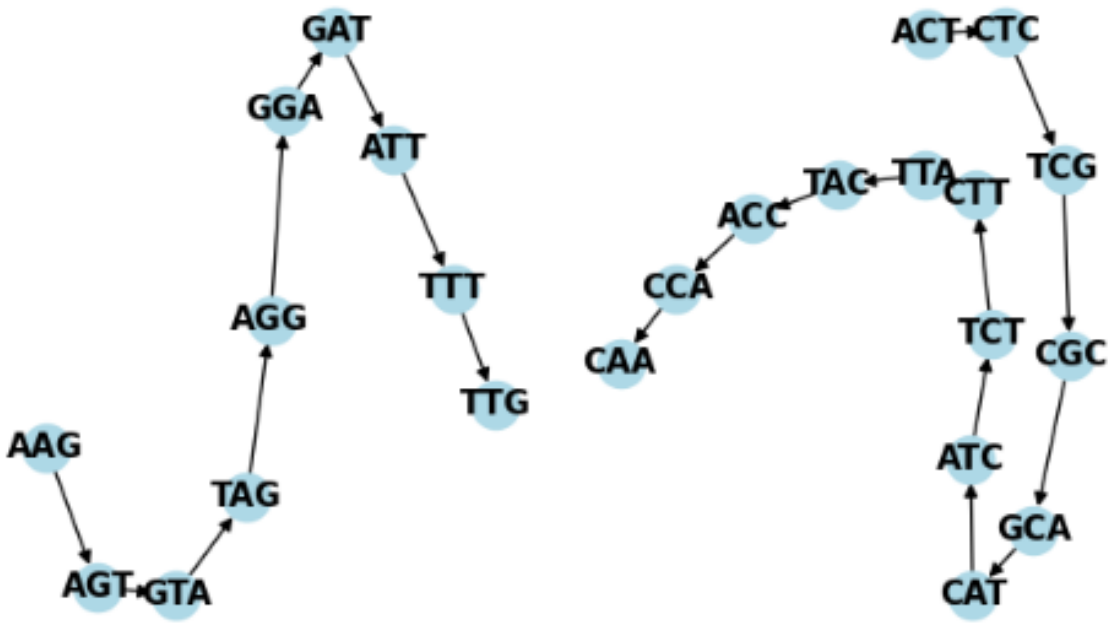
پاسخنامه تمرین دوم

سوال ۲:

با توجه به read های داده شده، ابتدا گراف de bruijn را می‌سازیم که به صورت زیر به نمایش می‌آید:



با توجه به گراف داده شده، ژن‌ها به صورت زیر هستند:

AAGTAGGATTTG
ACTCGCATCTTACCAA

سوال ۳:

S1 = ACCCTGAACC
S2 = ACTCGGAGC
S3 = CTGGAATCT
S4 = GCTAGGACC

برای align کردن هر دو رشته از رابطه زیر استفاده میکنیم.

dp[i][0] = dp[i−1][0] + score(seq1[i-1], ’-’)
dp[0][j] = dp[0][j−1] + score(’-’, seq2[j-1])

Match/Mismatch: dp[i−1][j−1] + score(seq1[i-1], seq2[j-1])
Gap in Sequence 1: dp[i−1][j] + score(seq1[i-1], '-')
Gap in Sequence 2: dp[i][j−1] + score('-', seq2[j-1])
dp[i][j] = max(match,gap_1,gap_2)

	S1	S2	S3	S4
S1	26	17	10	15
S2	17	24	14	18
S3	10	14	22	12
S4	15	18	12	24

محاسبه مرکز:

S1 = 17 + 10 + 15 = 42
S2 = 17 + 14 +18 = 49
S3 = 10 + 14 + 12 = 36
S4 = 15 + 18 + 12 = 45

بنابراین رشته S2 را بعنوان مرکز در نظر میگیریم. سپس هر رشته را با رشته S2 تراز میکنیم.

S1: ACCCTGAACC
S2: ACTCGG-AGC

	-	A	C	T	C	G	G	A	G	C
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
A	-1	2	1	0	-1	-2	-3	-4	-5	-6
C	-2	1	5	4	3	2	1	0	-1	-2
C	-3	0	4	4	7	6	5	4	3	2
C	-4	-1	3	3	7	9	8	7	6	6
T	-5	-2	2	5	6	8	10	9	8	7
G	-6	-3	1	4	7	9	11	10	12	11
A	-7	-4	0	3	6	8	10	13	12	12
A	-8	-5	-1	2	5	7	9	12	13	12
C	-9	-6	-2	1	5	7	9	11	14	16
C	-10	-7	-3	0	4	7	9	10	13	17

