

Sains Malaysiana 47(8)(2018): 1731–1739
<http://dx.doi.org/10.17576/jsm-2018-4708-12>

Analisis dan Penyaringan Data Lewah Interaksi Kelompok Bes Berikatan Hidrogen dalam Struktur RNA 3-Dimensi (Analysis and Filtering for Redundant Data of Hydrogen-bonded Base Interaction Clusters in RNA 3-Dimensional Structures)

HAZRINA YUSOF HAMDANI & MOHD FIRDAUS-RAIH*

ABSTRAK

Susun atur 3-dimensi (3D) yang sama boleh disalah cerap sebagai berbeza dari sudut penglihatan yang berlainan. Bagi makromolekul biologi, permasalahan ini juga dihadapi oleh algoritma pencarian susun atur 3D. Keputusan hasil larian yang sama akan diperoleh berulang kali (data lewah) kerana hasil tersebut boleh mempunyai susun atur jujukan berbeza. Permasalahan ini tidak ditemui di dalam pencarian jujukan. Dalam kajian ini, dua kaedah untuk menyaring data lewah tersebut telah dibangunkan dan dibandingkan iaitu kod Prüfer (berasaskan teori graf) dan kaedah saringan data lewah (dibangunkan secara khusus untuk kajian). Model hasil carian pula adalah menggunakan COnnection tables Graphs for Nucleic ACids (COGNAC) bagi pencarian interaksi kelompok bes berikatan hidrogen. Perbandingan yang dilakukan menunjukkan bahawa kaedah saringan data lewah mampu untuk mengenal pasti dan menyaring antara 50.5% sehingga 80% data lewah daripada hasil larian asal COGNAC berbanding kod Prüfer yang hanya mengenal pasti dan menyaring 50% data lewah daripada hasil larian asal COGNAC. Oleh itu, kaedah saringan data lewah telah diimplementasi ke dalam COGNAC. Selain itu, kaedah saringan data lewah ini juga boleh diguna pakai untuk algoritma yang tidak bergantung kepada jujukan bagi pencarian motif 3D dalam struktur protein.

Kata kunci: COGNAC; interaksi kelompok bes berikatan hidrogen; motif 3D RNA

ABSTRACT

There is a tendency for 3-dimensional (3D) arrangements to appear differently from different viewing angles. In biological macromolecules, this is a problem that is also encountered by algorithms searching for 3D arrangements. This results in the same output being retrieved repeatedly because they are not considered redundant from different points of the search. This is a problem that is not encountered for sequence level searches. In this study, we developed two approaches to filter such redundant data that are the Prüfer code (based on graph theory) and our own redundant data filtering method. The search results model uses the COnnection tables Graphs for Nucleic ACids (COGNAC) algorithm to search for clusters of bases that are connected by hydrogen bonds. The comparison between the two methods showed that the redundant data filtering method developed in this work was able to identify and filter between 50.5% and 80% of the redundant data from the original COGNAC results compared to the Prüfer code that identified and filtered 50% of the redundant data from the original COGNAC results. This redundant data filtering method was then integrated into the COGNAC computer program. Furthermore, the redundant data filtering method can also be deployed to the results of other sequence independent 3D search algorithms including those for protein 3D structures.

Keywords: COGNAC; hydrogen-bonded base interactions clusters; 3D RNA motive

PENGENALAN

Susun atur yang sama bagi residu bes dalam asid nukleik atau residu asid amino dalam protein boleh dilihat secara 3-dimensi (3D) apabila cerapan dilakukan dari sudut yang berbeza. Begitu juga pencarian secara 3D yang tidak mengambil kira jujukan dengan susun atur yang sama boleh dijumpai secara berulang kerana dianggap sebagai susun atur yang berbeza. Namun demikian, pencarian yang tidak bergantung kepada jujukan lebih berkesan bagi pencarian 3D disebabkan susun atur yang mungkin melibatkan residu yang berjauhan pada peringkat jujukan (1D) tetapi berdekatan secara 3D.

Pencarian jujukan adalah bergantung kepada kaedah penjajaran dan ini menghadkan pencarian kepada jujukan yang tidak terlalu berjauhan bagi mendapatkan kepastian jajaran. Hasil yang berulang dalam pencarian 3D merupakan satu masalah yang menghasilkan banyak data lewah. Sebagai contoh, tiga residu asid amino - His 57, Asp 102 dan Ser 195 - yang merupakan tapak aktif enzim protease serina hanya akan dilaporkan sebagai His 57, Asp 102 dan Ser 195 daripada hasil penjajaran jujukan. Sebaliknya, pencarian 3D yang tidak mengambil kira jujukan boleh menjumpai susun atur 3D yang sama tetapi dalam kombinasi yang berbeza seperti (Asp 102, His 57, Ser 195) atau (His 57, Ser 195, Asp 102).

Pencarian corak susun atur 3D yang berulang di dalam struktur 3D asid ribonukleik (RNA) atau dikenali sebagai motif struktur 3D RNA adalah penting bagi pengenalpastian fungsi sesuatu molekul RNA. Perisian pencarian sedia ada iaitu *Nucleic Acid Search for Substructures and Motifs* (NASSAM), *WebFR3D* dan *The RNA 3D Motif Atlas* menggunakan takrifan yang memerlukan susun atur yang sesatah antara bes dan sekurang-kurangnya dua ikatan hidrogen yang menghubungkan antara bes (Hamdani et al. 2012; Parlea et al. 2016; Petrov et al. 2011). Takrifan ini telah dipelopori oleh Leontis dan Westhof (2001) dan Burkard et al. (1999).

