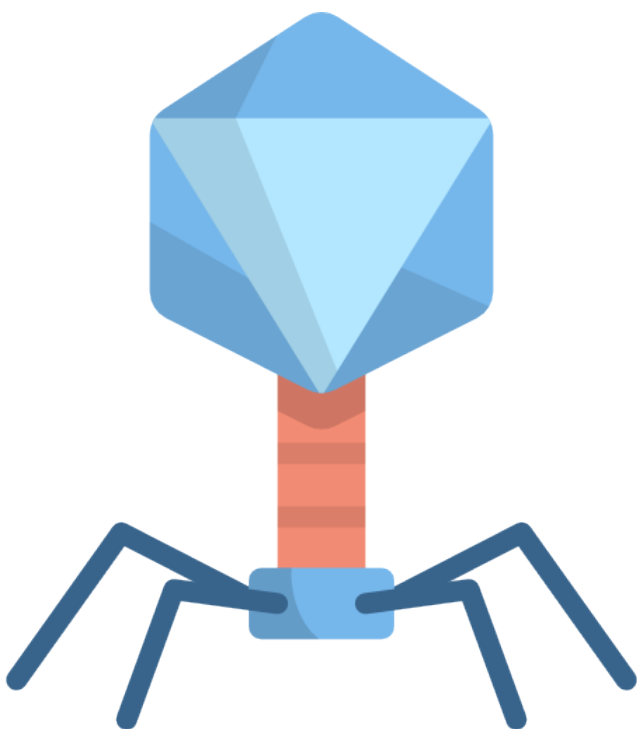


# تسک بیوانفورماتیک

پس از فاجعه ای که در یکی از مراکز تحقیقاتی ایالت جورجیا رخ داد، یک ابرویروس مهندسی شده به جهان بیرون خود آزاد شده و تمام انسان ها را بیمار کرده است! شما به عنوان یک ویروولوژیست جوان آخرین امید بشریت هستید تا با مطالعات همه جانبه این ویروس و خویشاوندانش، راه درمانی پیدا کنید و هموعانتان را نجات دهید. بدین منظور مراحل زیر را طی خواهید کرد.

- بخش اول: توالی یابی ژنوم ویروس
- بخش دوم: بررسی کپسیدهای ویروسی
- بخش سوم: تخمین زمان اشتقاق ویروس بیماری زا
- بخش چهارم: بررسی نوع زندگی ویروس ها
- بخش پنجم: مدل سازی بیماری AIDS



موفق باشید.

## مواد و وسایل مورد نیاز:

فولدر Bioinformatics Task حاوی فایل های:

- sequence\_A.txt تا sequence\_K.txt
- day\_0\_var\_1.txt تا day\_60\_var\_1.txt به تعداد کلی ۶ فایل
- day\_0\_var\_2.txt تا day\_60\_var\_2.txt به تعداد کلی ۶ فایل
- protein\_1.txt تا protein\_3.txt
- appendix.pdf ضمیمه آموزش کار با ابزار های مورد استفاده

صفحه مرورگر حاوی تب های سایت [rcsb.org](https://rcsb.org):

- 3J6R
- 1QGT
- 1M4X
- 5VF3
- 5OAC
- 3J2J
- 1QBE
- 6C21
- 5WK1
- 2ZTN

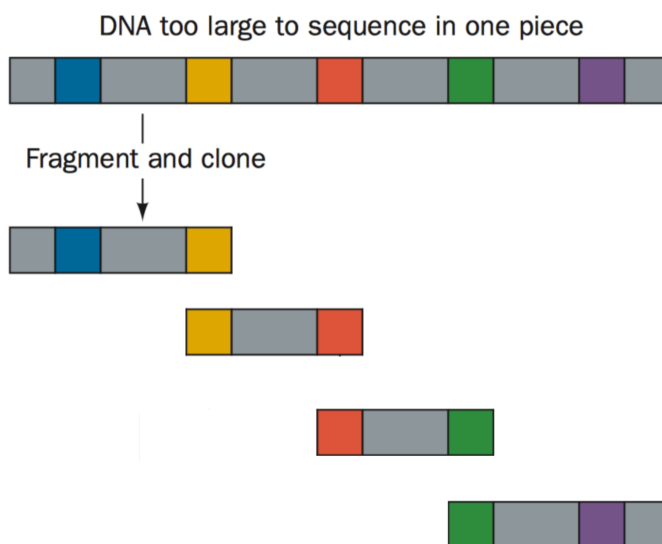
صفحه مرورگر حاوی تب های:

