of the University of Chicago, Dr. Mirzakhani[3]. examined billiards tables of more complicated shapes, and in fact considered the dynamics of balls bouncing around all possible tables that fit certain criteria.
http://www.rcsb.org/pdb/

Anastassis Perrakis, Structural Biologist Netherlands Cancer Institute.
`https://www.youtube.com/watch?v=cx7l9ZGFZkw`

This problem is similar to, if not idential to, the *image model*

Comment: Thm:
Since 7 and 17 both have the same period, their ratio must also have a period of 6. Thus 17/7 must also be periodic, with a period of 6. Obviously this leads to a theorem, that needs to be identified in the literature.

References

Allen, J. B. and Berkley, D. A. (1979) ”Image method for efficiently simulating small-room acoustics”; J. Acoust. Soc. Am., 65 pp 943-950,

---

[3] *https://www.nytimes.com/2017/07/16/us/maryam-mirzakhani-dead.html*