Takrifan sedemikian adalah berbeza daripada yang digunakan oleh perisian *CONNECTION TABLES GRAPHS FOR NUCLEIC ACIDS* (COGNAC) yang menggunakan takrifan pencarian secara lebih longgar iaitu susun atur antara bes tidak semestinya mengikut satah yang terlalu mendatar dan hanya memerlukan satu ikatan hidrogen antara bes untuk dikira sebagai interaksi bes. Hal ini dapat memberi kelebihan kepada pencarian susun atur motif yang baharu dan pencarian interaksi kelompok bes berikatan hidrogen (Firdaus-Raih et al. 2014; Jeffrey & Saenger 2012). Interaksi kelompok bes berikatan hidrogen ditakrifkan sebagai interaksi bes-bes yang menghubungkan lebih daripada tiga bes.

COGNAC terdiri daripada tiga aturcara komputer iaitu: aturcara *Hydrogen Bonding Predictor* (HBPRED) bagi menjanakan ikatan hidrogen; aturcara penukaran hasil HBPRED kepada jadual hubungan; dan algoritma teras COGNAC iaitu pencarian corak interaksi dalam jadual hubungan yang sepadan dengan kunci carian. Kunci carian COGNAC ialah hubungan corak antara bes-bes atau dikenali sebagai susunan hubungan pokok.

Perisian ini menggunakan perwakilan graf dengan bes diwakili oleh titik (nod) dan sisi yang menghubungkan antara titik merupakan interaksi ikatan hidrogen yang menghubungkan antara bes dengan bes dan menggunakan algoritma isomorfisme subgraf Ullman bagi pencarian corak interaksi dalam jadual hubungan yang sepadan dengan kunci carian.

COGNAC menggunakan kaedah carian terbuka iaitu parameter bagi jenis bes dan bilangan nombor bagi hubungan ikatan hidrogen tidak perlu ditetapkan terlebih dahulu. Hanya satu parameter yang perlu ditetapkan iaitu jenis susunan hubungan pokok bagi bilangan interaksi bes yang ingin dicari.

Keputusan daripada kaedah carian terbuka ini menghasilkan hasil larian COGNAC yang mempunyai data lewah. COGNAC akan mencerpap setiap bes terhubung yang menepati ciri kunci carian yang telah dipilih tanpa mengambil kira susunan bes-bes tersebut. Maka, data lewah ditakrifkan sebagai anotasi yang mempunyai jenis dan nombor bes yang sama tetapi susunan kedudukan bes yang berbeza (secara jujukan). Apabila dicerap dalam struktur 3D RNA, anotasi yang mempunyai susunan bes yang berbeza ini (secara jujukan), terletak pada lokasi yang sama. Oleh itu, satu pembangunan kaedah penyaringan

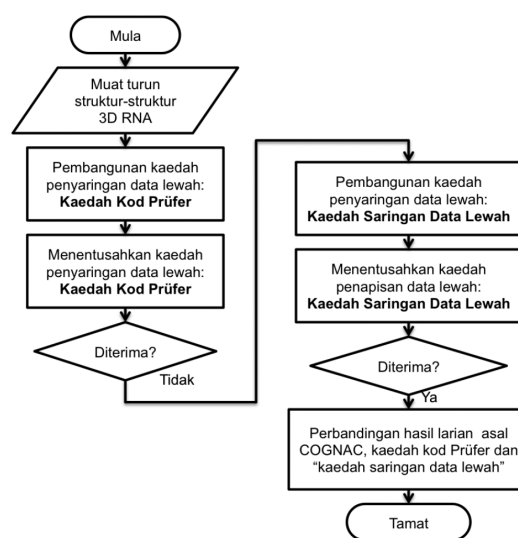
hasil larian COGNAC diperlukan bagi menyaring data lewah ini.

Keputusan kajian literasi mendapati bahawa secara hipotesisnya algoritma kod Prüfer mampu untuk menyaring data lewah COGNAC. Kod Prüfer ialah satu jujukan yang unik bagi sesebuah pokok yang diwakili. Dalam kajian ini, satu pokok mewakili satu anotasi dalam struktur 3D RNA. Perbandingan antara kod Prüfer yang terhasil bagi setiap anotasi dalam struktur 3D RNA dijalankan bagi menyaring data lewah. Oleh itu, pembangunan kod Prüfer sebagai algoritma penyaringan data lewah COGNAC diperlukan. Namun begitu, hasil larian kod Prüfer masih perlu disemak bagi memastikan tiada data lewah yang dihasilkan agar anotasi yang terhasil mempunyai jenis dan nombor bes yang berbeza serta lokasi anotasi di dalam struktur 3D juga berbeza.

KAEDAH KAJIAN

MUAT TURUN STRUKTUR 3D RNA

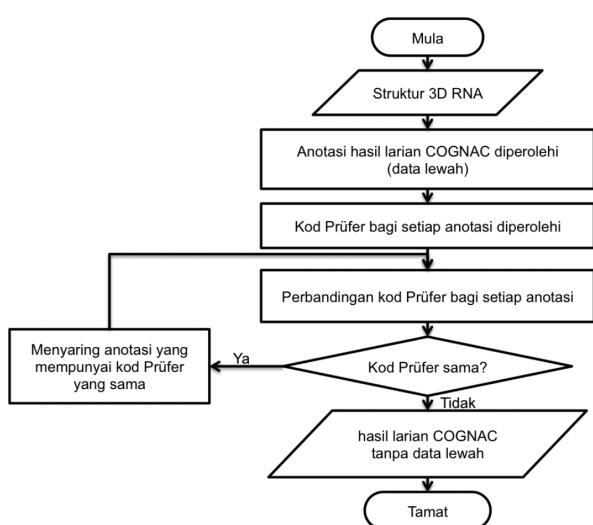
Aliran kerja penyelidikan bermula dengan muat turun struktur 3D RNA dan digambarkan secara keseluruhan dalam Rajah 1. Struktur 3D kompleks RNA telah dimuat turun dari pangkalan data *Protein Data Bank* (PDB) dalam format PDB (Berman et al. 2000; Rose et al. 2017). Struktur 3D kompleks RNA ialah struktur 3D yang mengandungi rantaian RNA dan boleh juga mengandungi rantaian molekul lain seperti protein, ligan, ion dan asid deoksiribonukleik (DNA). Struktur 3D kompleks RNA ini merupakan hasil pemprosesan data pembelauan sinar-X ke atas hablur yang mengandungi kompleks RNA yang beresolusi sinar-X antara 0 Å sehingga 4 Å. Bilangan struktur yang dimuat turun adalah sebanyak 2158 seperti dilaporkan bagi pangkalan data InterRNA (Appasamy et al. 2015).



