AUTOMATED ALIGNMENT OF RNA 3D STRUCTURES

Ryan Robert Rahrig

A Dissertation

Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Committee: Craig Zirbel, Advisor

Jane Chang,
Graduate Faculty Representative

John Chen

James Albert
ABSTRACT

Craig Zirbel, Advisor

RNA sequence databases contain sequences from hundreds to thousands of homologous molecules. After alignment, these sequences have been successfully mined to reconstruct secondary structures, infer phylogenies, and estimate mutation rates of individual nucleotides and basepairs. Atomic-resolution RNA 3D structures are less numerous, but far more informative than sequence data, as they show, in the ideal case of well-ordered and well-resolved structures, every basepair, base stacking, and base-backbone hydrogen bond.

To function correctly, structured RNA molecules must fold into the correct 3D structure. Since RNA 3D structure is more conserved than RNA primary sequence, detecting structural similarities (and dissimilarities) among RNA 3D structures can produce a wealth of information regarding the functional and evolutionary properties that could not be found by analyzing sequence data alone. The goal of RNA 3D structural alignment is to establish correspondences between the individual nucleotides that are similar in the two 3D structures.

Due to rapid technological developments within the past decade, there has been a dramatic increase in both the size and number of RNA 3D structures that have been crystallized and made available. It is no longer feasible to solely rely on manual comparison of two RNA 3D structures, which can be a labor-intensive and time-consuming process. Consequently, it has become essential to develop tools that are capable of accurately and automatically discovering structural similarities among RNA molecules, which is accomplished through 3D-to-3D alignment.
I first thank God for providing me with the abilities and opportunities to make this accomplishment possible.

Next I would like to thank my advisor, Dr. Zirbel, for all his help, encouragement, and enthusiasm throughout this process. It was a joy having him as an advisor. I am deeply grateful for all the time and effort he put into helping me with this dissertation. I truly admire his passion and dedication to helping students as both a teacher and researcher. I know I will be a better teacher and researcher in the future because of him.

I am also extremely grateful for my loving wife, Jessica, who walked into my life right at the beginning stages of the dissertation process and stuck with me. Her constant love throughout this process has been remarkable. She has been a great source of motivation and perseverance for me during all of the stages of writing this dissertation. I thank God for blessing me with such a wonderful and caring woman as my wife.

I also thank my family for all their love and support throughout my entire life. I especially thank my parents for allowing me to always choose my path in life and supporting me all along the way. They have always been there when I have needed them. They have been great examples of love for me, and I am a better and stronger person as a result. I could not have asked for better parents.

I wish to thank all my friends who helped make my years as a graduate student a more enjoyable and fruitful experience, especially all the friends I made through Creed on Campus. They have been such good friends in faith and have provided so many great memories along the way.
# Table of Contents

**Introduction** 1

**CHAPTER 1: RNA Structure and Alignment** 6

1.1 Basics of RNA Structure ................................. 6
  1.1.1 Comparison with DNA ................................. 7
  1.1.2 RNA Structure—Primary, Secondary, Tertiary ........ 8

1.2 Availability of RNA 3D Structure Data .................. 11

1.3 RNA Alignment ........................................... 12

1.4 Sequence Alignment and Structure Alignment ............ 14

1.5 Alignment Regimes ....................................... 15
  1.5.1 1D-to-1D ............................................ 15
  1.5.2 1D-to-1D, using an inferred 2D structure .......... 16
  1.5.3 1D-to-2D ............................................ 16
  1.5.4 1D-to-3D ............................................ 17
  1.5.5 3D-to-3D ............................................ 17

1.6 Current 3D-to-3D Alignment Methods ................... 17
  1.6.1 By Hand .............................................. 17
  1.6.2 ARTS ................................................ 18
  1.6.3 SARA ................................................ 18
  1.6.4 DIAL ............................................... 19
CHAPTER 2: Distance Matrix Methods

2.1 Introduction to Distance Matrices .......................................................... 23
2.2 Benefits of Using Distance Matrices ......................................................... 24
2.3 Theory of NP-Completeness ....................................................................... 30
2.4 SMP is NP-complete .................................................................................. 32
2.5 ε-SMP is NP-complete ............................................................................... 34
2.6 ε-SMP implementation .............................................................................. 35
2.7 Relative Deviations .................................................................................... 35
2.8 RD-SMP is NP-Complete .......................................................................... 36
2.9 Modified Distance Matrices ...................................................................... 38
2.10 Distance Matrix Overlap .......................................................................... 38

CHAPTER 3: A New Approach to 3D-to-3D Alignment

3.1 Goals of the New Approach ..................................................................... 39
3.2 Local Alignments and the Local Alignment Graph ................................. 41
3.2.1 Form Local Neighborhoods ................................................................. 41
3.2.2 Measuring Structural Similarity .......................................................... 44
3.2.3 Alignment Objectives .......................................................................... 45
3.2.4 Graph Theory Terminology ................................................................ 45
3.2.5 Vertices of the Local Alignment Graph .............................................. 46
CHAPTER 4: R3D Align

4.1 Vertex Preprocessing
   4.1.1 Lower Bound on the Size of the Maximum Clique
   4.1.2 Graph Coloring
   4.1.3 Neighbor Counting
   4.1.4 Specialized Pre-processing

4.2 Controlling the Number of Neighborhoods

4.3 Seed Alignments

4.4 Screening Criterion

4.5 Greedy Cliques Produce Alignments

4.6 Branch and Bound

CHAPTER 5: Alignment of Bacterial 16S rRNA Structures

5.1 Tools and Diagnostics for Displaying and Evaluating Alignments
   5.1.1 Displaying Alignments
   5.1.2 Spreadsheet vs. 2D matrix and 3D Superposition

5.2 Comparison of 16S Alignments in Spreadsheet Form

5.3 Alignment Diagnostics

5.4 “Bar Diagram” Display and Evaluation Tool

CHAPTER 6: Applications of R3D Align

6.1 Building Consensus Interaction Lists

6.2 Improving Existing Alignments

6.3 RNase P Alignment

6.4 Alignment of 23S rRNA Structures
List of Figures

1.1 23S rRNA crystal structure ............................................. 6
1.2 5S rRNA sequences ......................................................... 7
1.3 DNA double helix, found in crystal structure 1hcd ................. 8
1.4 *E. c.* secondary structure diagram .................................. 9
1.5 Edges available for basepairing ........................................ 10
1.6 *cis* and *trans* orientations of the glycosidic bonds ............ 11
1.7 Aligned 5S rRNA sequences in standard form ....................... 13
1.8 Aligned 5S rRNA sequences in alternative form .................... 14
2.1 Helical region in *T. Th.* consisting of 3 basepairs and one extruding base .... 23
2.2 Helical region in *T. Th.* consisting of 3 basepairs ............... 25
3.1 5S rRNA Loop B 3D structures ........................................ 42
3.2 Secondary structure diagram for *E. c.* 22-27, 60-61 ................ 43
3.3 Local neighborhoods of nucleotide 22^A^ ............................. 43
3.4 Neighborhoods list ......................................................... 44
3.5 Pairwise geometric discrepancies ..................................... 46
3.6 Vertices of the local alignment graph ................................ 47
3.7 Incompatible local alignments .......................................... 49
3.8 Compatible local alignments ............................................ 49
3.9 Complete local alignment graph ....................................... 50
6.10 D.r. (2zjr) and E.c. (2aw4) 23S Alignments ........................................ 105
6.11 H.m. (1s72) and E.c. (2aw4) 23S Alignments ........................................ 106

7.1 3D scatterplot of translation vectors for files 2uub and 2uuc .................. 112
7.2 Color bar for 3D scatterplots ................................................................. 113
7.3 Neighborhoods of nucleotides 1257 in 2uub and 2uuc in global superposition 113
7.4 Nucleotides 1019-1045 of 2uub and 2uuc in global superposition .......... 114
7.5 3D scatterplot of rotation vectors for files 1n33 and 1n34 ..................... 115
7.6 Superposition of local neighborhoods of nucleotide 1141 of 1n33 and 1n34 . 116
7.7 3D scatterplot of translation vectors for files 1j5e and 1fjf ................... 120
7.8 3D scatterplot of rotation vectors for files 1j5e and 1fjf ....................... 121
7.9 Extreme values for 1j5e and 1fjf data ...................................................... 122
7.10 3D scatterplot of translation vectors for files 1j5e and 2zm6 ............... 123
7.11 3D scatterplot of rotation vectors for files 1j5e and 2zm6 ................. 124
7.12 Extreme values for 1j5e and 2zm6 data ............................................... 125
7.13 3D scatterplot of translation vectors for files 1j5e and 1hnw ............... 127
7.14 3D scatterplot of rotation vectors for files 1j5e and 1hnw ................. 128
7.15 3D scatterplot of translation vectors for files 1j5e and 1hnx ............... 129
7.16 3D scatterplot of rotation vectors for files 1j5e and 1hnx .................. 130
7.17 Largest values of $D^2$ for 1j5e and 1hnx based on the translation vector data 131
7.18 3D scatterplot of translation vectors for files 1j5e and 2avy .......... 132
7.19 3D scatterplot of rotation vectors for files 1j5e and 2avy ................. 133
7.20 Extreme values for 1j5e and 2avy data ............................................... 134
7.21 Mutual translation vector total sample variance matrix ..................... 135
7.22 Dendrogram of structures equivalent sequentially with 1j5e ............ 136
Introduction

This dissertation is a work in the field of bioinformatics. Bioinformatics is an emerging interdisciplinary field which combines mathematical and computer sciences with the life sciences. It requires the development and implementation of algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.

This work focuses on using mathematical and computational techniques to study the structure of RNA molecules. Specifically, the focus will be on the problem of aligning two RNA molecules, using their known three-dimensional (3D) structures which have been determined by a process known as x-ray crystallography. The objective in RNA alignment is to establish correspondences between individual nucleotides that are similar in the two structures. Aligned nucleotides may be similar in sequence, structure, evolution, and/or function. When performing a structural alignment (the type of alignment we focus on), the correspondences are based on structural similarities, although additional types of similarities are also present. The formal definition of an RNA alignment is given in Section 1.3. When we align two RNA molecules whose 3D structures are both known, we refer to this as 3D-to-3D alignment (Section 1.5).

To function correctly, structured RNA molecules must fold into the correct 3D structure. Even so, various changes to the sequence may be structurally neutral. RNA structures are therefore more conserved over the course of evolution than the sequences are. Detecting structural similarities (and dissimilarities) among RNA 3D structures can then produce a
wealth of information regarding the functional and evolutionary properties that could not be found by analyzing sequence data alone.

Due to rapid technological developments within the past decade, there has been a dramatic increase in both the size and number of RNA 3D structures that have been crystallized and made available. It is no longer feasible to solely rely on manual comparison of two RNA 3D structures, which can be a labor-intensive and time-consuming process. Consequently, it has become essential to develop tools that are capable of accurately and automatically discovering structural similarities among RNA molecules, which is accomplished through 3D-to-3D alignment.

Due to the complex nature of RNA structure (Section 1.1), aligning two molecules based on their structures is a difficult task. While efficient sequence alignment methods have been available for decades, only within the last several years have methods for 3D-to-3D alignment been developed. However, as shown Chapter 5, these methods are often inaccurate because they ignore so much of the information contained in the 3D structures in order to achieve computational efficiency. This is often the trade-off one is faced with when producing a 3D-to-3D RNA alignment. It is difficult to incorporate enough information to produce an accurate alignment while maintaining computational feasibility. The main focus of this dissertation will be doing just that—designing and implementing a new method that is capable of producing highly accurate alignments in a reasonable amount of time.

This dissertation focuses on both the theoretical and practical aspects of RNA 3D-to-3D alignment. On the theoretical side, it is shown that detecting structural similarities of two RNA structures is truly indeed a difficult problem by proving four different formulations of the problem to be NP-complete (theory of NP-completeness is described in Section 2.3).

In the first series of results, discussed in Chapter 2, the 3D coordinates of each RNA molecule are used to calculate nucleotide-to-nucleotide Euclidean distance matrices. We attempt to derive a 3D-to-3D alignment by aligning the distance matrices of the two structures. Since our goal is to produce a structural alignment, the subsets of aligned nucleotides have
similar distance matrices. Since the distance matrices of the subsets of nucleotides are indeed submatrices of the original distance matrices, we reconstruct our problem to one of finding similar submatrices. Specifically, we make our objective to find the largest submatrices such that the corresponding entries are all within $\varepsilon$. We note that the submatrices must be formed in such a way that row $i$ is removed whenever column $i$ is removed in order to correspond to the removal of nucleotide $i$ from the alignment. However, it is shown that this problem is, in fact, NP-complete. By proving this problem is NP-complete, it is shown that even if all of the information contained within an RNA 3D structure is distilled into a simpler two-dimensional form, optimal alignment may still not be efficiently achieved.