- NCBI BLAST
- EBI pairwise alignment
- EBI multiple alignment

ماشین حساب

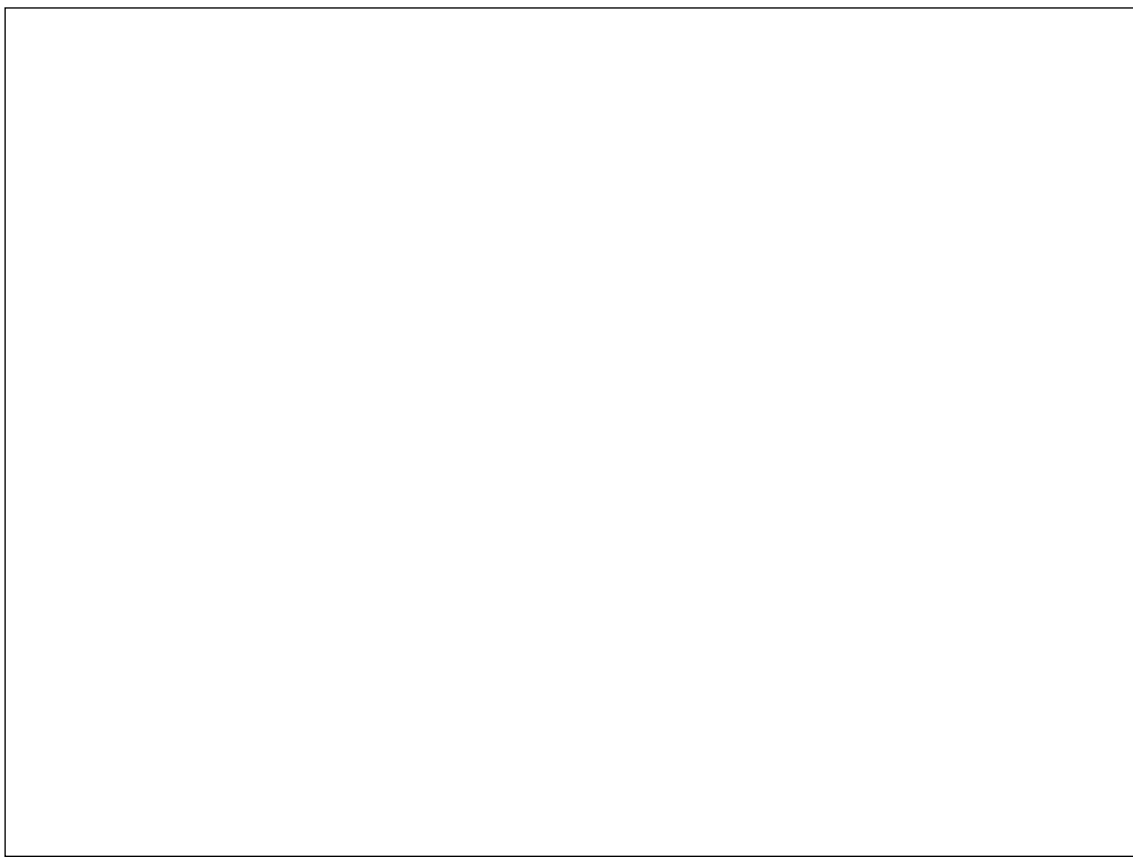
## بخش اول: توالی یابی ژنوم ویروس

با توجه به اینکه شما هنوز در آزمایشگاه خود هستید و تجهیزات بسیار پیشرفته ای در اختیار دارید، تصمیم میگیرید ژنوم ویروس مورد نظر را توالی یابی کنید بلکه اطلاعات جدیدی به دست آورید. دستگاه توالی یابی که در اختیار دارید بسیار دقیق است؛ بدین معنی که هیچوقت در تشخیص یک نوکلئوتید اشتباه نمیکند. اما مشکلی که پیش روی شماست این است که این دستگاه تنها امکان خواندن توالی های کوتاه را دارد. بنابراین نمیتوانید تمام ژنوم ویروس را در یک بار آزمایش توالی یابی کنید. برای حل این مشکل ژنوم ویروس را به چندین قطعه تقسیم میکنید که آنقدر کوچک هستند که دستگاه توانایی تعیین توالی آنها را داشته باشد. پس از به دست آوردن تمام