RAJAH 1. Aliran aktiviti penyelidikan

PEMBANGUNAN KAEDAH PENYARINGAN DATA LEWAH:
KAEDAH KOD PRÜFER

Kaedah kod Prüfer dibangunkan untuk menyaring data lewah dalam hasil larian COGNAC. Bahasa pengaturcaraan yang digunakan untuk membangunkan kaedah kod ialah bahasa pengaturcaraan Perl. Kaedah penyaringan data lewah menggunakan kaedah kod Prüfer bermula dengan mendapatkan hasil larian COGNAC yang mempunyai data lewah. Seterusnya, kod Prüfer bagi setiap anotasi di dalam struktur 3D RNA diperoleh dengan menggunakan algoritma jujukan Prüfer. Setelah itu, perbandingan kod Prüfer yang mewakili setiap anotasi dijalankan. Jika terdapat kod Prüfer yang sama, maka hanya satu sahaja anotasi yang diwakili oleh kod Prüfer akan dipilih dan dimasukkan ke dalam hasil larian COGNAC (Rajah 2).

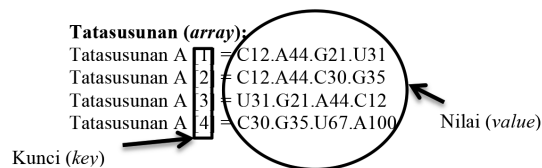


RAJAH 2. Carta alir algoritma kaedah penapisan data lewah menggunakan kod Prüfer

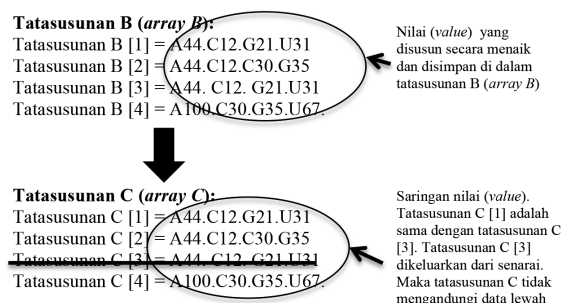
PEMBANGUNAN KAEDAH PENYARINGAN DATA LEWAH:
KAEDAH SARINGAN DATA LEWAH

Pembangunan kaedah penyaringan data lewah yang baru dinamakan sebagai ‘kaedah saringan data lewah’ kerana kaedah penyaringan data lewah menggunakan kod Prüfer masih menghasilkan data lewah. Kaedah saringan data lewah bermula dengan mendapatkan hasil larian COGNAC. Seterusnya, hasil larian COGNAC yang mempunyai data lewah disimpan di dalam tatasusunan A. Tatasusunan mempunyai kunci atau nombor indeks bagi menentukan bilangan yang keberapa bagi sesebuah anotasi hasil larian COGNAC yang disimpan di dalam tatasusunan. Anotasi hasil larian COGNAC pula ialah nilai yang disimpan di dalam tatasusunan (Rajah 3).

Setiap bes di dalam anotasi struktur 3D RNA disusun secara menaik dan seterusnya disimpan di dalam tatasusunan B (Rajah 4). Saringan nilai tatasusunan B menggunakan fungsi yang dibangunkan bagi mencari nilai yang unik di dalam tatasusunan B. Nilai yang unik kemudiannya disimpan di dalam tatasusunan C (Rajah 4).



RAJAH 3. Istilah kunci (*key*) dan nilai (*value*) bagi tatasusunan (*array*)



RAJAH 4. Contoh nilai (*value*) yang disusun secara menaik dan saringan nilai (*value*) yang mempunyai susun atur bes (jujukan bes) yang sama

Seterusnya kunci tatasusunan C diperoleh. Hal ini bertujuan untuk menggunakan kunci tersebut bagi memanggil dan mencetak nilai di dalam tatasusunan A. Nilai tatasusunan A digunakan bagi mencetak susunan bes yang asal (Rajah 5).

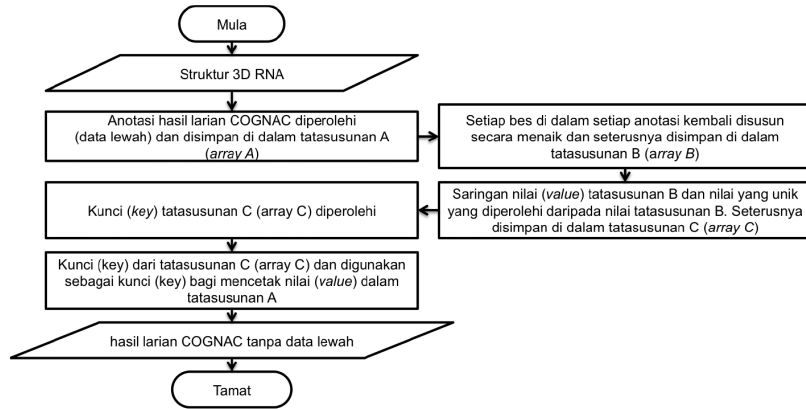
MENENTUSAHKAN KAEDAH PENYARINGAN DATA LEWAH

Setelah saringan data lewah dijalankan, lokasi hasil larian (kaedah kod Prüfer dan kaedah saringan data lewah) seterusnya disemak menggunakan kaedah visualisasi. Kaedah visualisasi ini digunakan bagi memastikan data lewah telah disaring sepenuhnya dengan membandingkan lokasi bagi setiap anotasi dalam hasil larian. Perisian Chimera telah digunakan bagi menyemak lokasi hasil larian ini (Pettersen et al. 2004).