As mentioned above, the main focus of this work is the design and implementation of a new 3D-to-3D method that is capable of producing highly accurate alignments in a reasonable amount of time. The framework for an algorithm which accomplishes this task is presented in Chapter 3. Unlike the other methods available, the algorithm is able to incorporate all types of interactions between nucleotides, whether they be base-pairing or base-stacking, local or long-range. It is also based on local 3D conformations rather than global conformations which accommodates the flexibility of RNA molecules.

In the algorithm, the alignment problem is transformed into the problem of finding the maximum clique of a specially defined graph, called the local alignment graph. To form this local alignment graph, first all 4-nucleotide neighborhoods are calculated that can be formed such that the maximum distance between any two of the four nucleotides in the neighborhood is less than the user-defined diameter, $D$. These neighborhoods are formed in such a way to include interactions of all types, as outlined above. The neighborhoods of the two different structures are then compared and evaluated for structural similarity. The objective is then to find the alignment that includes the maximum number of structurally similar neighborhoods. This is another problem that is shown to be NP-complete (Section 3.3).

Because of its practical importance, an efficient solution to the problem is still sought
despite it being NP-complete. A graph is defined in which the vertices of the graph repre-
sent structurally similar neighborhoods and edges are drawn between vertices if the two 
4-nucleotide local alignment they represent can be merged to form a valid larger alignment. 
Finding the maximum clique of this local alignment graph is then the critical step to finding 
the maximum alignment. A branch and bound technique is used to find the maximum clique, 
which is then translated back into the alignment it represents.

The implementation of this algorithm is presented in Chapter 4 along with numerous 
techniques that have been developed and implemented that increase the efficiency of the 
algorithm. For example, since we are not working with arbitrary graphs, but instead with 
local alignment graphs which by design have particular features that can be found in all local 
alignment graphs, we can exploit some of these features to find the maximum clique more 
quickly than with arbitrary graphs. These techniques are described as are other techniques 
used to increase efficiency, such as using graph colorings, seed alignments, greedy maximum 
clique algorithms, and a branch and bound algorithm. The algorithm has been implemented 
in the Matlab programming language in a suite of programs called “R3D Align.”

The R3D Align program suite also includes new diagnostic tools for the structural eval-
uation and display of RNA alignments, which are presented in Chapter 5. These tools were 
used to evaluate the alignment of two 16S rRNA molecules as produced by R3D Align. The 
alignments produced by other structure-based and sequence-based alignment methods were 
also evaluated and compared with R3D Align. The results showed the R3D Align align-
ment to be the most accurate. Furthermore, R3D Align was shown to have computational 
feasibility as the alignment of the two large structures was produced in thirteen minutes.

In Chapter 6 further uses and applications of the alignment program are presented. 
In Section 6.1 it is discussed how a 3D-to-3D alignment can be used help create a better 
probabilistic model for the stochastic context-free grammar/Markov random field 1D-to-3D 
alignment program, JAR3D. In Section 6.2 it is shown that R3D Align can actually be used 
to improve existing alignments produced by other 3D-to-3D alignment programs. Section
shows that how R3D Align can successfully be used to align non-ribosomal RNA, such as RNaseP.

Chapter 7 describes a new technique to quickly detect and analyze conformational changes among two RNA 3D structures. The translations and rotations of the optimal superpositions of local neighborhoods are found and compared with the translation and rotation of the optimal superposition of the entire structure using multivariate statistical techniques.

In the last chapter, Chapter 8, the user manual for R3D Align is presented.
CHAPTER 1

RNA Structure and Alignment

1.1 Basics of RNA Structure

Figure 1.1: 23S ribosomal RNA molecule, from the 2aw4 crystal structure, which consists of 2841 nucleotides [60]

RNA, a nucleic acid molecule that plays an important role in protein synthesis and many
other cellular functions, is a single stranded molecule made up of individual units called nucleotides. Each nucleotide consists of three components—the base, the sugar, and the phosphate. The sugars and phosphates of adjacent nucleotides covalently bond to form a sugar-phosphate backbone for the RNA molecule. There are four types of bases found in RNA: Adenine, Cytosine, Guanine, and Uracil. Typically, these bases are represented by their first letters: A, C, G, and U. Because RNA molecules are single-stranded, each molecule has a distinct beginning (known as the 5’ end) and end (the 3’ end). The sequence of the bases then has a unique ordering when starting from the 5’ end. The sequences of the 5S ribosomal RNA (rRNA) molecules, Escherichia coli (E.c.) and Thermus Thermophilus (T.th.) are given in Figure 1.2.

```
E.c. GCCUGGCGGCCGUAGCGCGGUGGUCCCACCUGACCCCAUGCCGAACUCAGAAGUGAAACGCCGUAGCGCCGAUGGUAGUGUGGGGUCUCCCCAUGCGAGAGUAGGGAACUGCCAGGC
T.th. UCCCCUGCGGCCGUAGCGCGGUGGAACCACCCGUUCCCAUUGCAGCCGAACUUGGCGCCGAUGCCAGCGGAGAGUAGGUCGGUGCGGGGG
```

Figure 1.2: Sequences of bases for the 5S rRNA molecules, E.c. and T.th.; E.c. has 117 bases and T.th. has 119 bases.

### 1.1.1 Comparison with DNA

RNA is similar to DNA, but there are a few important structural differences: 1) RNA is single-stranded, while DNA is double-stranded; 2) the sugar in RNA contains ribose while the sugar in DNA contains deoxyribose; and, 3) RNA has the base Uracil (U) rather than Thymine (T) that is present in DNA.

The two strands in DNA are held together by covalent bonds between bases. The strands are complementary in that G binds with C and A binds with U with very few exceptions. The complementary strands join together and form a double helix, as in Figure 1.3. The structure of RNA, on the other hand, is formed as the single strand of nucleotides folds hierarchically to form specific 3D structures, helical and non-helical, necessary for molecular function.
1.1.2 RNA Structure—Primary, Secondary, Tertiary

When the RNA molecules fold, the nucleotides interact in a variety of ways, but the best understood interactions involve bonds between the bases, although recent studies have focused on the interactions between bases and phosphates [73]. Two bases that are covalently bonded are said to be forming a basepair. The nucleotide sequences (the primary structure) fold back on themselves to align Watson-Crick complementary regions, forming Watson-Crick helices. The helices are mainly composed of the canonical Watson-Crick basepairs, AU, UA, CG, and GC. The formation of these helices provides the basis for the RNA secondary structure. The secondary structure of the E.c. 5S rRNA molecule is given in Figure 1.4.

The helices found in RNA are generally short (usually less than 15 Watson-Crick pairs). On the ends of helices are found nominally unpaired stretches of sequence that are called “loops”. These loops are classified into several types, depending on where they occur in the secondary structure. Hairpin loops serve as caps on the ends of helices, internal loops are found between two helices, and multi-helix junction loops are found linking three or more
helices. In Figure 1.4, Loop C and Loop D are hairpin loops, Loops B and E are internal loops, and Loop A is a multi-helix junction loop.

While bases in loops are usually depicted in secondary structures as not forming basepairs, these “loops” in fact are very often highly structured by the formation of several types of non-Watson-Crick basepairs [63]. Basepairs are grouped into families according to the interacting edges of the bases. RNA bases have three edges available for hydrogen-bonding interactions (Figure 1.5). These edges are known as the Watson-Crick edge, the Hoogsteen edge, and the Sugar edge. Basepairs in helices are formed by interactions between the Watson-Crick edge of each base. These are known as Watson-Crick basepairs. However, basepairs can be formed between any of the six possible combinations of the three edges. For example, a Watson-Crick edge can interact with a Sugar edge, two Sugar edges can interact, a Hoogsteen edge can interact with a Watson-Crick edge, etc. Basepairs in which at least one of the interacting base edges is not the Watson-Crick edge are known as non-Watson-Crick basepairs. Non-Watson-Crick basepairing is another RNA characteristic that is not typically found in DNA.
In addition, each of the 6 combinations of edges can interact in two different ways, according to the orientations of their glycosidic bonds (the bonds between the base and sugar of a nucleotide). In Figure 1.6, the orientations of the glycosidic bonds are indicated by arrows. In a cis basepair, both arrows point to the same side of the gray line, which is drawn according to the hydrogen bonds that hold the bases together. In a trans basepair, the arrows point to opposite sides of the gray line.

So there are 12 different geometric basepair families in total. Basepair families are abbreviated according to the orientation (cis or trans) and the two interacting edges. For example, the abbreviation ‘cWW’ is used to denote a cis Watson-Crick Watson-Crick basepair, the type of basepairs found in helices; ‘tSH’ denotes a basepair with the trans orientation in which the Sugar edge of one base is interacting with the Hoogsteen edge of the other. A tSH basepair is an example of a non-Watson-Crick basepair, typically found in loops.

The final three-dimensional structure of RNA is formed as the distant elements of the
secondary structure make “tertiary” contacts and fold up into the full 3D structure of the molecule, called the tertiary structure (as in Figure 1.1). Tertiary contacts may be formed by two hairpin loops folding in and interacting with one another, or by a hairpin interacting with a helix, or by interactions between two different helices. Non-Watson-Crick basepairs are typical among these long range tertiary interactions.

1.2 Availability of RNA 3D Structure Data

RNA primary sequences are readily available and have been so for decades. Thousands and thousands of RNA sequences have been determined and are accessible in sequence databases. Knowledge of the secondary structure and especially the 3D structure is more difficult to obtain. It has only been in recent years that full knowledge of the 3D structures of certain RNA molecules has become available.

Roughly 350 distinct RNA molecules from various organisms have been successfully crystallized and their 3D structures determined to atomic precision by x-ray crystallography [74].
While these atomic-resolution RNA 3D structures are less numerous, they are far more informative than sequence data, as they show, in the ideal case of well-ordered and well-resolved structures, every basepair, base stacking, and base-backbone hydrogen bond.

The 3D structure of small RNA molecules can also be determined by nuclear magnetic resonance spectroscopy (NMR). Some structures have been determined, but not to atomic precision, by cryo-electron microscopy techniques. 3D structures produced by any of these methods are generally deposited with the Protein Data Bank (PDB) \cite{PDB} and its partner, the Nucleic Acid Data Bank (NDB) \cite{NDB}. Although less numerous than sequence data, the size and number of RNA 3D structures have been increasing steadily throughout the past decade.

It is important to note that the PDB/NDB archives all structures deposited, which means that some structures are later superseded by improved versions. These resources have about 1700 3D structures containing RNA data, although only about 350 of them are distinct. For example, the original PDB entry for the 3D structure of the 23S ribosomal RNA molecule of *Haloarcula marismortui* (*H.m.*) was 1ffk, later superseded by 1jj2, and most recently by 1s72.

Also, in macromolecular crystallography it is often possible to diffuse a small molecule into the crystal to bind to the crystallized macromolecule at specific sites without changing the crystal form or drastically affecting the structure \cite{66}. This results in closely-related 3D structures, not brand-new crystallizations. For example, there are roughly 50 3D structures of *H.m.* 23S rRNA.

### 1.3 RNA Alignment

An alignment of RNA molecules is a list of correspondences between individual nucleotides from two or more RNAs. Two nucleotides that correspond are said to be aligned with one another. Not all nucleotides need to have a correspondence. Any nucleotide that is not
aligned with another nucleotide is said to be aligned with a ‘gap’, which is represented by the symbol ‘-’. Two nucleotides may be aligned because they are related by structure, by function, and/or by evolution.

In every alignment, we require that the correspondences be well-ordered (if nucleotides \(i\) and \(j\) are aligned with nucleotides \(i'\) and \(j'\), respectively, then \(i < j\) if and only if \(i' < j'\)) and uniquely-assigned (no nucleotide should be aligned to more than one nucleotide from the other RNA). Alignments are typically represented as matrices in which each row contains an individual sequence and columns identify nucleotides that are aligned. In Figure 1.7, an alignment of the sequences of the 5S rRNA molecules, \(E.c\) and \(T.th\). is given. The unaligned sequences were given in Figure 1.2.