S3: -CTGGA-ATCT
S2: ACTCGG-AGC-

	-	A	C	T	C	G	G	-	A	G	C
-	0	-1	-2	-3	-4	-5	-6	-6	-7	-8	-9

C	-1	0	2	1	0	-1	-2	-2	-3	-4	-5
T	-2	0	1	4	3	2	1	1	0	-1	-2
G	-3	-1	2	3	6	6	5	5	4	3	2
G	-4	-2	1	3	5	9	9	9	8	7	6
A	-5	-2	0	2	4	8	9	9	11	10	9
A	-6	-3	-1	1	3	7	8	8	11	11	10
T	-7	-4	-2	1	2	6	8	8	10	12	11
C	-8	-5	-1	0	4	5	8	8	9	12	15
T	-9	-6	-2	1	3	5	7	7	9	11	14

S4: GCTAGG-ACC-
S2: ACTCGG-AGC-

	-	A	C	T	C	G	G	-	A	G	C	-
-	0	-1	-2	-3	-4	-5	-6	-6	-7	-8	-9	-9
G	-1	0	1	0	-1	-1	-2	-2	-3	-4	-5	-5
C	-2	-1	3	2	3	2	1	1	0	-1	-1	-1
T	-3	-1	2	5	4	4	3	3	2	1	0	0
A	-4	-1	1	4	5	4	4	4	5	4	3	3
G	-5	-2	1	3	6	8	7	7	6	8	7	7
G	-6	-3	0	2	5	9	11	11	10	9	10	10
A	-7	-4	-1	1	4	8	10	10	13	12	11	11
C	-8	-5	-1	0	4	7	10	10	12	15	15	15
C	-9	-6	-2	-1	3	6	9	9	11	14	18	18

در نهایت رشته‌های تراز شده به صورت زیر هستند:

S1	A	C	C	C	T	G	A	A	C	C	-
S2(center)	A	C	T	C	G	G	-	A	G	C	-
S3	-	C	T	G	G	A	-	A	T	C	T
S4	G	C	T	A	G	G	-	A	C	C	-

سوال ۴:

الف) روش‌های همترازی ساختاری تکنیک‌هایی هستند که برای همتراز کردن ماکرومولکول‌های زیستی (مانند پروتئین‌ها یا اسیدهای نوکلئیک) بر اساس ساختار سه‌بعدی (3D) آن‌ها به جای صرفاً بر اساس توالی آمینواسید

یا نوکلئوتید استفاده می‌شوند. این نوع همترازی به ویژه در هنگام تحلیل پروتئین‌ها مفید است، زیرا می‌تواند اغلب روابط تکاملی و عملکردی را که تنها از داده‌های توالی قابل مشاهده نیستند، آشکار کند.

نکات کلیدی روش‌های همترازی ساختاری

1. حفظ ساختار 3D:
   - در پروتئین‌ها، ساختار 3D (یا تاخوردگی) اغلب بیشتر از توالی واقعی حفظ می‌شود، به ویژه در فاصله‌های تکاملی بزرگ. پروتئین‌هایی با شباهت توالی کم ممکن است هنوز ساختارهای مشابهی داشته باشند اگر از یک جد مشترک تکامل یافته باشند.
   - بنابراین، روش‌های همترازی ساختاری می‌توانند روابطی را که توسط همترازی‌های توالی از دست رفته‌اند، شناسایی کنند و آن‌ها را برای مطالعه پروتئین‌های دور از هم ارزشمند می‌سازند.
2. همترازی بر اساس مختصات فضایی:
   - روش‌های همترازی ساختاری با مقایسه مختصات فضایی اتم‌ها (معمولاً کربن‌های آلفا در پروتئین‌ها) برای همتراز کردن دو یا چند ساختار در فضای 3D کار می‌کنند. آن‌ها تلاش می‌کنند تا یک ساختار را روی دیگری همپوشانی دهند یا "ابرهم‌گذاری" کنند تا بهترین تناسب را از نظر جهت‌گیری فضایی پیدا کنند.
   - برخلاف همترازی توالی که بقایا را به صورت خطی همتراز می‌کند، همترازی ساختاری به جابجایی‌ها و چرخش‌های پیچیده برای دستیابی به یک همترازی فضایی اجازه می‌دهد.
3. بینش عملکردی:
   - همترازی‌های ساختاری در شناسایی موتیف‌های ساختاری محافظت‌شده که برای عملکرد یک پروتئین حیاتی هستند، مانند سایت‌های فعال، جیب‌های اتصال یا سایر مناطق ضروری برای فعالیت، کمک می‌کنند. این موتیف‌ها ممکن است فقط در توالی حفظ نشده باشند.
   - به عنوان مثال، آنزیم‌هایی با سایت‌های فعال مشابه اما توالی‌های متفاوت می‌توانند به دلیل حفظ ساختاری آن سایت‌ها، همچنان عملکردهای مشابهی داشته باشند.