این توالی ها، قصد دارید توالی ژنوم کامل را بازسازی کنید. میدانید که حاصل تکه تکه کردن ژنوم، قطعات پشت سر هم نیست، بلکه قطعات کوچک با هم همپوشانی (overlap) دارند. پس میتوانید با شناسایی این همپوشانی ها، توالی اولیه را بسازید (assemble کنید). مطالب بالا در این تصویر خلاصه شده اند.



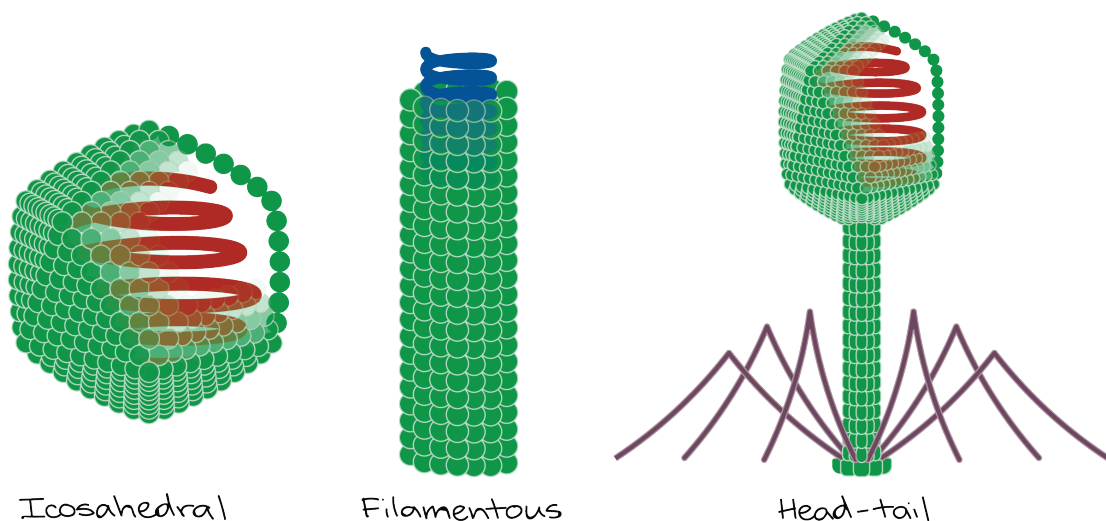
در فولدر پیش روی شما ۱۱ فایل متن sequence\_A.txt تا sequence\_K.txt وجود دارد که توالی ۱۱ قطعه‌ی حاصل از تکه تکه کردن ژنوم ویروس است. با استفاده از ابزار هایی که در اختیار دارید، ژنوم اصلی را assemble کنید و ترتیب درست قطعات کوچک را در جدول پایین بنویسید. بدیهی است که کد گذاری توالی ها ربطی به ترتیب آنها در ژنوم ندارد و تصادفی است.

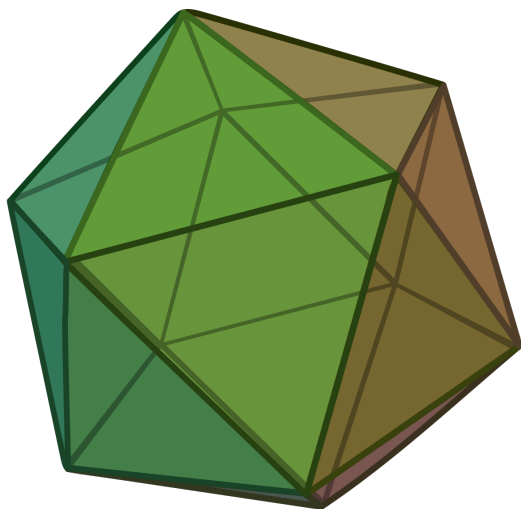
روشی که برای حل این مسئله به کار بردید را توضیح دهید.