![Figure 1.7: An alignment of the 5S rRNA sequences \(E.c\) and \(T.th\). Nucleotides in the same column are aligned, while a ‘-’ in a column represents a nucleotide aligned with a gap (i.e., it is unaligned).](image)

Most of the time two aligned nucleotides are regarded as being homologous. That is, they are believed to have a common ancestor. Nucleotides without a correspondence (i.e., unaligned nucleotides) are the result of insertion and/or deletion mutations over the course of the evolution of the molecules. Insertions (deletions) occur when nucleotides are added to (removed from) the molecule from one generation to the next. Aligned nucleotides that are not the same base occur as a result of base substitutions that take place over time. Base substitutions occur when a base of one type is replaced by a base of another type from one generation to the next. For example, an A may be replaced by a G in the structure. The well-ordered criterion has its basis in the observation that while insertions, deletions, and base substitutions occur relatively frequently, it is much more rare for two nucleotides to swap positions in the sequence from one generation to the next.
With a pairwise alignment, it is sometimes useful to display the alignment in the alternative fashion of listing the nucleotides of each molecule and drawing lines to connect corresponding nucleotides. By the well-ordered property, the lines must be non-crossing in order for the alignment to be valid. By the uniquely-assigned property, no nucleotide shall be connected to more than one line. Figure 1.8 gives this alternative representation of the alignment of \textit{E.c.} and \textit{T.th.} given in Figure 1.7.

Figure 1.8: The alignment of the 5S rRNA sequences \textit{E.c.} and \textit{T.th} using lines to connect aligned nucleotides.

### 1.4 Sequence Alignment and Structure Alignment

The first alignment methods developed used only the primary sequence data. Alignments of RNA sequences can provide great insight into the structures, functions, and evolutionary histories of RNA molecules. For highly similar sequences, e.g., 5S rRNA \cite{22} \cite{53}, an alignment based solely on sequence similarity will also correctly align higher-order structural features. However, because there are only four bases, the ability to produce good alignments by sequence similarity diminishes rapidly as sequence conservation decreases \cite{22}. Knowledge of the underlying structure then becomes an essential guide to the alignment procedure \cite{9}.

To function properly, structured RNA molecules must fold into the correct 3D structure, with the result that 3D structures are highly conserved over many, many generations. Secondary structures are also typically conserved, although less so than 3D structures. The primary nucleotide sequence is the most variable.
RNA can accommodate high sequence variation with low structural variation because *isosteric* basepair substitutions can take place. Two different basepairs are regarded as being isosteric if one of the basepairs can replace the other without affecting the overall 3D structure. The case of covariation between CG, GC, AU, UA Watson-Crick basepairs is well known, but all 12 basepair families have mutually isosteric subgroups [41]. However, base stacking, base-backbone interactions, and other factors limit base variability in ways that are not yet fully understood.

Because there is less variability among the structures than the sequences, better alignments can be produced when more of the structural information is used. As new technologies have been developed that have made more RNA structural information more available, new alignment methods have been developed that take into account this structural information. We will now distinguish a number of different alignment regimes based on the type of information that is taken into account.

### 1.5 Alignment Regimes

#### 1.5.1 1D-to-1D

Here, only the primary sequence information of each molecule is used to produce the alignment. Given two sequences, determine the set of correspondences which produce an optimal score where identical or similar bases give a positive contribution and unaligned or dissimilar bases give a negative contribution. Characters are scored one at a time, and thus, no secondary structure or other interaction information is taken into account. Optimal global sequence alignment is exemplified by the Needleman-Wunsch algorithm [46], which is based on dynamic programming and runs in \( O(n^2) \) time.
1.5.2 1D-to-1D, using an inferred 2D structure

Given two RNA sequences, a consensus secondary structure and alignment of the sequences is simultaneously determined. As above, only the primary sequence information of each molecule is used as input, but predicted folding possibilities are used to aid in the alignment of the sequences. For example, if one set of nucleotides is predicted to fold into a helix based on its sequential information, then that set should correspond with a set of nucleotides in the other structure that is also predicted to have the capability of folding into a helix. The optimal solution to this approach is exemplified by the Sankoff algorithm [57]. Being expensive in both time and memory requirements, $(O(N^6)$ and $O(N^4)$, respectively), more efficient, although restricted, versions of Sankoff’s algorithm are implemented in programs such as FoldAlign [28] [31] and Dynalign [43]. Another approach falling in this category is PMmatch [32] [33], which finds an alignment using the base pairing probability matrices predicted by McCaskill’s algorithm [44].

1.5.3 1D-to-2D

Here, a new RNA sequence is aligned to a consensus secondary structure model, which is used to represent a set of related organisms. Alignment of two RNA sequences to the model implicitly makes an alignment of the two sequences to each other. This approach is exemplified by Infernal [18] [45], which uses an RNA Stochastic Context-Free Grammar based covariance model [19] [56]. The approach, which runs in $O(n^2)$ time, is also able to take into account some of the long-range interactions found in the helices of RNA molecules. Pseudoknots, interactions within loops, and tertiary interactions are not accounted for since secondary structures do not include that information. (Pseudoknots are non-nested basepairs; i.e., two basepairs $(i, j)$ and $(i', j')$ such that either $i < i' < j < j'$ or $i' < i < j' < j$.)
1.5.4 1D-to-3D

The 3D structure of one molecule is used to determine a probabilistic model in conjunction with basepair frequencies and then new RNA sequences are aligned to that model. Aligning two sequences to the model implies an alignment of the sequences to one another. This approach is described in [58] and exemplified by the $O(n^3)$ method, JAR3D (under development). The Infernal alignment project has been very successful in turning a manual alignment and consensus secondary structure into a model that can be used to accurately align new sequences. This is the basis for the RFAM alignments [29]. The JAR3D project aims to do similar things with models based on 3D structure.

1.5.5 3D-to-3D

In a 3D-to-3D alignment, the full 3D structures of both molecules to be aligned are known and used in the alignment process. This is the focus of this work. Current 3D-to-3D alignment methods will now be described.

1.6 Current 3D-to-3D Alignment Methods

As we’ve discussed, better alignments should be achieved if more of the 3D structural information is considered. But it has only been within the last decade that the number and size of the atomic-resolution RNA 3D structures available has increased to a point where automated methods for 3D structural alignment are needed. There are currently several methods and programs used for performing RNA 3D-to-3D alignment, which are briefly reviewed in this section.

1.6.1 By Hand

Like sequence alignments, the highest quality 3D-to-3D alignments for RNA still require hand curation by experts, in a labor-intensive effort. However, hand alignment is time-
consuming and labor intensive. Such an alignment may take several weeks to complete for large RNA structures, reinforcing the need for automatic methods, such as those described next.

1.6.2 ARTS

ARTS [15] represents each nucleotide by a single point in space, using the position of the phosphorus atom in the backbone. ARTS seeks the rigid transformation that superimposes the largest number of phosphate atoms of one structure onto the phosphate atoms of another structure. A weighted scoring function is maximized that is based on both the number of corresponding nucleotides and basepairs. This method can be used to detect previously unknown common substructures, which can contain hundreds and even thousands of nucleotides. They could also be small local motifs, as long as they contain at least two successive base pairs. Unlike many programs, ARTS is not constrained to preserve linear order within the alignment (the alignments need not be well-ordered). Because of this, ARTS is able to discover non-sequential common substructures. ARTS is a heuristic algorithm with theoretical worst case complexity of $O(n^3)$, where $n$ is the number of nucleotides in the larger of the two structures. ARTS takes no account of nucleotide identity or similarity.

1.6.3 SARA

SARA [10] performs pairwise RNA 3D-to-3D alignment by implementing a unit-vector representation of RNA structures that calculates a set of vectors between consecutive atoms of a user-selected type [37]. $O(n^2)$ time dynamic programming algorithms are then applied to the reduced unit-vector representation of the structure. SARA does not require the assignment of basepairs as ARTS does, although a more accurate alignment is produced when such information is provided. SARA provides a statistical assessment of the significance of the resulting alignment. The SARA algorithm was inspired by the MAMMOTH program for protein structure alignment [50].
1.6.4 DIAL

DIAL [20] is an extension of the program PRIMOS [17] that aligns RNA backbone conformations and accounts for dihedral angle similarity, sequence similarity, and base pairing similarity. Basepairs are only accounted for in a simplistic way: each nucleotide is assigned a ‘0’, ‘L’, or ‘R’ based on whether it is not part of any basepairs, paired with a base to the left of it in the sequence, or paired with a base to the right of it in the sequence, respectively. DIAL calculates a representation of the RNA 3D structure in terms of the six dihedral angles $(\alpha, \beta, \gamma, \delta, \epsilon, \zeta)$ used to define RNA backbone conformations and the dihedral angle $(\chi)$ that describes rotation about the glycosidic bond. Two additional virtual angles $\eta$ and $\theta$, first introduced by Olson [49] and later reintroduced by Duarte, Wadey and Pyle [17] offer a reduced but sufficient conformational description of the RNA backbone [67].

DIAL devises a scoring system based on the dihedral angle similarity, sequence similarity, and base pairing similarity and then employs classical dynamic programming algorithms to achieve pairwise global and local alignment between two 3D RNA structures in $O(n^3)$ time using a scoring function that includes terms for similarity in nucleotide sequence, basepairing, and the nine torsional angles. A third type of alignment, known as ‘semiglobal’ alignment, uses a variation of the Needleman-Wunsch algorithm in which gaps are not penalized at the beginning or end of the alignment and can be used to perform motif searches. DIAL runs RNAVIEW [71] to determine the basepairing status of each nucleotide. Unlike the PRIMOS algorithm, which can only consider gapless alignments of pseudo-dihedral angles for contiguous sequences, DIAL can handle fragmented queries and alignments with bulging nucleotides by means of gap insertion.

1.6.5 SARSA

SARSA [12] is a web tool that can be used to align two RNA structures using the PARTS application. PARTS uses an approach that reduces the 3D structure into a 1D sequence of letters selected from an RNA structural alphabet of 23 letters that represent distinct and
common backbone conformations. Classical sequence alignment methods are then employed on these 1D encoded sequences. Like DIAL, PARTS can perform global, local, and semi-global alignment in quadratic time. SARSA also has a tool, MARTS, which is able to perform global multiple alignment of RNA 3D structures using multiple sequence alignment algorithms.

1.7 Uses for a 3D-to-3D alignment

There are a wide variety of uses and applications of RNA 3D-to-3D alignment, some of which will now be described.

1.7.1 Provide Sequence Variability Information

3D-to-3D alignment methods that do not heavily rely on sequence data and can produce highly accurate alignments at the individual nucleotide level, such as the method we have developed (Chapter 3), produce a wealth of information regarding natural variations which occur between the sequences of the aligned RNA molecules. Sequence variability information is highly relevant to sequence-based endeavors such as multiple sequence alignment and phylogenetic tree building. Consequently, the more we learn about RNA 3D structure and the variations which occur between the 3D structures of homologous molecules, the more we can bring to bear on the alignment problem.

1.7.2 Compare Large Structures from Different Organisms

The case of 23S rRNA, previously discussed, illustrates another instance in which automated 3D-to-3D alignment is useful. The 23S has been crystallized in four organisms: Haloarcula marismortui (H.m.), Thermus thermophilus (T.th.), Escherichia coli (E.c.), and Deinococcus radiodurans (D.r.). H.m. is an archaeon, the rest are bacteria. An alignment of these four organisms is provided in Section 6.4. As already noted, in each case, the PDB file may
contain multiple versions of the molecule from the same research lab. In addition, the *T. th.*
23S has been solved to atomic precision by two different groups [62] [38], and the two 3D
structures resolve different regions to different levels of resolution. For example, 1vsa [38]
contains helices 42, 43, and 44, which are not resolved in 2j01 [62], but 2j01 has more
basepairs per nucleotide as identified by FR3D [59], suggesting that it may be more useful
for certain purposes. Because of primary sequence variability, it is non-trivial to align the
sequences of 23S rRNAs from different organisms. Thus, 3D-to-3D methods that utilize
structural information and do not rely solely on sequential information are necessary. Also,
since the size of the 23S is large (2800 nucleotides), manual comparison of two structures
would be time-intensive. And since there are four organisms to compare, with multiple
crystallizations of each, it is not feasible to rely on manual comparison of the organisms; an
automated 3D-to-3D method is needed.

1.7.3 Reveal Functional and Evolutionary Properties

3D-to-3D alignments are more robust than sequence alignment, because the structures are
less variable. Thus, 3D-to-3D alignment can produce a wealth of information regarding the
functional and evolutionary properties that could not be found by analyzing sequence data
alone.