ب) الگوریتم‌ها و ابزارهای رایج همترازی ساختاری

چندین ابزار و الگوریتم برای همترازی ساختاری استفاده می‌شوند، از جمله:

- **DALI (Distance-matrix ALIgnment):** از ماتریس‌های فاصله برای مقایسه ساختارها و شناسایی همسانی ساختاری استفاده می‌کند.
- **CE (Combinatorial Extension):** روشی که همترازی ساختاری را با مقایسه قطعات پیوسته ساختارها و گسترش همترازی‌ها در صورت سازگاری انجام می‌دهد.
- **TM-align:** ابزاری که ساختارهای پروتئین را بر اساس TM-score، که کیفیت همترازی ساختاری را ارزیابی می‌کند، همتراز می‌کند.
- **STAMP (Structural Alignment of Multiple Proteins):** اغلب در همترازی‌های ساختاری چندگانه برای شناسایی شباهت ساختاری بین چندین پروتئین به طور همزمان استفاده می‌شود.

ج)

ویژگی	هم‌ترازی مبتنی بر توالی	هم‌ترازی مبتنی بر ساختار
مبنای مقایسه	توالی آمینواسید یا نوکلئوتید	ساختار سه بعدی
هدف	شناسایی روابط تکاملی و عملکردی	شناسایی شباهت‌های ساختاری
روش	مقایسه مستقیم توالی‌ها	مقایسه مختصات فضایی اتم‌ها
مزایا	داده‌های توالی در دسترس، الگوریتم‌های کارآمد	شناسایی روابط دور، اطلاعات دقیق‌تر درباره عملکرد
محدودیت‌ها	حساس به تغییرات ساختاری بدون تغییر در توالی، ممکن است روابط بین پروتئین‌های با توالی‌های بسیار متفاوت را از دست بدهد	نیاز به ساختارهای سه بعدی با کیفیت بالا، محاسبات پیچیده، ممکن است برای مولکول‌های با ساختارهای بسیار متفاوت چالش‌برانگیز باشد

د)ادغام داده‌های همترازی ساختاری در همترازی چندگانه توالی (MSA) به ایجاد همترازی‌هایی کمک می‌کند که برای مطالعات عملکردی و تکاملی دقیق‌تر هستند. با همتراز کردن هم توالی و هم ساختار، محققان می‌توانند شکاف‌ها را دقیق‌تر قرار دهند و ویژگی‌های ساختاری مهم را حفظ کنند، که می‌تواند برای مطالعه خانواده‌های پروتئینی با واگرایی بالا در سطح توالی اما ساختارهای محافظت‌شده، حیاتی باشد.

مزایای ادغام روش‌های تراز ساختاری در MSA

- **شناسایی روابط دور:** پروتئین‌هایی که توالی‌های بسیار متفاوتی دارند ممکن است ساختارهای سه بعدی مشابهی داشته باشند که نشان‌دهنده یک جد مشترک است. تراز ساختاری می‌تواند این روابط دور را آشکار کند که با استفاده از MSA به تنهایی قابل تشخیص نیستند.
- **دقت بالاتر در تعیین همسانی:** با ترکیب اطلاعات ساختاری و توالی، می‌توان همترازی‌های دقیق‌تری ایجاد کرد. این امر به ویژه برای مناطقی که در آن‌ها تغییرات توالی زیادی رخ داده اما ساختار حفظ شده است، اهمیت دارد.
- **درک بهتر عملکرد:** با مقایسه ساختارهای سه بعدی پروتئین‌ها، می‌توان مناطقی را شناسایی کرد که برای عملکرد پروتئین حیاتی هستند. این اطلاعات می‌تواند به پیش‌بینی عملکرد پروتئین‌های ناشناخته کمک کند.
- **مدل‌سازی ساختاری:** با استفاده از همترازی‌های ساختاری، می‌توان مدل‌های ساختاری دقیق‌تری از پروتئین‌ها ایجاد کرد، که برای طراحی دارو و مهندسی پروتئین بسیار مهم است.

چالش‌ها و محدودیت‌ها

- **در دسترس بودن ساختارهای سه بعدی:** برای استفاده از روش‌های تراز ساختاری، به ساختارهای سه بعدی با کیفیت بالا نیاز است. این ساختارها معمولاً از طریق کریستالوگرافی اشعه ایکس یا میکروسکوپ الکترونی به دست می‌آیند و برای همه پروتئین‌ها در دسترس نیستند.

- **پیچیدگی محاسباتی:** همترازی ساختاری معمولاً محاسباتی گران است، به ویژه برای پروتئین‌های بزرگ یا مجموعه‌های داده‌های بزرگ.
- **عدم قطعیت در همترازی:** همترازی ساختاری می‌تواند به پارامترهای ورودی و الگوریتم‌های استفاده شده حساس باشد، که می‌تواند منجر به عدم قطعیت در نتایج شود.
- **تغییرات ساختاری محلی:** برخی از تغییرات ساختاری ممکن است بسیار محلی باشند و به راحتی توسط روش‌های همترازی ساختاری شناسایی نشوند.
- **تفسیر نتایج:** تفسیر نتایج همترازی ساختاری می‌تواند پیچیده باشد، به ویژه هنگامی که چندین همترازی ممکن وجود داشته باشد.