KEPUTUSAN DAN PERBINCANGAN

KAEDAH KOD PRÜFER

Penamaan susun atur bes berinteraksi melalui rangkaian ikatan hidrogen ‘interaksi lima bes Jenis 1’ - telah dilaporkan oleh kumpulan pembangunan perisian COGNAC (Firdaus-Raih et al. 2014). Contohnya anotasi bagi interaksi lima bes Jenis 1 dapat dikesan pada struktur ribozim Kumpulan 1 (ID PDB: 1GID) (Cate et al. 1996) dan terdapat empat anotasi bagi interaksi lima bes Jenis 1 ini. Anotasi yang pertama ialah A121.A200.A197.A124.A201; anotasi yang kedua ialah A201.A124.A197.A200.A121; anotasi yang ketiga ialah A200.A197.A124.A201.A120; dan anotasi keempat ialah A120.A201.A124.A197.A200.



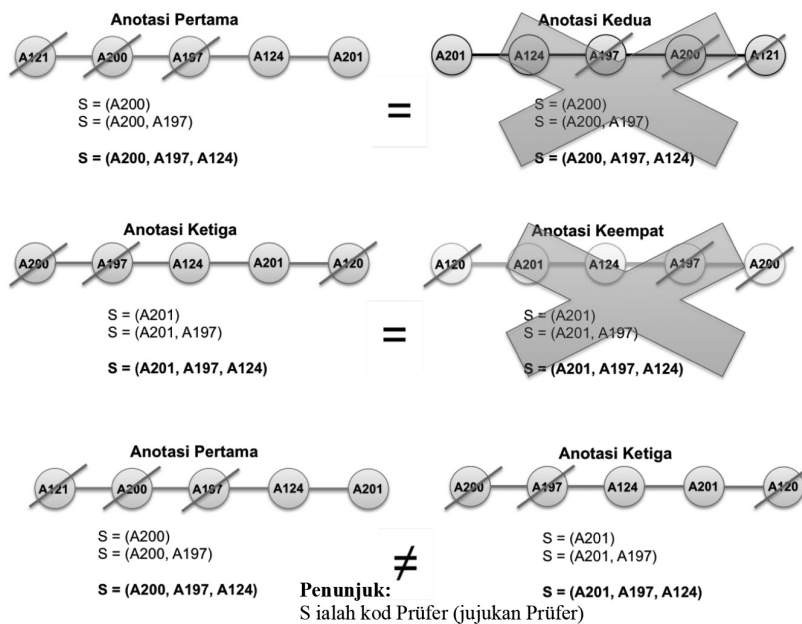
RAJAH 5. Carta alir kaedah penapisan data lewah menggunakan kaedah tapisan data lewah

Kod Prüfer bagi setiap anotasi ini diperolehi dan seterusnya perbandingan kod Prüfer dijalankan (Rajah 6).

Anotasi pertama mempunyai kod Prüfer yang sama dengan anotasi yang kedua iaitu $S = (A200, A197, A124)$. Ini menunjukkan anotasi pertama dan kedua merupakan data lewah. Oleh itu, hanya anotasi pertama telah dipilih bagi anotasi lewah ini. Seterusnya, perbandingan kod Prüfer antara anotasi ketiga dan anotasi keempat. Anotasi ketiga mempunyai kod Prüfer yang sama dengan anotasi yang keempat iaitu $S = (A201, A197, A124)$. Oleh itu, hanya anotasi ketiga telah dipilih bagi anotasi lewah ini. Seterusnya, perbandingan kod Prüfer antara anotasi pertama dengan anotasi ketiga telah dijalankan. Kod Prüfer antara anotasi pertama dan anotasi ketiga berbeza iaitu $S = (A200, A197, A124)$ dan $S = (A201, A197, A124)$. Oleh itu, anotasi pertama dan ketiga telah dipilih ke dalam hasil larian COGNAC yang baharu.

Walaupun kod Prüfer berbeza bagi setiap anotasi, terdapat juga situasi dengan lokasi anotasi adalah pada lokasi yang sama dalam struktur 3D RNA. Ini dapat dikesan setelah kaedah visualisasi digunakan bagi menentusahkan kaedah penyaringan data lewah. Kaedah visualisasi ini menggunakan perisian Chimera. Sebagai contoh, empat anotasi daripada interaksi empat bes Jenis 1 yang ditemui pada struktur 3D subunit besar ribosom *Haloarcula marismortui* (ID PDB: 1K8A) mempunyai kod Prüfer yang berbeza (Jadual 1 dan Rajah 7) (Hansen et al. 2002). Huruf A pada permulaan nombor bes merupakan rantaian yang mewakili subunit besar ribosom *H. marismortui* (ID PDB: 1K8A).

Analisis menggunakan kaedah visualisasi mendapati anotasi pertama iaitu A153.A184.A150.A185 mempunyai kod Prüfer $S = (A184, A150)$, anotasi kedua iaitu A153.A185.A150.A184 mempunyai kod Prüfer $S = (A185,$



RAJAH 6. Contoh kaedah penyaringan data lewah menggunakan kod Prüfer bagi interaksi lima bes Jenis 1 pada struktur ribozim Kumpulan 1 (ID PDB: 1GID)