1.7.4 Identify Motif Swaps

3D-to-3D alignments can be used to identify motif swaps. By noticing regions which align
well at a local level, they can be used to identify conserved local RNA interaction structure,
such as basepair interactions. They also help identify sequence variations that preserve local
3D structure between different organisms.
1.7.5 Create Consensus Models

3D-to-3D alignments could also be used to create a consensus model of the structure and basepairs for a sequence variability model. Consensus models are used in models that make use of a stochastic context free grammar covariance model, such as the 1D-to-2D alignment method Infernal and the 1D-to-3D method JAR3D. This will be explained in detail in Section 6.1.
CHAPTER 2

Distance Matrix Methods

2.1 Introduction to Distance Matrices

An instructive approach to alignment is to reduce an RNA 3D structure with \( m \) nucleotides to an \( m \times m \) “distance” matrix, in which the \((i, j)\) entry represents the Euclidean distance between the geometric centers of the bases of nucleotides \( i \) and \( j \). Distance between bases are typically measured in Ångstroms (1 Ångstrom = 1.0 \( \times \) 10\(^{-10}\) meters). The distance matrix then is always a square, symmetric matrix, with entries equal to zero on the diagonal. Figure 2.1 displays a helical region of the 1j5e crystal structure [70] of \textit{T.thermophilus}, which consists of three Watson-Crick basepairs with an extruding base between two of the basepairs.

![Figure 2.1: Helical region in T.Th. consisting of 3 basepairs and one extruding base. Nucleotides 30 and 553 are forming a cWW basepair, as are nucleotides 32 and 552, and 33 and 553, respectively. Nucleotide 31 is the bulged base.](image)
The corresponding distance matrix, $D^A$, is given by

$$
D^A = \begin{bmatrix}
U30 & G31 & A32 & A33 & U551 & U552 & A553 \\
0.000 & 12.757 & 4.885 & 8.322 & 9.853 & 8.203 & 5.706 \\
12.757 & 0.000 & 12.351 & 14.634 & 19.610 & 17.510 & 15.023 \\
4.885 & 12.351 & 0.000 & 3.722 & 7.325 & 5.681 & 4.955 \\
8.322 & 14.634 & 3.722 & 0.000 & 5.634 & 5.375 & 7.081 \\
9.853 & 19.610 & 7.325 & 5.634 & 0.000 & 3.901 & 7.553 \\
8.203 & 17.510 & 5.681 & 5.375 & 3.901 & 0.000 & 4.085 \\
5.706 & 15.023 & 4.955 & 7.081 & 7.553 & 4.085 & 0.000
\end{bmatrix}
$$

### 2.2 Benefits of Using Distance Matrices

Using distance matrices helps reduce the information contained in a 3D structure into a more manageable two-dimensional form. Although the distance matrix is a greatly reduced representation of the structure, it still contains enough information to reconstruct the 3D structure by distance geometry methods, except for overall chirality [30].

Distance matrices have long been used for comparing protein conformations. Phillips [54] and Nishikawa and Ooi [47] were among the first to recognize important uses for distance matrices in certain aspects of molecular biological analysis. A popular protein alignment tool, DALI [34], uses a Metropolis Monte Carlo optimization algorithm to maximize a similarity score based on the corresponding entries of the two distance matrices obtained by including only the nucleotides in the structures that are aligned with corresponding nucleotide.

Distance matrices can be of great use in the alignment of structures because structurally similar regions have similar inter-nucleotide distances and thus have similar distance matrices. Given an alignment, two distance matrices can be formed by only using the nucleotides in the alignment with a correspondence. We can then evaluate the quality of an alignment by comparing the distance matrices of the two sets of aligned nucleotides. We note that
these distance matrices are submatrices of the original distance matrices, which are formed by deleting rows and columns that correspond to nucleotides aligned that are not aligned to a nucleotide in the other structure. The better the alignment is, the more similar the two distance matrices will be.

In the first step of analyzing the similarity of two distance matrices $A$ and $B$, we obtain a matrix $S$ consisting of the absolute differences; that is, $S(i, j) = |A(i, j) - B(i, j)|$ for each $i, j$. If the two structures match fairly well, then the entries of $S$ will be small. We define the distance between two matrices to be the maximum entry of $S$. That is,

$$d(A, B) = \max_{i,j} |A(i, j) - B(i, j)|.$$

Figure 2.2 displays a helical region with three Watson-Crick basepairs with no bulged bases.

![Figure 2.2: Helical region in $T. Th.$ consisting of 3 basepairs. Nucleotides 612 and 628 are forming a cWW basepair, as are nucleotides 613 and 627, and 614 and 626, respectively.](image)

The corresponding distance matrix, $D^B$, is given by
If we were to align the two structures shown in Figures 2.1 and 2.2, the most obvious alignment would be to align each of the bases involved in a basepairing interaction and exclude the bulged base (align it with a gap). That is, we propose the following alignment:

\[
\begin{align*}
U & \quad G & \quad A & \quad A & \quad U & \quad U & \quad A \\
C & \quad - & \quad C & \quad C & \quad G & \quad G & \quad G
\end{align*}
\]

We obtain a distance matrix $D^A^*$, which is obtained by using only the six nucleotides of A that are involved in a basepair. $D^A^*$ is given by

\[
D^A^* = \begin{bmatrix}
U30 & A32 & A33 & U551 & U552 & A553 \\
U30 & 0.000 & 4.885 & 8.322 & 9.853 & 8.203 & 5.706 \\
A32 & 4.885 & 0.000 & 3.722 & 7.325 & 5.681 & 4.955 \\
A33 & 8.322 & 3.722 & 0.000 & 5.634 & 5.375 & 7.081 \\
U551 & 9.853 & 7.325 & 5.634 & 0.000 & 3.901 & 7.553 \\
U552 & 8.203 & 5.681 & 5.375 & 3.901 & 0.000 & 4.085 \\
A553 & 5.706 & 4.955 & 7.081 & 7.553 & 4.085 & 0.000
\end{bmatrix}
\]
We compare $D^A$ and $D^B$ to assess the quality of this alignment. $S_{A^*,B}$ is given by

\[
S_{A^*,B} = \begin{bmatrix}
0.000 & 0.367 & 0.197 & 0.318 & 0.157 & 0.093 \\
0.367 & 0.000 & 0.140 & 0.341 & 0.006 & 0.043 \\
0.197 & 0.140 & 0.000 & 0.015 & 0.208 & 0.206 \\
0.318 & 0.341 & 0.015 & 0.000 & 0.132 & 0.153 \\
0.157 & 0.006 & 0.208 & 0.132 & 0.000 & 0.209 \\
0.093 & 0.043 & 0.206 & 0.153 & 0.209 & 0.000
\end{bmatrix}
\]

We observe that the maximum entry of $S$ is 0.3348; that is, $d(A^*, B) = 0.3348$. Since this number is fairly small, it is an indication that the aligned structures are similar. Many of the other entries are very small, which affirms that the structures are similar in shape.

Ultimately, our main purpose for working with distance matrices is not to use them to evaluate alignments, but to use them to actually determine the alignments. We note that when we have two distance matrices of the same size, a valid alignment can always be formed by aligning the nucleotides in the order they appear in the distance matrices, from left to right. Keeping this in mind, given the two distance matrices of the structures to be aligned, any alignment of the two structures corresponds to finding submatrices of the two distance matrices that are the same size. Different alignments correspond to different nucleotides of the original structures not being included in the alignment which corresponds to different nucleotides being removed from the distance matrices. Deletion of nucleotide $i$ from the alignment corresponds to the deletion of both row $i$ and column $i$ from the distance matrix. Thus, the submatrices can only be formed by removing row $i$ whenever column $i$ is removed. By the well-ordered criterion, permutations of rows and columns are not allowed.

So instead of checking whether the distance matrices of the aligned nucleotides satisfy a given property, we find submatrices that satisfy the property and then infer the alignment. A first inclination when considering how to find the best alignment using distance matrices may be to find the submatrices that have a minimum distance between them. But this
would simply result in trivial alignments in which no nucleotides are aligned. An alternative approach is to define a cutoff discrepancy parameter $\varepsilon$ and then find the submatrices of largest size such that the difference between them is less than $\varepsilon$. This corresponds to finding the alignment that includes the maximum number of nucleotides such that the $S$ matrix of absolute differences contains entries that are all less than $\varepsilon$.

For example, suppose we want to align two structures, $A$ and $B$, whose distance matrices are:

$$D^A = \begin{bmatrix}
0 & 4 & 8 & 10 & 5 \\
4 & 0 & 4 & 8 & 5 \\
8 & 4 & 0 & 6 & 7 \\
10 & 8 & 6 & 0 & 9 \\
5 & 5 & 7 & 9 & 0
\end{bmatrix}$$

and

$$D^B = \begin{bmatrix}
0 & 8 & 10 & 5 & 5 \\
8 & 0 & 6 & 8 & 7 \\
10 & 6 & 0 & 4 & 9 \\
5 & 8 & 4 & 0 & 4 \\
5 & 7 & 9 & 4 & 0
\end{bmatrix}$$

The $S$ matrix for $D^A$ and $D^B$ is

$$S_{A,B} = \begin{bmatrix}
0 & 4 & 2 & 5 & 0 \\
4 & 0 & 2 & 0 & 2 \\
2 & 2 & 0 & 2 & 2 \\
5 & 0 & 2 & 0 & 5 \\
0 & 2 & 2 & 5 & 0
\end{bmatrix}$$
With such large entries of $S$ is clear that the two structures do not align well. However, if the 2nd row and column is deleted from $D^A$, we obtain the following submatrix:

$$D^A_{\{1,3,4,5\}} = \begin{bmatrix}
0 & 8 & 10 & 5 \\
8 & 0 & 6 & 7 \\
10 & 6 & 0 & 9 \\
5 & 7 & 9 & 0
\end{bmatrix}$$

New notation was introduced above which will be used below as well. $D^A_{\{1,3,4,5\}}$ is the submatrix of $D^A$ formed by keeping the 1st, 3rd, 4th, and 5th rows and columns of $D^A$. If the 4th row and column is deleted from $D^B$, then we have the submatrix:

$$D^B_{\{1,2,3,5\}} = \begin{bmatrix}
0 & 8 & 10 & 5 \\
8 & 0 & 6 & 7 \\
10 & 6 & 0 & 9 \\
5 & 7 & 9 & 0
\end{bmatrix}$$

So $D^A$ and $D^B$ have identical $4 \times 4$ submatrices, $D^A_{\{1,3,4,5\}}$ and $D^B_{\{1,2,3,5\}}$, respectively. Since they are identical, each absolute difference is zero and hence $d(D^A_{\{1,3,4,5\}}, D^B_{\{1,2,3,5\}}) < \varepsilon$ for all $\varepsilon > 0$. This implies that a good alignment of the original structures is the alignment of the nucleotides 1, 3, 4, and 5 of structure $A$ and 1, 2, 3, and 5 of structure $B$, respectively.

Given two large, square, symmetrical matrices $A$ and $B$, how do we find the largest submatrices that have the desired property? As the size of the distance matrix increases, the number of possible submatrices increases exponentially. Is it possible to solve this problem within a reasonable amount of time in all cases? Knowing whether an efficient algorithm exists for solving the problem is important as it directly relates to the efficiency of finding the best alignment using the distance matrix approach.

It will shortly be proven that the problem at hand is in the computational complexity class, NP-complete. Before the proof is given, a short discussion on the theory of NP-
completeness is provided.

2.3 Theory of NP-Completeness

Complexity theory is related to the scalability of problems and consequently any algorithms that could be used to solve them. Different complexity classes have been created for the categorizing of computational problems and algorithms.

Two of the important time complexity classes that exist for decision problems are P and NP. A decision problem is one with two possible solutions, either “yes” or “no”. Time complexity is determined by the number of steps it takes to solve the problem. Time complexity is typically given as a function of the size of the input. The class P consists of all those decision problems that can be deterministically solved using a polynomial amount of time. That is, there exists \( k \) such that the problem can be solved in less than \( n^k \) steps, where \( n \) is the size of the input. In this case, we say the complexity class is \( O(n^k) \). A decision problem is in the class NP if a proposed solution to the problem can be verified in a polynomial amount of time. Although a proposed solution can be verified in polynomial time, a solution may or may not be able to be determined in polynomial time. Note that any problem in P is in NP but the converse is not necessarily true.

Some problems in NP are in a special class of problems that are specified as being NP-complete.