### بخش دوم: بررسی کپسیدهای ویروسی

پس از توالی یابی ژنوم به فکران میرسد که شاید بررسی کپسید ویروس اطاعات جدیدی به شما بدهد، بنابراین شروع به تحقیق میکنید. همانطور که میدانید کپسید یک پوشش پروتئینی است که ماده‌ی ژنتیکی ویروس را در بر گرفته است. به صورت کلی سه نوع اصلی کپسید در ویروس ها دیده میشوند که عبارتند از کپسید های مارپیچی helical، بیست وجهی icosahedral و پیچیده complex که میتوان گفت دارای سر icosahedral و دم helical هستند. میدانیم که ویروس مورد نظر ما دارای کپسید icosahedral است.





دانشمندان یونان باستان کشف کرده بودند که تنها اشکالی با ویژگی های هندسی خاص توانایی تشکیل یک محیط سه بعدی بسته را دارند. به عنوان مثال هرم، مکعب و بیست وجهی منتظم. پس از پیشرفت میکروسکوپ ها، مشاهده شد که ویروس هایی که پیشتر به شکل کره دیده میشدند در اصل بیست وجهی منتظم هستند. این نوع آرایش مزایای بسیاری برای ویروس دارد که از جمله آنها میتوان به نیاز به انواع کمی از زیرواحد های سازنده برای ساخت یک کپسید مستحکم و پیچیده اشاره کرد. میتوان تنها با کنار هم گذاشتن پنج ضلعی های منتظم یک بیست وجهی ساخت؛ همانطور که در شکل روبرو و در کپسید ساده ترین ویروس ها مشاهده میکنیم. اما در کپسید های پیچیده تر ترکیبی از پنج ضلعی ها و شش ضلعی های منتظم را

نیز مشاهده میکنیم. پس هر یک از زیر واحد ها میتواند مرکز یک پنج ضلعی یا یک شش ضلعی باشد. یکی از روش هایی که برای بررسی پیچیدگی کپسید ویروس ها وجود دارد، استفاده از پارامتر  $T$  است که بدین صورت تعریف میشود.

$$T = h^2 + hk + k^2$$

که در آن  $h$  و  $k$  مقادیر صحیحی (شامل صفر) هستند که از بررسی کپسید به دست می آیند. برای به دست آوردن این مقادیر باید از مرکز یک پنج ضلعی (دسته ی پنج زیرواحدی که به طور منظم قرار گرفته است) در سطح کپسید به مرکز نزدیک ترین پنج ضلعی برسیم، به طوری که در راه بر روی مراکز شش ضلعی (دسته ی شش زیرواحدی که به طور منظم قرار گرفته است) سطح کپسید فرود بیاییم. تعداد قدم هایی که در یک راستا داریم برابر با  $h$  و در صورت نیاز به تغییر جهت حرکت، تعداد قدم هایی که در راستای دیگر داریم برابر با  $k$  است.

به عنوان مثال مقدار  $T$  برای یک بیست وجهی ساده (شکل بالا) برابر با ۱ است. زیرا تمامی زیرواحد در دسته های پنج ضلعی مرتب شده اند و بنابراین تنها یک جهش برای رسیدن از مرکز یک پنج ضلعی به مرکز پنج ضلعی بعد نیاز است. پس:

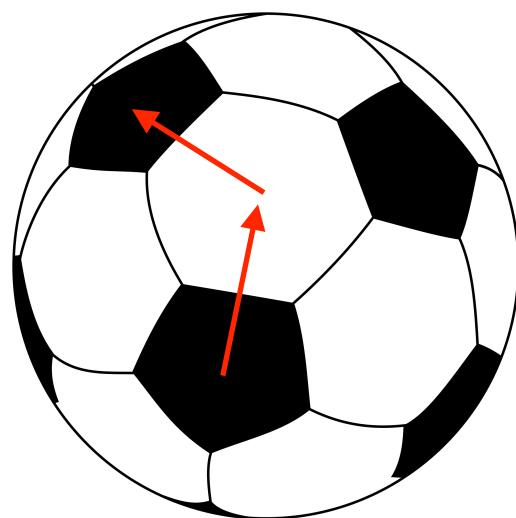
$$h = 1, k = 0$$

$$T = 1^2 + 1 \times 0 + 0^2 = 1$$

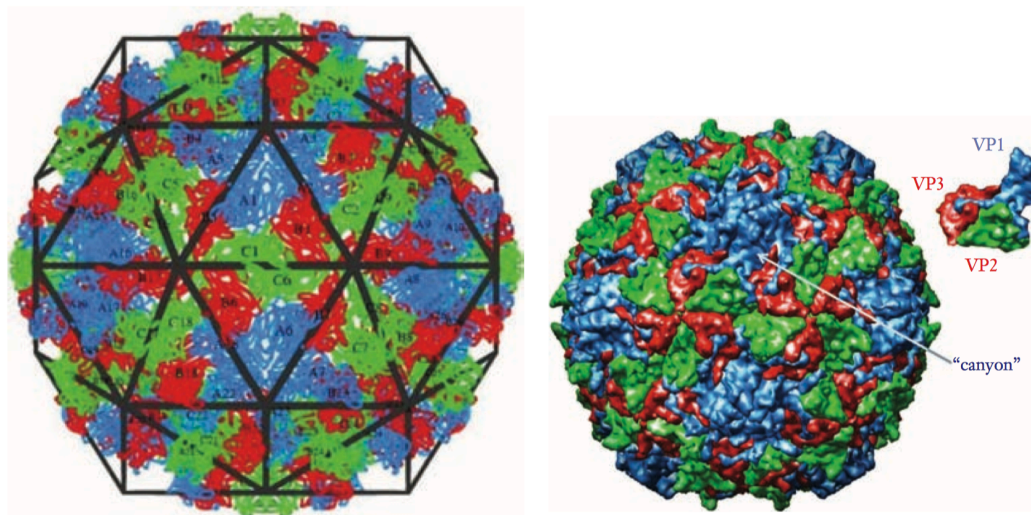
برای مثال بعد میتوان به توپ فوتبال اشاره کرد. در این حجم هندسی، هم دسته های پنج ضلعی مشاهده میشوند و هم شش ضلعی. برای رسیدن از یک پنج ضلعی به مرکز پنج ضلعی دیگر نیاز به دو حرکت داریم که در دو راستای متفاوتند. پس:

$$h = 1, k = 1$$

$$T = 1^2 + 1 \times 1 + 1^2 = 3$$



به عنوان مثال آخر کپسید یک نوع ویروس را بررسی خواهیم کرد (شکل سمت راست). همانطور که مشاهده میکنید این کپسید از تکرار یک زیرواحد ۳ تایی تشکیل شده است. اگر بر روی این کپسید الگوی های پنج ضلعی و شش ضلعی را مشخص کنیم به شکل سمت چپ خواهیم رسید. بنابراین:



$$h = 1, k = 1$$

$$T = 1^2 + 1 \times 1 + 1^2 = 3$$

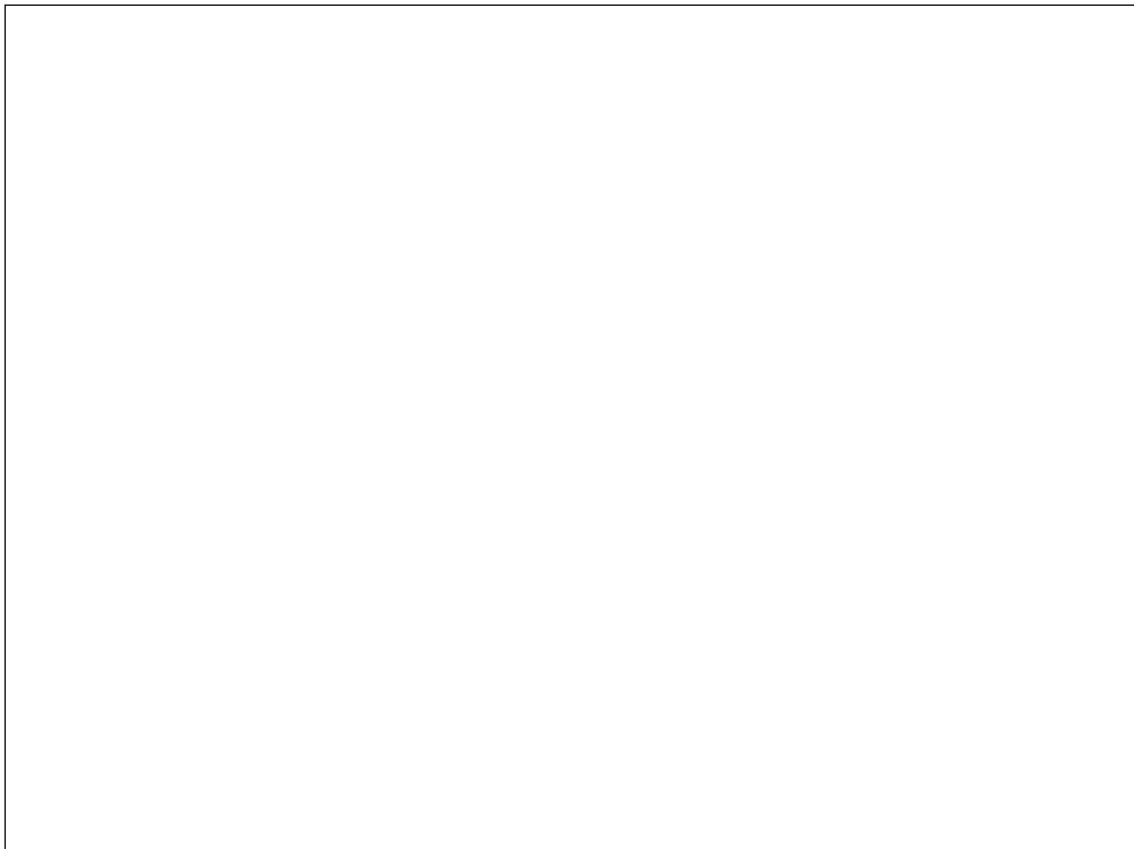
حال پارامتر T را برای چندین ویروس مختلف محاسبه خواهید کرد. بر روی کامپیوتر شما ۱۰ تب از سایت PDB باز است که ساختار سه بعدی کپسید ویروس های مختلف را نشان میدهد. با بررسی آن مقدار T هر کپسید را جلوی نام ویروس وارد کنید.

	human papilloma virus type 16
	human hepatitis B
	PBCV-1
	bacteriophage T4
	FLiP
	coxsackievirus A9
	BACTERIOPHAGE Q BETA
	Staphylococcus aureus phage 80alpha
	marine siphovirus TW1
	Hepatitis E virus ORF2

بخش سوم: تخمین زمان اشتقاق ویروس بیماری زا

با بررسی نمونه هایی که از این ویروس دارید، متوجه میشوید که در نمونه خون بیماری دو "واريته" از این نوع ویروس وجود دارد که از لحاظ ژنتیکی با هم تفاوت هایی دارند. از آنجا که نمونه های این بیمار را در زمان های مختلف دارید به این فکر میفتید که تغییرات ژنتیکی این ویروس را نسبت به زمان بررسی کنید. در فولدر شما ۱۲ فایل متن وجود دارد که حاوی توالی یک قطعه پروتئینی از این ویروس برای هر دو "واريته" و در زمان های متفاوت نمونه گیری است. فایل day\_0\_var\_1.txt به معنای توالی "واريته" اول در روز صفر است و فایل day\_0\_var\_2.txt به معنای توالی "واريته" دوم در روز صفر. با مقایسه این دو "واريته" در طی زمان به سوالات زیر پاسخ دهید.