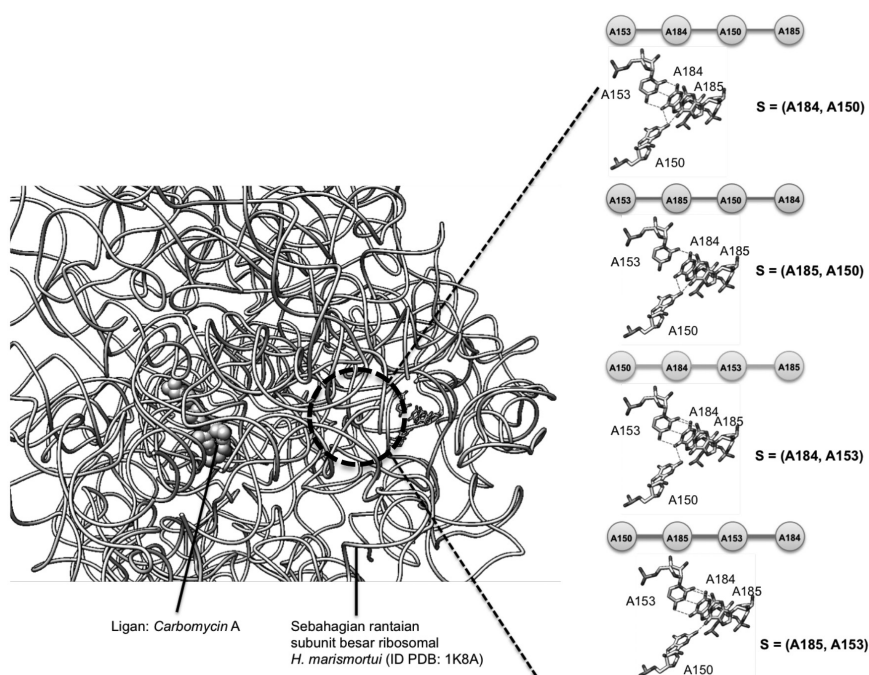
A150), anotasi ketiga iaitu A150.A184.A153.A185 mempunyai kod Prüfer $S = (A184, A153)$ dan anotasi keempat iaitu A150.A185.A153.A184 mempunyai kod Prüfer $S = (A185, A153)$ mempunyai susun atur 3D yang sama. Perbezaan hanya pada susunan kedudukan bes pada jujukan anotasi dan ikatan hidrogen yang terlibat bagi menghubungkan antara bes bagi setiap anotasi (Rajah 7). Terdapat lima ikatan hidrogen terhasil antara bes bagi anotasi pertama (A153.A184.A150.A185). Untuk pasangan bes A153.A184 terdapat tiga ikatan hidrogen yang terhasil iaitu ikatan hidrogen pertama: atom penyumbang N4 bes C153 kepada atom penerima O6 bes G0184, ikatan hidrogen kedua: atom penyumbang N2 bes G184 kepada atom penerima O2 bes C153 dan ikatan hidrogen ketiga: atom penyumbang N1 bes G184 kepada atom penerima N3 bes C153. Untuk pasangan bes A184.A150 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G184. Untuk pasangan bes yang terakhir A150.A185, terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G185.

Terdapat tiga ikatan hidrogen terhasil antara bes bagi anotasi kedua (A153.A185.A150.A184). Untuk pasangan bes A153.A185 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G185 kepada atom O2 bes C153. Untuk pasangan bes A185.A150 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G185. Untuk pasangan bes yang terakhir A150.A184 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G184.

Terdapat lima ikatan hidrogen terhasil antara bes bagi anotasi ketiga (A150.A184.A153.A185). Untuk pasangan bes A150.A184 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G184. Untuk pasangan bes A184.A153 terdapat tiga ikatan hidrogen yang terhasil iaitu ikatan hidrogen pertama: atom penyumbang N4 bes C153 kepada atom penerima O6 bes G0184, ikatan hidrogen kedua: atom penyumbang N2 bes G184 kepada atom penerima O2 bes C153 dan ikatan hidrogen ketiga: atom penyumbang N1 bes G184 kepada atom penerima N3 bes C153. Untuk pasangan bes yang terakhir A153.A185 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G185 kepada atom O2 bes C153.

Terdapat lima ikatan hidrogen terhasil antara bes bagi anotasi keempat (A150.A185.A153.A184). Untuk pasangan bes A150.A185, terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G150 kepada atom penerima O6 bes G185. Untuk pasangan bes A185.A153 terdapat satu ikatan hidrogen yang terhasil iaitu atom penyumbang N2 bes G185 kepada atom O2 bes C153. Untuk pasangan bes yang terakhir A153.A184 terdapat tiga ikatan hidrogen yang terhasil iaitu ikatan hidrogen pertama: atom penyumbang N4 bes C153 kepada atom penerima O6 bes G0184, ikatan hidrogen kedua: atom penyumbang N2 bes G184 kepada atom penerima O2 bes C153 dan ikatan hidrogen ketiga: atom penyumbang N1 bes G184 kepada atom penerima N3 bes C153.

Empat anotasi ini (Jadual 1) terletak pada lokasi yang sama iaitu pada Domain I dalam struktur subunit besar ribosom *H. marismortui* (ID PDB: 1K8A) (Hansen



RAJAH 7. Lokasi empat anotasi yang terhasil adalah sama pada struktur 3D subunit besar *H. marismortui* walaupun selepas kaedah kod Prüfer telah digunakan bagi menyaring data lewah. Garis terputus yang menghubungkan antara bes mewakili ikatan hidrogen

JADUAL 1. Contoh kod Prüfer daripada Interaksi Empat Bes Jenis Pertama (*Quadruple Pattern 1*) struktur subunit besar *H. marismortui* (ID PDB: 1K8A)

Rantai1.No_bes1	Rantai2.No_bes2	Rantai3.No_bes3	Rantai3.No_bes3	Kod Prüfer (S: kod Prüfer)
A153	A184	A150	A185	S = (A184, A150)
A153	A185	A150	A184	S = (A185, A150)
A150	A184	A153	A185	S = (A184, A153)
A150	A185	A153	A184	S = (A185, A153)

et al. 2002; Petrov et al. 2014). Domain I yang terletak bersebelahan Domain IV ialah domain terkecil antara enam domain yang ada dalam subunit besar ribosom *H. marismortui* (ID PDB: 1K8A). Walaupun Domain I tidak mempunyai fungsi yang khusus, tetapi Domain I berkemungkinan meneguhkan binaan struktur 3D bagi Domain IV. Ini kerana Domain IV mempunyai lokasi tapak aktif bagi struktur subunit besar ribosom *H. marismortui* (Hansen et al. 2002; Petrov et al. 2014).