**Definition 2.3.1.** A decision problem \( C \) is regarded as being **NP-complete** if it satisfies the following two properties [23]:

1. Any proposed solution to \( C \) can be verified in polynomial time (\( C \) is in NP).

2. For every other problem \( C' \) that is in NP, there is a polynomial time transformation from \( C' \) to \( C \).

The second item essentially states that if \( C \) is NP-complete, then it is at least as hard as every other problem in NP. Although solutions to NP-complete problems can be verified
in a polynomial time, it is not known whether or not they can be solved deterministically in polynomial time (whether or not they are in P). By the properties of the NP-complete class, if one NP-complete problem is in P, then all problems in NP are in P (P=NP). This is apparent since if one problem has a polynomial solution and all other problems can be transformed into that problem in polynomial time, then all problems have a polynomial solution. But if one NP-complete problem is shown to not be in P, then no problems in the NP-complete class are in P (P≠NP).

Whether or not P=NP is an unsolved problem; in fact, a million dollar prize has been offered for its solution [13]. Although it is an unsolved problem, it is generally believed to not be true because of the amount of research that has been put into it. Thus, problems that are NP-complete are most often considered to be the hardest problems in NP and to not have a solution that can be determined in polynomial time.

To show a problem is NP-complete, two things must be shown:

1. The problem is in NP.

2. There exists another problem that is known to be NP-complete that can be reduced (transformed) to it by a function that is bounded in time by a polynomial function.

What about problems that are not decision problems? For example, how does one prove that an optimization problem is NP-complete? Consider the traveling salesman problem, which is an optimization problem. In this classic optimization problem, one is given a collection of cities and the distance between each pair of them, and the task is to find the shortest path to take in order to visit all of the cities and then return to the starting point. The decision problem formulation of this problem is given as, “Does a route exist in which the total distance traveled is less than $k$?” Proving a decision problem to be NP-complete necessarily proves its corresponding optimization problem to be NP-complete as well since the decision problem can be no harder than the corresponding optimization problem. For if we could find the shortest route in polynomial time, then we could solve the decision
problem in polynomial time by simply finding the shortest route and then comparing it with the given bound \( k \).

We again use the following notation: \( A_{\{a_1, a_2, \ldots, a_m\}} \) is the submatrix of \( A \) formed by the intersection of rows \( a_1, a_2, \ldots, a_m \) and columns \( a_1, a_2, \ldots, a_m \). We assume \( a_1 < a_2 < \ldots < a_m \).

The problem we want to solve is the following: Given two square matrices, \( A \) and \( B \), find the largest submatrices \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \) such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon \) for all \( i, j \).

We will refer to this problem as the \( \varepsilon \)-Submatrix Matching Problem (\( \varepsilon \)-SMP). The standard Submatrix Matching Problem (SMP) is very similar: Given two square matrices, \( A \) and \( B \), find the largest submatrices \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \) such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0 \) for all \( i, j \). As explained above, only decision problems are NP-complete, so we formulate the SMP and as decision problems.

**Definition 2.3.2.** Given two square matrices, \( A \) and \( B \) and positive integer \( K \), the **Submatrix Matching Problem (SMP)** asks the following question: Do there exist submatrices \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \), where \( m \geq K \), such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0 \) for all \( (i, j) \)?

**Definition 2.3.3.** Given two square matrices, \( A \) and \( B \) and positive integer \( K \), the **\( \varepsilon \)-Submatrix Matching Problem (\( \varepsilon \)-SMP)** asks the following question: Do there exist submatrices \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \), where \( m \geq K \), such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon \) for all \( (i, j) \)?

### 2.4 SMP is NP-complete

**Theorem 2.4.1.** SMP is NP-complete.

**Proof.** SMP is in NP since given two \( m \times m \) matrices, \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \), it can be determined whether \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0 \) for all \( (i, j) \) in \( O(m^2) \) time.
Next we must show that a known NP-complete problem can be transformed into the SMP in polynomial time. We use the CLIQUE problem, one of the original 21 problems shown to be NP-complete [23]. The CLIQUE problem is the following: given a graph $G = (V, E)$ and positive integer $J \leq |V|$, does $G$ contain a clique of size $J$ or more, that is a subset $V' \subseteq V$ such that $|V'| \geq J$ and every two vertices in $V'$ are joined by an edge in $E$? The maximum clique problem is discussed further in Section 3.2.4.

To show that CLIQUE can be transformed into SMP, it must be shown that given a graph $G = (V, E)$ and positive integer $J \leq |V|$, we can construct two matrices $A$ and $B$ and find a positive integer $K$ in polynomial time such that there exists $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, where $m \geq K$, such that $|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0$ for all $(i, j)$ if and only if $G$ contains a clique of size greater than or equal to $J$.

Let $K = J$, let $n = |V|$, and let $A$ and $B$ be $n \times n$ matrices where $A(i, j) = n(i-1) + j$ and $B(i, j) = n^2 + n(i-1) + j$. At this point every entry of each matrix is distinct and all entries of matrix $A$ and matrix $B$ are distinct. We now modify $A$ and $B$ by setting $A(i, i) = 0$ and $B(i, i) = 0$ for $i = 1 \ldots n$. Then for each $i, j$ such that $\{v_i, v_j\} \in E$, we set $B(i, j) = A(i, j)$ and $B(j, i) = A(j, i)$.

First, we suppose that $G$ contains a clique of size greater than or equal to $J$. Then $V$ has a subset $\{v_{i_1}, v_{i_2}, \ldots, v_{i_r}\}$ such that $i_1 < i_2 < \ldots < i_J$ and $\{v_{i_r}, v_{i_s}\} \in E$ for all $r, s \in \{1, 2, \ldots, J\}, r \neq s$. We now show that $|A_{\{i_1, i_2, \ldots, i_J\}}(p, q) - B_{\{i_1, i_2, \ldots, i_J\}}(p, q)| = 0$ for all $p, q \in \{1, 2, \ldots, J\}, p \neq q$. Since for every $p, q$, $\{v_{i_p}, v_{i_q}\} \in E$, it is true that $A(i_p, i_q) = B(i_p, i_q)$. Also, $B_{\{i_1, i_2, \ldots, i_J\}}(p, q) = B(i_p, i_q)$ and $A_{\{i_1, i_2, \ldots, i_J\}}(p, q) = A(i_p, i_q)$, so we have $|A_{\{i_1, i_2, \ldots, i_J\}}(p, q) - B_{\{i_1, i_2, \ldots, i_J\}}(p, q)| = 0$ for all $p, q \in \{1, 2, \ldots, J\}, p \neq q$. Since all entries on the diagonal matrices are 0, the result holds when $p = q$.

Conversely, suppose there exists $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, where $m \geq K$, such that $|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0$ for all $(i, j)$. If $A_{\{a_1, a_2, \ldots, a_m\}}(i, j) = B_{\{b_1, b_2, \ldots, b_m\}}(i, j)$ for all $i, j$, then $A(a_i, a_j) = B(b_i, b_j)$ for all $i, j$. Since $A(a_i, a_j) = B(b_i, b_j)$ if and only if $\{v_{a_i}, v_{a_j}\} \in E$, this implies that $\{v_{a_i}, v_{a_j}\} \in E$ for all $i, j \in \{1, 2, \ldots, m\}$. Thus,
\( \{v_{a_1}, v_{a_2}, \ldots, v_{a_m} \} \) is a clique of size \( m \geq K = J \).

2.5 \( \varepsilon \)-SMP is NP-complete

Theorem 2.5.1. \( \varepsilon \)-SMP is NP-complete

Proof. As with SMP, it can easily be verified that \( \varepsilon \)-SMP is in NP. It remains to be shown that there exists a problem already known to be NP-complete that can be transformed into it in polynomial time. As above, we could use the CLIQUE problem, although a different construction of \( A \) and \( B \) is required.

However, because we have shown that SMP is NP-complete, we can now use it to prove that \( \varepsilon \)-SMP is NP-complete. We must transform SMP into an instance of the \( \varepsilon \)-SMP in polynomial time in such a way that \( \varepsilon \)-SMP can be used to provide a solution for the SMP.

Suppose we have arbitrary (square) matrices \( A \) and \( B \) and positive integer \( K \) and we want to know if there exists \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \), where \( m \geq K \), such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0 \) for all \((i, j)\). We use the same \( A, B, \) and \( K \) for \( \varepsilon \)-SMP.

Suppose \( A \) is \( r \times r \) and \( B \) is \( s \times s \). Then there are \( r^2 \times s^2 \) pairwise absolute differences between the entries of \( A \) and \( B \). We let \( \varepsilon \) be equal to the minimum positive difference divided by two. That is,

\[
2\varepsilon = \min_{i,j,p,q} \{|A(i, j) - B(p, q)| : |A(i, j) - B(p, q)| > 0\}.
\]

It is easy to see that for the given \( \varepsilon \) (and in fact, any \( \varepsilon > 0 \)), if the \( \varepsilon \)-SMP does not have a solution, then SMP does not have a solution.

Now suppose \( \varepsilon \)-SMP does have a solution. That is, there exists \( A_{\{a_1, a_2, \ldots, a_m\}} \) and \( B_{\{b_1, b_2, \ldots, b_m\}} \), with \( m \geq K \) such that \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon \) for all \( i, j \). But since \( \varepsilon \) is less than the smallest of the positive absolute differences between entries, \( |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon \) for all \( i, j \).
\[ B_{\{b_1, b_2, \ldots, b_m\}}(i, j) < \varepsilon \] implies \[ |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0. \] Thus, SMP has a solution as well, which proves that \( \varepsilon \)-SMP is NP-complete.

2.6 \( \varepsilon \)-SMP implementation

Despite being NP-complete, a program was written to solve \( \varepsilon \)-SMP. Test cases found that in addition to being computationally expensive, this approach does not produce favorable alignments because the requirement that every absolute difference must be within \( \varepsilon \) is too strict and consequently not many nucleotides will be included in the alignment because it is unlikely to find two large structural regions fulfilling the strict requirement. For example, a nucleotide will be excluded from the alignment if just one of its inter-nucleotide distances differs from the corresponding inter-nucleotide distance by more than the predefined value. This can easily happen with some of the distances between nucleotides that are far apart. For example, if \( \varepsilon = 1 \text{Å} \), then two nucleotides that are 60Å apart cannot be aligned with two nucleotides that are 62Å apart even though the distances are relatively close. However, if \( \varepsilon \) is simply increased, then it will be too large to discriminate on a local level.

2.7 Relative Deviations

One way to remedy this is to impose a requirement based on relative deviations instead of absolute. That is, rather than the requirement \[ |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon \] for all \( i, j \), we have \[ \frac{|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)|}{AB^*(i, j)} \leq \varepsilon \] for all \( i, j \) where \( AB^*(i, j) \) is the average of \( A_{\{a_1, a_2, \ldots, a_m\}}(i, j) \) and \( B_{\{b_1, b_2, \ldots, b_m\}}(i, j) \). Suppose here that \( \varepsilon = .10 \). Then if the average of the two corresponding inter-distances in \( A \) and \( B \) is 10Å, their difference should differ by no more than 1Å. But a deviation of 6Å is acceptable if the average of the two inter-distances is 60Å.

Definition 2.7.1. Given two square nonnegative matrices, \( A \) and \( B \) and positive integer \( K \),
the Relative Deviations-Submatrix Matching Problem (RD-SMP) asks the following question: Do there exist submatrices $A_{\{a_1, a_2, ..., a_m\}}$ and $B_{\{b_1, b_2, ..., b_m\}}$, where $m \geq K$, such that $|A_{\{a_1, a_2, ..., a_m\}}(i,j) - B_{\{b_1, b_2, ..., b_m\}}(i,j)|/AB^*(i,j) < \varepsilon$ for all $i, j$ where $AB^*(i,j)$ is the average of $A_{\{a_1, a_2, ..., a_m\}}(i,j)$ and $B_{\{b_1, b_2, ..., b_m\}}(i,j)$? If $A_{\{a_1, a_2, ..., a_m\}}(i,j) = 0$ and $B_{\{b_1, b_2, ..., b_m\}}(i,j) = 0$, then $AB^*(i,j)$ is defined to equal 1.

Although this approach, may produce more favorable alignments, it is still NP-complete, as shown in the next section.

### 2.8 RD-SMP is NP-Complete

**Theorem 2.8.1.** RD-SMP is NP-complete

*Proof.* It is again easy to verify that RD-SMP is in NP. Here we again use the fact that SMP is NP-complete to prove that RD-SMP is NP-complete. We must transform SMP into an instance of RD-SMP in polynomial time in such a way that RD-SMP can be used to provide a solution for SMP.

Suppose we have arbitrary (square) matrices $A$ and $B$, positive integer $K$, and we want to know if there exists $A_{\{a_1, a_2, ..., a_m\}}$ and $B_{\{b_1, b_2, ..., b_m\}}$, where $m \geq K$, such that $|A_{\{a_1, a_2, ..., a_m\}}(i,j) - B_{\{b_1, b_2, ..., b_m\}}(i,j)| = 0$ for all $(i,j)$. We use the same $A, B, \text{ and } K$ for RD-SMP.