۱- آیا میزان تغییرات similarity بین این دو توالی از نظریه ساعت مولکولی پیروی میکند؟ اگر بله معادله خط را بنویسید و نمودار آن را در پایین بکشید (با توجه به تعداد نقاطی که دارید، ضریب همبستگی بالای ۰.۹۵ از لحاظ آماری معنی دار است) و تخمین بزنید زمان ظهور این ویروس کی بوده است.



روشی که برای حل این مسئله به کار بردید را توضیح دهید.



بخش چهارم: بررسی نوع زندگی ویروس ها

برای آشنایی بیشتر با چرخه زندگی این ویروس، چند گونه ویروس نزدیک را بررسی خواهید کرد. در پایگاه داده خود توالی های مربوط به یک پروتئین کلیدی در چرخه زندگی برای سه گونه ویروس را پیدا کرده اید. عملکرد این پروتئین ها را در جدول پایین بنویسید و مشخص کنید هر یک از چرخه های زندگی که در پایین نشان داده شده است میتواند متعلق به کدام یک از این ویروس ها باشد. در فولدری که در اختیار دارید سه فایل متن protein\_1.txt تا protein\_3.txt مربوط به این قسمت هستند.

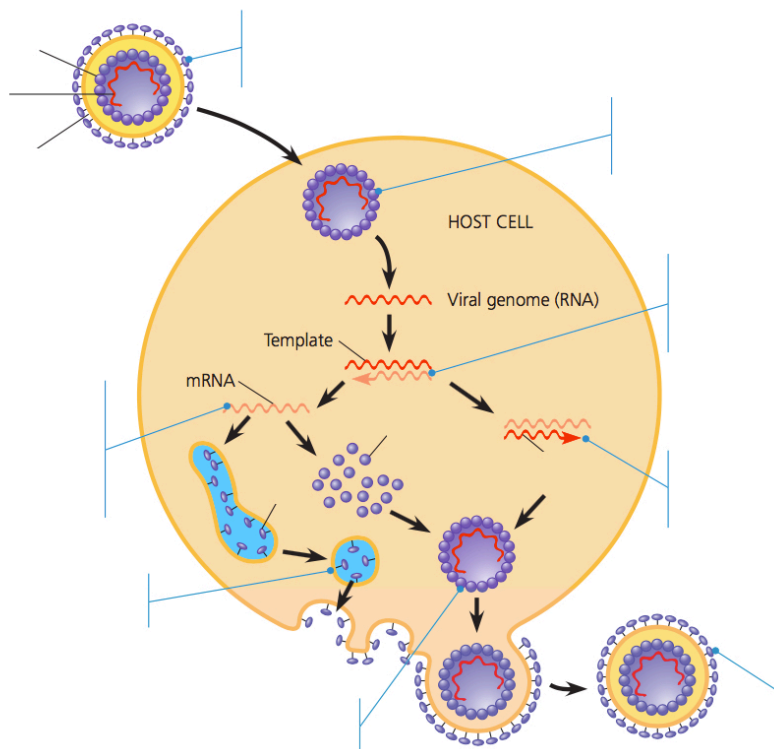
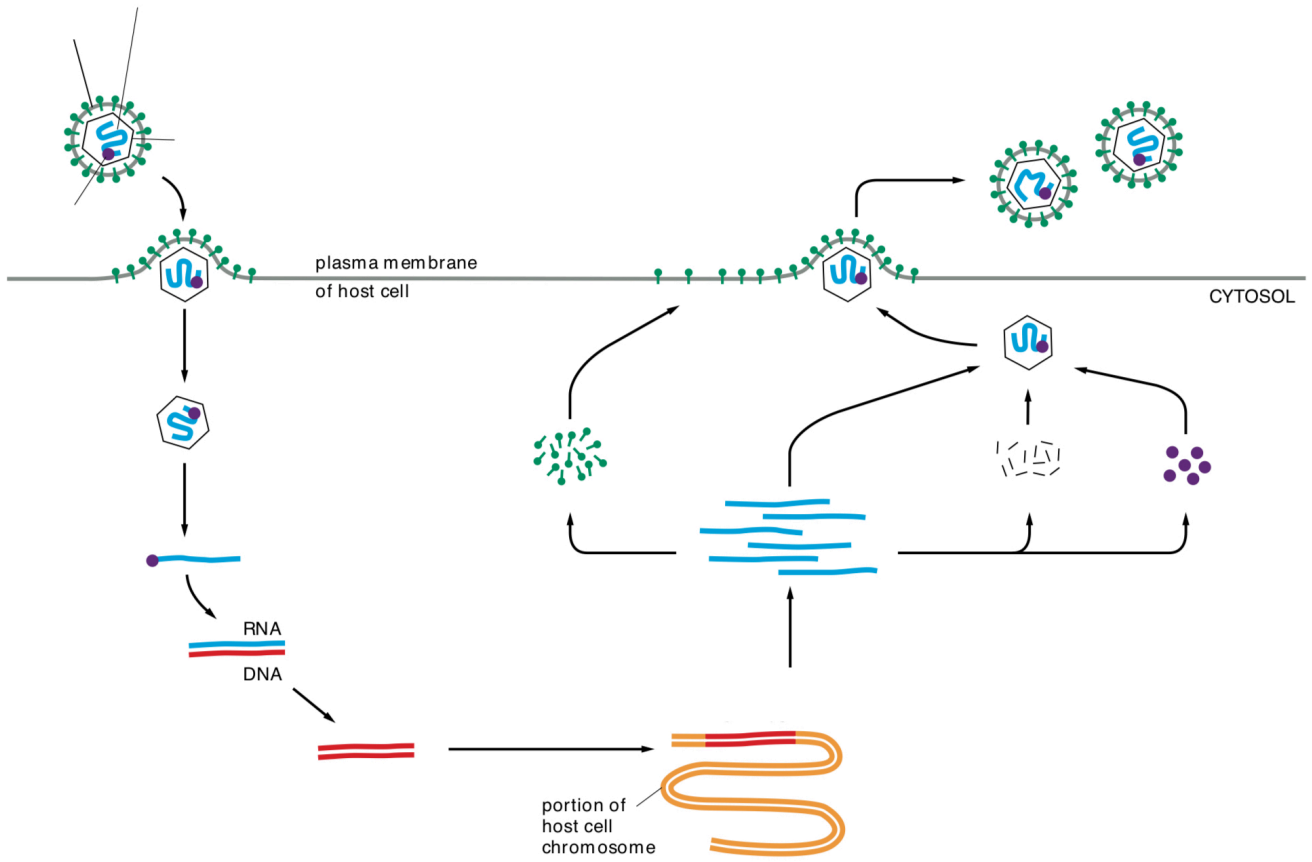
عملکرد	پروتئین
	1
	2
	3