Oleh kerana kaedah kod Prüfer masih menghasilkan data lewah, maka satu kaedah baharu telah dibangunkan dan diberi nama sebagai 'kaedah saringan data lewah'.

KAEDAH SARINGAN DATA LEWAH

Pencarian anotasi unik menggunakan fungsi Perl yang dibangunkan dalam kaedah ini adalah bagi menyaring anotasi yang lewah (Rajah 8). Contohnya seperti anotasi pada interaksi enam bes Jenis 1. Interaksi ini dapat dikesan pada struktur ribosuis kotak T (ID PDB: 4MGN) (Grigg & Ke 2013). Terdapat enam anotasi yang mempunyai data lewah pada interaksi enam bes Jenis 1 ini. Anotasi yang pertama ialah A49.A65.A50.A64.A51.A63; anotasi yang kedua ialah A63.A51.A64.A50.A65.A49; anotasi yang ketiga ialah A29.A75.A32.A30.A73.A33; anotasi keempat ialah A31.A75.A32.30.A73.A33; anotasi kelima ialah A33.A73.A30.32.A75.A29; dan yang terakhir anotasi keenam iaitu A33.A73.A30.A32.A75.A31. Huruf A di hadapan nombor bes merujuk rantaian A yang merupakan rantaian kepada struktur ribosuis kotak T. Anotasi yang unik diperoleh dan seterusnya anotasi ini dicerap di dalam struktur ribosuis kotak T (Rajah 8).

Hasil larian COGNAC setelah menggunakan kaedah saringan data lewah menunjukkan anotasi pertama: A49.A65.A50.A64.A51.A63; anotasi ketiga: A29.A75.A32.A30.A73.A33 dan anotasi keempat: A31.A75.A32.30.A73.A33 sahaja yang masih kekal. Anotasi kedua: A63.A51.A64.A50.A65.A49, anotasi kelima: A33.A73.A30.32.A75.A29; dan anotasi keenam: A33.A73.A30.A32.A75.A31 telah dikeluarkan daripada senarai hasil larian. Ini kerana anotasi kedua merupakan lewah kepada anotasi pertama. Manakala anotasi kelima merupakan lewah kepada anotasi ketiga dan anotasi keenam merupakan lewah kepada anotasi keempat.

Seterusnya hasil larian COGNAC disemak menggunakan kaedah visualisasi bagi menentusahkan hasil saringan data lewah. Bagi interaksi enam bes Jenis 1 ini, anotasi pertama: A49.A65.A50.A64.A51.A63 berkedudukan pada heliks

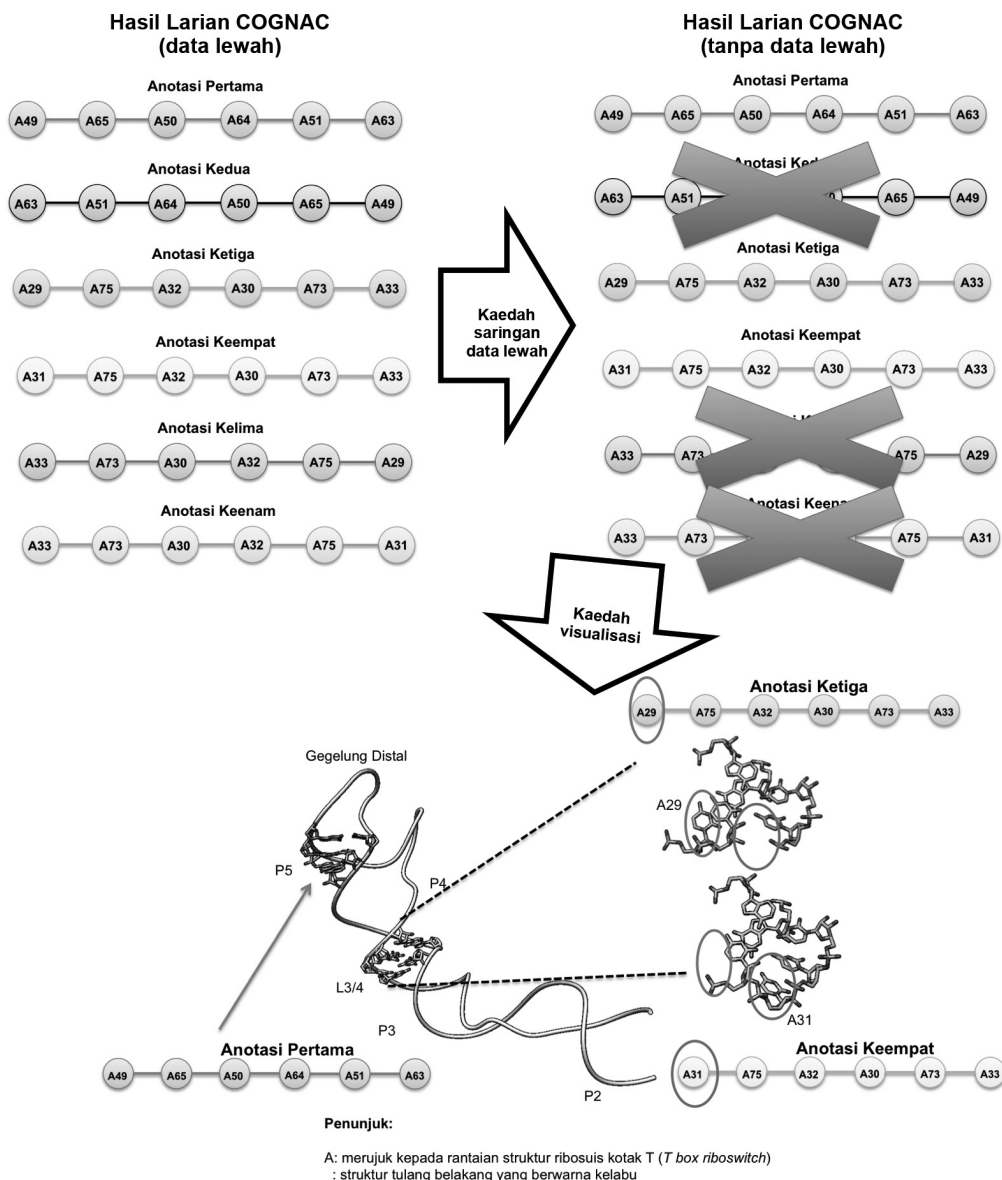
kelima (P5) iaitu kedudukan sebelum gegelung distal (Grigg & Ke 2013). Gegelung distal merupakan lokasi yang penting kerana merupakan lokasi berlakunya interaksi kodon-antikodon (Grigg & Ke 2013). Maka, anotasi pertama: A49.A65.A50.A64.A51.A63 berkemungkinan berfungsi sebagai pelantar yang menyokong struktur gegelung distal bagi membenarkan berlakunya interaksi kodon-antikodon.