Suppose $A$ is $r \times r$ and $B$ is $s \times s$. Then there are $r^2 \times s^2$ pairwise averages between the entries of $A$ and $B$. We let $\delta$ be equal to the maximum average between any two entries. That is,

$$\delta = \max_{i,j,p,q} \left\{ \frac{A(i,j) - B(p,q)}{2} \right\}. \quad (2.1)$$

Then we let $\varepsilon^*$ be equal to minimum positive difference of the entries divided by $2\delta$. That is,

$$\varepsilon^* = \min_{i,j,p,q} \left\{ \frac{|A(i,j) - B(p,q)|}{2\delta} : |A(i,j) - B(p,q)| > 0 \right\}. \quad (2.2)$$

Now it remains to be shown that the solution to RD-SMP using $\varepsilon^*$ implies the solution
to SMP. Suppose the answer to RD-SMP is ‘no’. That is, there do not exist $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, where $m \geq K$, such that $|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| / AB^*(i, j) < \varepsilon^*$ for all $(i, j)$. Then there do not exist $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, such that $|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon^* \times AB^*(i, j)$ for all $(i, j)$. Since $\varepsilon^* \times AB^*(i, j) \geq 0$ for all $(i, j)$, it follows that there do not exist $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, such that $|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0$ for all $(i, j)$ (and hence the answer to SMP is ‘no’ as well).

Now we suppose RD-SMP does have solution (the answer to RD-SMP is ‘yes’). That is, there do exist $A_{\{a_1, a_2, \ldots, a_m\}}$ and $B_{\{b_1, b_2, \ldots, b_m\}}$, where $m \geq K$, such that for all $(i, j)$, we have

$$|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| / AB^*(i, j) < \varepsilon^*.$$ 

Equivalently, we have for all $(i, j)$,

$$|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon^* \times AB^*(i, j)$$

By Equation 2.1

$$|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| < \varepsilon^* \delta$$

However, by the definition of $\varepsilon^*$ in Equation 2.2, we have that

$$\varepsilon^* \delta < |A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)|$$

for all $(i, j)$.

unless

$$|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0.$$

Thus,

$$|A_{\{a_1, a_2, \ldots, a_m\}}(i, j) - B_{\{b_1, b_2, \ldots, b_m\}}(i, j)| = 0$$

for all $(i, j)$

(and hence the answer to SMP is ‘yes’ as well).
2.9 Modified Distance Matrices

Another way to reduce the effect of pairs of nucleotides that are in the long distance range is to use modified distance matrices in which any distance greater than some threshold is set to 0. This way only the more discriminative distances are included and only short range interactions are considered. This will also allow for some flexibility in the overall shape of the two structures being aligned.

2.10 Distance Matrix Overlap

Yet another approach that will produce alignments with more aligned nucleotides is one that does not require all deviations to be less than $\varepsilon$, but instead maximizes the number of deviations that are less than $\varepsilon$. This will be referred to as the distance matrix overlap problem. A simplified version of this problem is the contact map overlap problem which has been used in the alignment of 3D protein structures \[25\]. In the contact map overlap problem, the 3D information is stored as a simplified 0-1 distance matrix (a contact map) in which the entry for two residues is 1 if they are close (“in contact”). A contact map overlap is defined as the number of contacts that overlap when two maps of the same size are placed on top of one another. The highest quality alignment according to the contact map overlap is then the alignment of bases which maximizes the number of overlapping contacts. This is equivalent to solving the distance matrix overlap problem using the simplified binary distance matrices and setting $\varepsilon = 0$. However, finding the optimal contact map overlap has been shown to be NP-complete \[27\]. The contact map overlap problem is explained in further detail in Section 3.3.
CHAPTER 3

A New Approach to 3D-to-3D Alignment

3.1 Goals of the New Approach

The most common approach for constructing a 3D-to-3D alignment of RNA, used by DIAL, SARA, and PARTS, is to first extract particular classes of information from the 3D structure, create a reduced, essentially linear, representation of the RNA structure, and then utilize a quadratic-time sequence alignment algorithm on the reduced structure. However, to achieve quadratic time, the secondary and tertiary structure of the RNA are ignored in order to reduce the structure to a linear representation, which often results in a decrease in the accuracy of the alignment. Given the complexity of RNA 3D structure, it is not surprising that quadratic time algorithms, although computationally convenient, are not as accurate as desired [20]. RNA 3D structures involve complex and long range interactions between nucleotides not adjacent in the primary sequence. We will develop a new alignment method that incorporates the long range interactions characteristic of RNA.

The other 3D-to-3D program, ARTS, is not designed to align large, conformationally flexible RNA 3D structures because it uses a single rigid transformation to find corresponding
nucleotides. Typically, ARTS leaves entire regions out of the alignment, even though the regions themselves may superpose very well within some local neighborhood. The distance matrix methods SMP and $\varepsilon$-SMP, considered in Chapter 2 are also in this category. It is therefore desirable to develop a method that focuses on local conformations rather than just the global shape, which is the approach of ARTS and some of the distance matrix approaches discussed in Chapter 2. The new alignment method should align nucleotides if there is some local neighborhood in which they correspond well structurally.

Since we want an alignment in which the correspondences imply that local geometries correspond instead of solely a correspondence between the overall global shape of the set of aligned nucleotides, we will focus on superpositions of local neighborhoods. That is, for any local neighborhood of aligned nucleotides in one structure, we should be able to infer structural similarity with the aligned nucleotides in the other structure. We can evaluate an alignment based on the number of local aligned neighborhoods that superimpose well. The goal of our approach is to find the alignment that maximizes this number; that is, we want the alignment that contains more local neighborhoods that are structurally similar than any other possible alignment.

Our approach focuses on the superpositions of local neighborhoods. We will form the local neighborhoods in such a way that will take into account pairwise nucleotide interactions of all types, including non-Watson-Crick as well as Watson-Crick base pairs. In addition to local basepair interactions, the approach will also consider also considers long-range tertiary base-pair and base-stacking contacts. Although we account for and use all local and long-range pairwise interaction information, we maintain independence from any particular program for annotating base interactions, such as FR3D [59], RNAView [71], or MC-ANNOTATE [24], in order to avoid the problem of having interactions fall just outside the limits of the classification region. In other words, we work directly with the geometry of the 3D structures, rather than reduce the structure to a network of symbolic interactions.
3.2 Local Alignments and the Local Alignment Graph

Our objective is to produce the global alignment of two homologous or structurally similar RNA 3D structures in which as many smaller local neighborhoods as possible superimpose well in 3D space. Since we are focusing solely on the structure data to make the alignment, a corollary is that nucleotides that do not superpose well are not aligned even though they may be homologous at the sequential level. For example, a site with a G in each structure may not be aligned because the two nucleotides do not correspond structurally.

3.2.1 Form Local Neighborhoods

The first step is to identify the local neighborhoods in each structure. Suppose that structure \( A \) is made up of nucleotides \( 1^A, 2^A, \ldots, n^A \), in 5’ to 3’ order. A list \( N^A \) is constructed that contains all neighborhoods consisting of four nucleotides each, such that the maximum Euclidean distance between any two of the four nucleotides in a neighborhood is less than a pre-specified distance threshold, \( D \). Inter-nucleotide distances are measured from base center to base center as previously described [59].

An example will be used to illustrate our method. In our example, we choose nucleotides representing a part of internal loop B of 5S rRNA. Let structure \( A \) be nucleotides 22-27 and 60-61 in \( T.\ thermophilus \) 5S rRNA (2j01 crystal structure [62]) and structure \( B \) be nucleotides 22-27 and 60-61 in \( E.\ Coli \) 5S rRNA (2aw4 crystal structure [60]). The 3D structures are displayed in Figure 3.1.

We use \( D = 12 \text{Å} \) as the default value to calculate \( N^A \), a value that groups together in the same neighborhood nucleotides that form any possible pairwise interaction (e.g. base-pairing, base-stacking, base-backbone) whether local or long-range with respect to the sequence or secondary structure.

To get a better idea about the types of neighborhoods that are formed we will look at the neighborhoods of structure \( A \) that include nucleotide 22\(^A\). We will illustrate the neigh-
Figure 3.1: Part of internal loop B of two 5S rRNA molecules.

Figure 3.1: Part of internal loop B of two 5S rRNA molecules.

(a) *T. thermophilus*, crystal structure 2j01
(b) *E. coli*, crystal structure 2aw4

When $D = 12\text{Å}$, $22^A$ is included in four different neighborhoods, which are illustrated in Figure 3.3. From the figure it is clear that the neighborhoods include bases that are making basepairing interactions. It should also be noted that the different combinations of nucleotides in the neighborhoods allow the program to distinguish single nucleotides that superimpose poorly.

Using a distance threshold of $12\text{Å}$, all the neighborhoods of structure $A$, $N^A$, are calculated. $N^A$ consists of 17 neighborhoods, which are listed in Figure 3.4. The four nucleotides in each neighborhood are listed in ascending order starting from the 5’-end. That is, for any neighborhood of $A$, $(i^A, j^A, k^A, l^A)$, it is true that $i < j < k < l$. Each nucleotide can and most likely will be contained in multiple neighborhoods, for if a sphere of diameter $D$ drawn about nucleotide $i^A$ contains $k$ other nucleotides, then $i^A$ will be included in at least
Figure 3.2: Secondary structure diagram for the *E.c.* region consisting of nucleotides 22-27 and 60-61. Nucleotide 22 is circled in the diagram.

Figure 3.3: The four neighborhoods in $N^4$ that include nucleotide $22^4$ that are formed when $D = 12\text{Å}$. 
two of the four nucleotides is $\binom{4}{k}$. Because the number of nucleotides within any radius is restricted due to the volume occupied by each nucleotide, the number of neighbors for any given nucleotide is limited. The total number of neighborhoods of a molecule of size $n$ is then of complexity $O(n)$.

### 3.2.2 Measuring Structural Similarity

Given another structure $B$ that is to be aligned with structure $A$, we construct its list of four-nucleotide neighborhoods, $N^B$. In our example, $N^B$ consists of 11 neighborhoods, which are listed in Figure 3.4. We then compare the neighborhoods of $N^A$ and $N^B$ to determine which neighborhoods are structurally similar.

$$
\begin{align*}
    n^A(1) &= (22^a, 23^a, 24^a, 60^a) \\
    n^A(2) &= (22^a, 23^a, 24^b, 61^a) \\
    n^A(3) &= (22^a, 23^a, 60^a, 61^a) \\
    n^A(4) &= (22^a, 24^a, 60^a, 61^a) \\
    n^A(5) &= (23^a, 24^a, 25^a, 26^a) \\
    n^A(6) &= (23^a, 24^a, 25^a, 27^a) \\
    n^A(7) &= (23^a, 24^a, 26^a, 27^a) \\
    n^A(8) &= (23^a, 24^a, 26^a, 61^a) \\
    n^A(9) &= (23^a, 24^a, 27^a, 60^a) \\
    n^A(10) &= (23^a, 24^a, 27^a, 61^a) \\
    n^A(11) &= (23^a, 24^a, 60^a, 61^a) \\
    n^A(12) &= (23^a, 25^a, 26^a, 27^a) \\
    n^A(13) &= (23^a, 26^a, 27^a, 61^a) \\
    n^A(14) &= (23^a, 27^a, 60^a, 61^a) \\
    n^A(15) &= (24^a, 25^a, 26^a, 27^a) \\
    n^A(16) &= (24^a, 26^a, 27^a, 61^a) \\
    n^A(17) &= (24^a, 27^a, 60^a, 61^a)
\end{align*}
$$

Figure 3.4: All 4-nucleotide neighborhoods such that the maximum distance between any two of the four nucleotides is $D = 12\text{Å}$.

We find that the geometric discrepancy (measured in Ångstroms) developed in [59] is suitable for determining the structural similarity between 4-nucleotide neighborhoods. The geometric discrepancy is an entirely geometric measure that takes into account the general
shape of the two structures being compared as well as the relative orientations of the bases. It is similar to the average root-mean-square deviations (RMSD) measure, but accounts for base substitutions between the two structures in a consistent way.

For each local neighborhood in $N^A$ we calculate the geometric discrepancy with every local neighborhood in $N^B$. Two neighborhoods are classified as being \textit{structurally similar} if their geometric discrepancy is less than the pre-defined discrepancy cutoff parameter $d$.