دو چرخه زندگی که در صفحه بعد مشاهده میکنید میتوانند مربوط به کدام یک از این ویروس ها باشند؟ (دقت کنید که عامل اصلی در تصمیم گیری پروتئینی که در اختیار دارید است)

ویروس حاوی پروتئین ۱، ۲ یا ۳	چرخه
	1
	2

کدام یک از این سه ویروس میتواند باعث بیماری سرطان در انسان شود؟







به پای ایستگاه بروید و بگویید آیا هر یک از موارد زیر در مورد دینامیک عفونت در بیمار های 1، 2 و 3 درست یا نادرست است. همچنین مقادیر اولیه X، Y و Z را برای هر بیمار مشخص کنید.

الف) در نهایت سلول های ایمنی به طور کامل از بین می روند

ب) افزایش اولیه ای در پاسخ به عفونت در تعداد لنفوسیت ها دیده می شود.

ج) در ابتدا کاهش می یابند CD8+ tcells

د) هردوی لنفوسیت ها با هم به صورت پایدار افزایش می یابند

بیمار اول:

الف					ب		ج		د		
صحيح											
غلط											

X				Y				Z			

بیمار دوم:

الف					ب	ج	د
صحیح							
غلط							
X			Y			Z	

بیمار سوم:

الف					ب	ج	د
صحیح							
غلط							
X			Y			Z	