Manakala bagi anotasi ketiga: A29.A75.A32.A30.A73.A33 dan anotasi keempat: A31.A75.A32.30.A73.A33, kedua-dua anotasi ini terletak pada lokasi yang sama iaitu pada gegelung 3/4 (L3/4) yang mana L3/4 merupakan motif 2-dimensi (2D) yang terpelihara di dalam struktur ribosuis kotak T (Grigg & Ke 2013). Walau bagaimanapun, kedua-dua anotasi ini tidak dianggap sebagai data lewah. Hal ini kerana, anotasi ketiga dan anotasi keempat hanya berkongsi lima bes yang sama sahaja iaitu A75.A32.A30.A73.A33. Manakala, satu lagi bes berbeza iaitu A29 dan A31. Situasi ini berbeza berbanding hasil larian kod Prüfer, dengan semua bes dalam interaksi empat bes Jenis 1 merupakan bes yang sama (Rajah 7). Oleh yang demikian, kaedah saringan data lewah dapat menghapuskan hasil larian yang mempunyai data lewah.

PERBANDINGAN HASIL LARIAN ASAL COGNAC DAN HASIL LARIAN KAEDAH-KAEDAH PENYARINGAN DATA LEWAH

Perbandingan keputusan antara hasil larian kaedah kod Prüfer, hasil larian kaedah saringan data lewah dan hasil larian asal COGNAC dijalankan bagi melihat keberkesanan kaedah-kaedah penyaringan data lewah yang telah dibangunkan (Jadual 2). Perbandingan ini dibuat dengan mengambil contoh bilangan hasil larian bagi interaksi empat bes Jenis 1 berdasarkan lapan struktur 3D RNA yang mewakili jenis molekul RNA yang berlainan seperti tRNA (ID PDB: 1ASY dan 6TNA), ribozim (ID PDB: 1CX0 dan 1HMH), ribosuis (ID PDB: 3FU2 dan 1Y26) dan ribosom (ID PDB: 4TPE dan 1FFK).

Daripada keputusan Jadual 2, empat struktur 3D iaitu 1ASY, 6TNA, 1CX0 dan 1HMH mempunyai bilangan hasil larian yang sama antara kaedah kod Prüfer dan kaedah saringan data lewah. Hal ini mungkin berlaku kerana struktur tRNA dan ribozim merupakan struktur yang kecil iaitu antara 75 sehingga 144 nukleotida, maka hasil larian asal COGNAC adalah sedikit dan tidak menghasilkan data lewah yang besar (Ferre-D'Amare et al. 1998; Pley et al. 1994; Ruff et al. 1991). Walau bagaimanapun, kaedah kod Prüfer dan kaedah saringan data lewah berjaya mengenal



RAJAH 8. Contoh interaksi enam bes Jenis 1 menggunakan kaedah saringan data lewah dan menentusahkan hasil larian menggunakan kaedah visualisasi bagi struktur ribosuis kotak T

JADUAL 2. Perbandingan bilangan interaksi empat bes Jenis 1 bagi hasil larian asal COGNAC berbanding hasil larian kaedah kod Prüfer dan hasil larian kaedah saringan data lewah

ID PDB	Molekul RNA	Hasil larian asal COGNAC	Hasil larian kaedah kod Prüfer	Hasil larian kaedah saringan data lewah
1ASY	tRNA	4	2	2
6TNA	tRNA	2	1	1
1HMH	Ribozim <i>Hammerhead</i>	8	4	4
1CX0	Ribozim Hepatitis B Virus (HDV)	14	7	7
3FU2	Ribosuis PreQ1 kelas 1	60	30	27
1Y26	Ribosuis Purina	10	5	2
4TPE	Ribosom subunit kecil	256	128	118
1FFK	Ribosom subunit besar	206	103	102

pasti 50% daripada hasil larian asal COGNAC merupakan data lewah. Seterusnya 50% data lewah tersebut telah disaring.

Manakala bagi struktur yang lebih besar iaitu struktur 3FU2, 1Y26, 4TPE dan 1FFK, kaedah kod Prüfer dapat mengenal pasti 50% daripada larian asal COGNAC

merupakan data lewah dan data lewah tersebut telah disaring. Manakala bagi kaedah saringan data lewah, kaedah ini berjaya mengenal pasti antara 50.5% sehingga 80% daripada hasil larian asal COGNAC merupakan data lewah dan data lewah tersebut telah disaring. Hal ini menunjukkan bahawa dengan menggunakan kaedah saringan data lewah dapat melakukan penyaringan data lewah dengan lebih tepat berbanding kaedah kod Prüfer. Oleh itu, kaedah saringan data lewah telah dipilih sebagai kaedah penyaringan data lewah hasil larian COGNAC.