### 3.2.3 Alignment Objectives

Having determined those local neighborhoods that are structurally similar, we want to find the alignment that maximizes the number of these well superimposed 4-nucleotide neighborhoods contained within the alignment. However, like many formulations of the 3D structural alignment problem, this problem is NP-complete. The proof is provided in Section 3.3.

While the problem is NP-complete, an efficient solution is still worth pursuing. This is not only because of the practical importance of the problem, but also since the size of RNA molecules are not arbitrarily large. If we can develop an algorithm that will efficiently align the largest crystallized RNA structures, such as 16S and 23S rRNA, we can be assured it can be used in many other situations as well.

Using the number of corresponding structurally similar 4-nucleotide neighborhoods as the measure of the quality of the alignment, it will be shown that finding the optimal alignment can be reduced to finding the maximum clique of a specially defined graph, which will be referred to as the \textquotedblleft local alignment graph\textquotedblright. First, a brief review of the terminology of graph theory and the maximum clique problem will be given.

### 3.2.4 Graph Theory Terminology

A undirected graph $G = (V, E)$ is an ordered pair with $V$ being a finite set of vertices and $E$ being the set of edges between the vertices. Each element of $E$ is an unordered pair $(u, v)$ of distinct vertices. If two vertices have an edge connecting them, they are said to be adjacent.
\( \delta(v) \) is the set of vertices adjacent to vertex \( v \), also known as the neighborhood of \( v \). Adjacent vertices of \( v \) will sometimes be referred to as neighbors of \( v \). A subgraph of a graph \( G \) is a graph \( G' = (V', E') \) whose vertex set \( V' \) is a subset of \( V \) and whose adjacency relation \( E' \) is the subset of \( E \) restricted to this \( V' \). A subset of vertices of a graph is said to be a clique if each pair of vertices in the subset are connected by an edge. A clique is said to be maximal if its vertices cannot be contained in a larger clique. A maximum clique of a graph is a clique with the property that no other clique contains more vertices.

### 3.2.5 Vertices of the Local Alignment Graph

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n^b(1) )</td>
<td>0.3</td>
<td>0.7</td>
<td>1.2</td>
<td>1.8</td>
<td>1.5</td>
<td>1.3</td>
<td>1.5</td>
<td>1.7</td>
<td>2.4</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>( n^a(2) )</td>
<td>0.7</td>
<td>0.2</td>
<td>1.2</td>
<td>1.9</td>
<td>1.7</td>
<td>1.2</td>
<td>1.2</td>
<td>1.8</td>
<td>2.6</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>( n^b(3) )</td>
<td>1.2</td>
<td>1.1</td>
<td>0.3</td>
<td>1.1</td>
<td>1.8</td>
<td>1.9</td>
<td>1.8</td>
<td>1.1</td>
<td>1.9</td>
<td>2.2</td>
<td>1.7</td>
</tr>
<tr>
<td>( n^a(4) )</td>
<td>1.7</td>
<td>1.7</td>
<td>0.9</td>
<td>0.2</td>
<td>2.4</td>
<td>2.2</td>
<td>2.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.9</td>
<td>1.5</td>
</tr>
<tr>
<td>( n^b(5) )</td>
<td>1.7</td>
<td>1.8</td>
<td>1.9</td>
<td>2.4</td>
<td>0.4</td>
<td>2.2</td>
<td>2.3</td>
<td>2.2</td>
<td>2.7</td>
<td>1.6</td>
<td>2.3</td>
</tr>
<tr>
<td>( n^a(6) )</td>
<td>1.3</td>
<td>1.5</td>
<td>1.9</td>
<td>2.6</td>
<td>0.7</td>
<td>1.7</td>
<td>2.0</td>
<td>2.2</td>
<td>3.0</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>( n^b(7) )</td>
<td>1.5</td>
<td>1.7</td>
<td>1.8</td>
<td>2.3</td>
<td>0.3</td>
<td>2.1</td>
<td>2.2</td>
<td>2.0</td>
<td>2.7</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>( n^a(8) )</td>
<td>1.5</td>
<td>1.2</td>
<td>2.0</td>
<td>2.2</td>
<td>2.2</td>
<td>0.7</td>
<td>0.4</td>
<td>1.9</td>
<td>2.5</td>
<td>2.2</td>
<td>2.6</td>
</tr>
<tr>
<td>( n^b(9) )</td>
<td>1.1</td>
<td>1.1</td>
<td>1.9</td>
<td>2.3</td>
<td>2.0</td>
<td>0.2</td>
<td>0.6</td>
<td>1.9</td>
<td>2.6</td>
<td>2.2</td>
<td>2.6</td>
</tr>
<tr>
<td>( n^a(10) )</td>
<td>1.3</td>
<td>1.1</td>
<td>1.8</td>
<td>2.1</td>
<td>2.2</td>
<td>0.6</td>
<td>0.1</td>
<td>1.8</td>
<td>2.4</td>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td>( n^b(11) )</td>
<td>1.6</td>
<td>1.7</td>
<td>1.0</td>
<td>1.1</td>
<td>2.1</td>
<td>1.9</td>
<td>1.8</td>
<td>0.1</td>
<td>1.1</td>
<td>2.0</td>
<td>1.2</td>
</tr>
<tr>
<td>( n^a(12) )</td>
<td>1.7</td>
<td>1.9</td>
<td>1.7</td>
<td>1.5</td>
<td>1.9</td>
<td>2.2</td>
<td>2.3</td>
<td>2.1</td>
<td>2.2</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>( n^b(13) )</td>
<td>1.7</td>
<td>1.6</td>
<td>2.0</td>
<td>1.7</td>
<td>3.0</td>
<td>1.5</td>
<td>1.4</td>
<td>2.0</td>
<td>2.1</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>( n^a(14) )</td>
<td>2.3</td>
<td>2.4</td>
<td>1.6</td>
<td>1.0</td>
<td>2.7</td>
<td>2.6</td>
<td>2.4</td>
<td>1.0</td>
<td>0.2</td>
<td>1.9</td>
<td>1.0</td>
</tr>
<tr>
<td>( n^b(15) )</td>
<td>1.6</td>
<td>1.7</td>
<td>1.6</td>
<td>1.4</td>
<td>2.0</td>
<td>1.9</td>
<td>2.0</td>
<td>1.7</td>
<td>1.9</td>
<td>1.6</td>
<td>2.2</td>
</tr>
<tr>
<td>( n^a(16) )</td>
<td>1.6</td>
<td>1.7</td>
<td>2.0</td>
<td>1.8</td>
<td>2.9</td>
<td>1.6</td>
<td>1.5</td>
<td>1.8</td>
<td>2.1</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>( n^b(17) )</td>
<td>2.0</td>
<td>2.3</td>
<td>1.5</td>
<td>1.4</td>
<td>2.2</td>
<td>2.5</td>
<td>2.4</td>
<td>1.0</td>
<td>0.9</td>
<td>2.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Figure 3.5: Pairwise geometric discrepancy calculations for between the neighborhoods of \( A \) and \( B \) given in Figure 3.4.

After the discrepancy values are calculated, they are organized into an \( m \times n \) array, with \( m = |N^A| \) and \( n = |N^B| \) being the number of 4-nucleotide neighborhoods from each
structure, respectively. See Figure 3.5. Vertices are created in positions of the array where the geometric discrepancy value is less than the discrepancy cutoff parameter, \( d \). In Figure 3.6, the discrepancies of Figure 3.5 have been removed and those discrepancies less than 0.5 have been replaced by vertices. We have found 0.5\( \text{Å} \) to be a sufficient value for this parameter, but the value may be lowered (raised) to find an alignment that is more (less) discriminative in determining which nucleotides sets are structurally similar. Lowering the value of \( d \) also has the effect of decreasing the overall runtime of the program since fewer pairs of neighborhoods will be considered to be structurally similar, causing fewer vertices to be created.

<table>
<thead>
<tr>
<th>( n^A(1) )</th>
<th>( n^B(1) )</th>
<th>( n^B(2) )</th>
<th>( n^B(3) )</th>
<th>( n^B(4) )</th>
<th>( n^B(5) )</th>
<th>( n^B(6) )</th>
<th>( n^B(7) )</th>
<th>( n^B(8) )</th>
<th>( n^B(9) )</th>
<th>( n^B(10) )</th>
<th>( n^B(11) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(2) )</td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(3) )</td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(4) )</td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(5) )</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(6) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(7) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(8) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(9) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(10) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(11) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>( n^A(12) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(13) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(14) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>( n^A(15) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(16) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n^A(17) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
</tr>
</tbody>
</table>

Figure 3.6: Vertices of the local alignment graph. Vertices are created in entries where the geometric discrepancy is less than the cutoff parameter \( d \). Here \( d = 0.5\text{Å} \).
Proposition 3.2.1. Each vertex represents a valid (well-ordered and uniquely-assigned) alignment of one neighborhood from A and one neighborhood from B.

Proof. Let $v$ be a vertex of the graph. Then $v$ corresponds to some entry $(q,r)$ of the array and consequently to neighborhood $q$ of $A$, $(i_q^A, j_q^A, k_q^A, l_q^A)$, and neighborhood $r$ of $B$, $(i_r^B, j_r^B, k_r^B, l_r^B)$. It was shown above that the order of the nucleotides is preserved within the neighborhood specification and hence, $i_q^A < j_q^A < k_q^A < l_q^A$ and $i_r^B < j_r^B < k_r^B < l_r^B$. Thus, an alignment of $i_q^A$ with $i_r^B$, $j_q^A$ with $j_r^B$, $k_q^A$ with $k_r^B$, and $l_q^A$ with $l_r^B$ will be a valid alignment where the ordering of the nucleotides is kept intact and no nucleotide is aligned with more than one other nucleotide.

3.2.6 Edges of the Local Alignment Graph

It has been established that each vertex represents a well-ordered and uniquely-assigned alignment of nucleotides in two structurally similar neighborhoods. Each of these alignments has only four correspondences, but we can merge together many such local alignments to form a larger alignment of $A$ and $B$. We want to combine together as many 4-nucleotide alignments as possible, but not all of these 4-nucleotide alignments are compatible; some lead to an alignment which violates the well-ordered and/or the uniquely-assigned criteria. Figure 3.7 gives an example of two 4-nucleotide alignments that are incompatible; the two alignments cannot be merged to form a larger valid alignment. However, Figure 3.8 gives an example of two 4-nucleotide local alignments that can be merged to form a valid 5-nucleotide alignment.

We draw edges between two vertices in the local alignment graph if the 4-nucleotide alignments which they represent are compatible with each other, in the sense that they could be merged together to form a larger valid alignment, as previously described. In Figure 3.9 edges have been drawn between the vertices of the graph given in Figure 3.6.

We note that a group of vertices represent a valid alignment if and only if each pair of vertices are compatible; that is, if and only if each pair of vertices are connected by an edge.
Figure 3.7: An example of two local alignments that cannot be merged to form a valid alignment because the uniquely-assigned criterion is violated.

Figure 3.8: An example of two local alignments that can be merged to form a valid alignment because neither the uniquely-assigned criterion nor the well-ordered criterion is violated.
The goal of selecting as many structurally similar local neighborhoods as possible that can be combined to form a larger alignment then corresponds to selecting the largest set of vertices such that every pair of vertices in the set are connected by an edge. In other words, we want to find the maximum clique of the local alignment graph. This idea was first introduced with 2-nucleotide alignments to solve the contact map overlap problem for proteins in [64].

In the example above, the maximum clique consists of all vertices except (5,5) and (8,7). The ten vertices of the clique combine to give a final alignment of \((22^A, 23^A, 24^A, 26^A, 27^A, 60^A, 61^A)\) and \((22^B, 23^B, 24^B, 26^B, 27^B, 60^B, 61^B)\), illustrated in Figure 3.10. Figure 3.11 displays the superposition of the aligned nucleotides. Nucleotides \(25^A\) and \(25^B\) are also included for illustrative purposes. From the figure, it is clear that nucleotides \(25^A\) and \(25^B\) do not superimpose on each other and so are correctly not aligned to each other.
Figure 3.10: Alignment of the regions produced by the maximum clique procedure

Figure 3.11: Aligned regions are superimposed along with the two unaligned nucleotides. The two nucleotides which are not aligned, circled in this figure, clearly do not superimpose well.
3.3 Maximizing Similar Neighborhoods is NP-complete

Definition 3.3.1. Given two structures $A$ and $B$, the corresponding lists of 4-nucleotide neighborhoods as well as what pairs of neighborhoods (taking one neighborhood from each structure) are structurally similar, and a positive integer $K$, the **Maximizing Similar Neighborhoods (MSN) problem** asks the following question: Does their exist an alignment of the nucleotides of $A$ and $B$ such that the number of aligned structurally similar neighborhoods is greater than or equal to $K$?