KESIMPULAN

Kehadiran data lewah hasil pencarian 3D yang tidak bersandarkan maklumat jujukan adalah masalah lazim dalam kajian pencarian susun atur 3D. Seringkali, penyelesaian secara khusus perlu dibangunkan yang boleh menangani hasil kaedah tertentu. Dengan pembangunan dua kaedah penyaringan data lewah iaitu kaedah teori graf yang berasaskan kod Prüfer dan kaedah saringan data lewah yang dibangunkan khusus untuk kajian ini, hasil carian bagi sesuatu perisian pencarian motif 3D seperti COGNAC dapat dibersihkan dan dipaparkan kepada pengguna melalui kaedah perantara web. Tanpa penyaringan sedemikian, perisian sebegini hanya dapat dioperasikan secara dalaman disebabkan hasil yang kompleks dan sukar ditakrif. Kaedah penyaringan yang dibangunkan khusus untuk kajian ini mampu untuk mengenal pasti dan menyaring antara 50.5% ke 80% daripada hasil larian asal COGNAC yang mempunyai data lewah. Hal ini kerana kaedah saringan data lewah dibangunkan secara khusus untuk mematuhi kriteria bahawa jika setiap anotasi mempunyai jujukan yang sama tetapi pada susunan yang berbeza, maka anotasi itu perlu berada di lokasi yang berbeza. Algoritma penyaringan ini juga sesuai diguna pakai bagi menyaring hasil pencarian 3D yang tidak bergantung kepada maklumat jujukan dalam struktur protein.

PENGHARGAAN

Terima kasih kepada Kementerian Pengajian Tinggi dan Universiti Sains Malaysia kerana memberi penajaan pengajian Hazrina Yusof Hamdani. Penyelidikan ini telah dibiayai geran MOSTI SciFund 02-01-02-SF1278 dan geran Universiti Kebangsaan Malaysia GUP-2014-008.

RUJUKAN

Appasamy, S.D., Hamdani, H.Y., Ramlan, E.I. & Firdaus-Raih, M. 2015. InterRNA: A database of base interactions in RNA structures. *Nucleic Acids Research* D1: D266-D271.

Burkard, M.E., Turner, D.H. & Tinoco Jr, I. 1999. Appendix 1: Structures of base pairs involving at least two hydrogen bonds. In *The RNA World*. 2nd ed., edited by Gesteland, R.F. & Cech, T. & Atkins, J.F. New York: Cold Spring Harbor Laboratory Press. pp. 675-680.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. & Doudna, J.A. 1996. Crystal

structure of a group I ribozyme domain: Principles of RNA packing. *Science* 273(5282): 1678-1685.

Ferre-D'amare, A.R., Zhou, K. & Doudna, J.A. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395(6702): 567-574.

Firdaus-Raih, M., Hamdani, H.Y., Nadzirin, N., Ramlan, E.I., Willett, P. & Artymiuk, P.J. 2014. Cognac: A web server for searching and annotating hydrogen-bonded base interactions in Rna three-dimensional structures. *Nucleic Acids Res.* 42(Web Server issue): W382-388.

Grigg, J.C. & Ke, A. 2013. Structural determinants for geometry and information decoding of Trna by T box leader Rna. *Structure* 21(11): 2025-2032.

Hamdani, H.Y., Appasamy, S.D., Willett, P., Artymiuk, P.J. & Firdaus-Raih, M. 2012. Nassam: A server to search for and annotate tertiary interactions and motifs in three-dimensional structures of complex Rna molecules. *Nucleic Acids Res.* 40(Web Server issue): W35-41.

Hansen, J.L., Ippolito, J.A., Ban, N., Nissen, P., Moore, P.B. & Steitz, T.A. 2002. The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Molecular Cell* 10(1): 117-128.

Jeffrey, G.A. & Saenger, W. 2012. *Hydrogen Bonding in Biological Structures*. New York: Springer Science & Business Media.

Leontis, N.B. & Westhof, E. 2001. Geometric nomenclature and classification of Rna base pairs. *RNA* 7(04): 499-512.

Parlea, L.G., Sweeney, B.A., Hosseini-Asanjan, M., Zirbel, C.L. & Leontis, N.B. 2016. The Rna 3d motif atlas: Computational methods for extraction, organization and evaluation of Rna motifs. *Methods* 103: 99-119.

Petrov, A.I., Zirbel, C.L. & Leontis, N.B. 2011. Webfr3d--a Server for finding, aligning and analyzing recurrent RNA 3D Motifs. *Nucleic Acids Research* 39(Web Server issue): W50-55.

Petrov, A.S., Bernier, C.R., Gulen, B., Waterbury, C.C., Hershkovits, E., Hsiao, C., Harvey, S.C., Hud, N.V., Fox, G.E. & Wartell, R.M. 2014. Secondary structures of rRNAs from all three domains of life. *PLoS One* 9(2): e88222.

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

Pley, H.W., Flaherty, K.M. & McKay, D.B. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* 372(6501): 68-74.

Rose, P.W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., Feng, Z., Green, R.K., Goodsell, D.S., Hudson, B., Kalro, T., Lowe, R., Peisach, E., Randle, C., Rose, A.S., Shao, C., Tao, Y.P., Valasatava, Y., Voigt, M., Westbrook, J.D., Woo, J., Yang, H., Young, J.Y., Zardecki, C., Berman, H.M. & Burley, S.K. 2017. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Research* 45(D1): D271-D281.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C. & Moras, D. 1991. Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA (Asp). *Science* 252(5013): 1682-1689.

Hazrina Yusof Hamdani & Mohd Firdaus-Raih*
Pusat Pengajian Biosains dan Bioteknologi
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

Hazrina Yusof Hamdani
Institut Perubatan dan Pergigian Termaju
Universiti Sains Malaysia
13200 Bertam, Kepala Batas, Pulau Pinang
Malaysia

*Pengarang untuk surat-menyurat; email: firdaus@mfrlab.org

Diserahkan: 21 September 2017
Diterima: 15 Mac 2018