To show the MSN problem is NP-complete, we show that the contact map overlap problem (shown to be NP-complete in [27]) can be reduced to it.

A brief description of the contact map overlap is now provided. The contact map overlap was first proposed in [25] as a measure of protein similarity although it can also be applied to RNA structures as well. As RNA folds, nucleotides that are not nearby in the sequence can become close to each other in the tertiary structure. The contact map overlap measures the similarity between the tertiary structures of two proteins by comparing the closeness of nucleotides.

The contact map of an RNA molecule represents the distance between every two nucleotides of the 3D structure using a binary 2D matrix. For two nucleotides $i$ and $j$, the $ij$ element of the matrix is 1 if the two nucleotides are closer than a predetermined threshold, and 0 otherwise. For the purposes of this section, we note that the contact map can also be represented as a graph containing one vertex for each nucleotide. An edge exists between two vertices if and only if the two vertices are closer than the predetermined threshold (are “in contact”). Contacts between consecutive nucleotides are ignored since they share a bond and are therefore automatically in contact.

Figure 3.12 show the contact maps of two small RNA molecules. In each, vertices represent the nucleotides in the order they appear in the RNA. Edges are drawn between vertices if the two nucleotides they represent are close in the tertiary structure. For example, in $A$,
the first nucleotide is in contact with nucleotide 3, but not with any others. Nucleotide 8 is in contact with three other nucleotides—nucleotides 4, 5, and 10.

Figure 3.12: Contact maps of two small RNA molecules. Each molecule consists of 10 nucleotides and 7 contacts although the contacts occur between different pairs of nucleotides.

For a given alignment of the nucleotides of $A$ and $B$, a contact map overlap is defined as a contact in $A$ whose endpoints are aligned with endpoints of a contact in $B$. In the contact map overlap problem, the objective is to find the alignment with that maximizes the number of contact map overlaps.

In Figure 3.13 overlaps are denoted by thick edges. For example, because the dashed lines show an alignment of nucleotides $2^A$ with $1^B$ and $5^A$ with $4^B$, the contact $(2^A, 5^A)$ and the contact $(1^B, 4^B)$ overlap. Figure 3.13 displays the optimal alignment (which has five contact overlaps) for the RNAs given in Figure 3.12.

**Theorem 3.3.2.** *MSN is NP-complete*

*Proof.* To show that MSN is NP-complete, we must show that given an arbitrary instance of the contact map overlap (CMO) problem, its solution can be determined by using the solution of an instance of the MSN problem.
Figure 3.13: Alignment that produces the optimal contact map overlap. Dashed lines connect aligned nucleotides. Thick lines represent overlapping contacts.

Suppose we have an instance of the CMO problem. That is, we have two structures $A$ and $B$, the contacts in $A$, the contacts in $B$, and a positive integer $k$. Let $m$ be the number of nucleotides in $A$, $n$ the number of nucleotides in $B$, $p$ the number of contacts in $A$, and $q$ the number of contacts in $B$. We want to know if there exists an alignment of the nucleotides of $A$ and $B$ such that $k$ or more contacts overlap.

Now we construct the instance of the MSN problem. Let $A^*$ and $B^*$ consist of $2m$ and $2n$ nucleotides, respectively. For each contact in $A$, $(i^A, j^A)$, we create the following 4-nucleotide neighborhood in $A^*$: $(2i^A, (2i + 1)^A, 2j^A, (2j + 1)^A)$. In this way, we create $p$ 4-nucleotide neighborhoods in $A^*$ and $q$ neighborhoods in $B^*$ using the contacts of $B$. It is also important to note that we suppose that all of the pairs (one from $A^*$ and one from $B^*$) of neighborhoods are structurally similar.

Suppose an alignment of $A^*$ and $B^*$ exists such that $k$ 4-nucleotide neighborhoods are aligned. Then we can list the $k$ aligned neighborhoods as follows:
\[(r_1^A, s_1^A, t_1^A, u_1^A) \quad (w_1^B, x_1^B, y_1^B, z_1^B)\]
\[(r_2^A, s_2^A, t_2^A, u_2^A) \quad (w_2^B, x_2^B, y_2^B, z_2^B)\]
\[\vdots\]
\[(r_k^A, s_k^A, t_k^A, u_k^A) \quad (w_k^B, x_k^B, y_k^B, z_k^B)\]

where for each \(i \in \{1, 2, \ldots, k\}\), \(r_i^A\) is aligned with \(w_i^B\), \(s_i^A\) is aligned with \(x_i^B\), \(t_i^A\) is aligned with \(y_i^B\), and \(u_i^A\) is aligned with \(z_i^B\).

Now we create an alignment of \(A\) and \(B\) in which for each \(i \in \{1, 2, \ldots, k\}\) we align \((\frac{r_i}{2})^A\) with \((\frac{w_i}{2})^B\) and \((\frac{t_i}{2})^A\) with \((\frac{y_i}{2})^B\). Since for each \(i\), \((\frac{r_i}{2})^A\) is in contact with \((\frac{t_i}{2})^A\) and \((\frac{w_i}{2})^B\) is in contact with \((\frac{y_i}{2})^B\), this alignment will indeed have \(k\) contact overlaps. It remains to be shown that this alignment is well-ordered and uniquely-assigned.

Suppose the alignment is not uniquely-assigned. Then there exist \(i^A\) aligned with \(j^B\) and \(j^B\), \(j \neq j'\). Then \((2i)^A\) is aligned with \((2j)^B\) and \((2j')^B\), which is a contradiction since the alignment of \(A^*\) and \(B^*\) is uniquely-assigned.

Suppose the alignment is not well-ordered. Then there exist \(i^A\) aligned with \(j^B\) and \(i'^A\) aligned with \(j^B\) such that \(i < i'\) and \(j > j'\). Then \((2i)^A\) is aligned with \((2j)^B\) and \((2i')^A\) is aligned with \((2j')^B\), with \(2i < 2j\) and \(2i' > 2j'\), which is a contradiction since the alignment of \(A^*\) and \(B^*\) is well-ordered.

Now we suppose that there does not exist an alignment of \(A^*\) and \(B^*\) such that \(k\) 4-nucleotide neighborhoods are aligned. It will be shown that this implies there is no alignment of \(A\) and \(B\) such that with \(k\) contact overlaps.

Suppose on the contrary that there is an alignment of \(A\) and \(B\) with \(k\) overlaps. Then \(k\) aligned contacts can be listed as follows:

\[(r_1^A, s_1^A) \quad (w_1^B, x_1^B)\]
\[(r_2^A, s_2^A) \quad (w_2^B, x_2^B)\]
\[\vdots\]
\[(r_k^A, s_k^A) \quad (w_k^B, x_k^B)\]
where for each $i \in \{1, 2, \ldots, k\}$, $r_i^{A*}$ is aligned with $w_i^{B*}$ and $s_i^{A*}$ is aligned with $x_i^{B*}$.

Now we create an alignment of $A^*$ and $B^*$ in which for each $i \in \{1, 2, \ldots, k\}$ we align $(2r)_i^A$ with $(2w)_i^B$, $(2r + 1)_i^A$ with $(2w + 1)_i^B$, $(2s)_i^A$ with $(2x)_i^B$, and $(2s + 1)_i^A$ with $(2w + 1)_i^B$. This alignment of $A^*$ and $B^*$ has $k$ aligned neighborhoods and is well-ordered and uniquely assigned, a contradiction. Thus, no such alignment of $A$ and $B$ exists.

Hence, the MSN problem is NP-complete. □
CHAPTER 4

R3D Align

The ideas described in the previous section have been implemented in “R3D Align” (“RNA 3D Align”), a suite of programs available for download at http://r3dalign.googlecode.com. R3D (pronounced “Red”) Align has been programmed in the Matlab programming language for easy inclusion in FR3D [59], the set of programs which it is built upon. R3D Align uses existing FR3D programs for auxiliary functions such as calculating geometric discrepancies and reading the 3D information from PDB files. Details of the implementation of R3D Align are provided in this chapter. A user manual is also available at the download location, and is provided in Chapter 8. Performance characteristics are discussed in Chapter 5.

As discussed in Section 2.4, the maximum clique problem is NP-complete [23]. So unless P=NP, there is no algorithm that can guarantee finding the maximum clique of a general graph in polynomial time. However, we are not dealing with arbitrary graphs; the local alignment graphs are formed in a specific way. Then it is quite possible these graphs possess specific features which allow the maximum clique to be found more quickly than in the general case.

Also, because RNA molecules are limited in size, the size of the local alignment graphs will be limited as well; they will not be arbitrarily large. Thus, if meaningful alignments can be obtained for the largest RNA molecules (such as 23S rRNA) in a reasonable amount of
time, we can be assured the method can be used in the alignment of smaller molecules as well.

R3D Align uses a branch and bound algorithm to find the maximum clique of the local alignment graph. The branch and bound procedure does guarantee finding the maximum clique of the graph, although it does not require explicit exploration of every possible solution, which is essentially impossible since the number of potential solutions increases exponentially as the problem size increases. However, before running the branch and bound algorithm, which will be described in Section 4.6, R3D Align implements a series of techniques and provides user options to ensure an alignment may be produced in a reasonable amount of time, which will now be described.

### 4.1 Vertex Preprocessing

Given a graph $G = (V, E)$, R3D Align uses a number of techniques to quickly eliminate vertices that cannot be part of the maximum clique of $G$, $MC_G$. For each of these techniques, a lower bound $\sigma$ on the size of the clique must first be found. Then for each vertex $v_i \in V$, we let $G_{v_i}$ be the subgraph induced by $\delta_G(v_i)$. That is, $G_{v_i}$ consists of the vertices adjacent to $v_i$ and edges between vertices as in $G$. An upper bound $k$ on the size of the maximum clique of $G_{v_i}$ is found and compared with the lower bound. If $k + 1 < \sigma$, then it is known that $v_i$ is not part of the maximum clique ($v_i \not\in MC_G$). A variety of techniques will be employed to calculate an upper bound, but first we will discuss how to find a lower bound on the size of the maximum clique.

#### 4.1.1 Lower Bound on the Size of the Maximum Clique

To find a lower bound we may use any heuristic procedure that will produce a maximal clique (Section 3.2.4). One algorithm that finds a maximal clique very quickly is the highest-degree-first algorithm. In this procedure, a candidate set is formed that contains all vertices which
can be added to the current clique to increase its size by one. The candidate set initially contains all the vertices as the current clique does not contain any vertices. The vertex in the candidate set with the maximum number of adjacent vertices in the candidate set is added to the current clique. The candidate set is then updated by removing all vertices not adjacent to the last vertex added to the clique. This procedure continues until the candidate set is empty. Pseudocode is given in Algorithm 1.

**Algorithm 1 Maximal Clique**

Input: $G = (V, E)$  
$W := V$  
$C = \emptyset$  

while $W \neq \emptyset$ do  
  Select $v \in W$ such that $|\delta_W(v)| \geq |\delta_W(w)| \forall w \in W$  
  $C = C \cup v$  
  $W = W \cap \delta_W(v)$  
end while

For a graph of size $n$, the computational time is $O(n^2)$ for this procedure. This means that this method is guaranteed to quickly find a maximal clique, whose size may be used as a lower bound for the maximum clique of the graph.

### 4.1.2 Graph Coloring

A graph coloring is an assignment of colors to each vertex of the graph such that no pair of adjacent vertices share the same color. That is, a graph coloring is a partition of the vertices into color classes $(C_1, C_2, \ldots, C_k)$, where each $C_i$ is an independent set (a set with no two vertices being adjacent) containing those vertices assigned to the $i^{th}$ color. Examples of validly colored graphs are given in Figures 4.1 and 4.2.

Because all pairs of vertices are adjacent in a clique, any valid coloring of a clique of size $M$ requires $M$ colors. Figure 4.2 makes this clear. It follows that a graph that can be colored by $k$ colors cannot contain a clique of size larger than $k$. 
Figure 4.1: Graph consisting of 10 vertices and 3 color classes, which are also numbered. No pair of adjacent vertices in the graph belong to the same color class.

We use this fact to identify vertices that are not part of the maximum clique. For each vertex, we color the subgraph induced by its neighboring vertices. If the number of colors used is less than $\sigma - 1$, then the vertex is not part of the maximum clique and can be removed from the local alignment. Each graph coloring requires $O(n^2)$ time, so this preprocessing step of coloring all subgraphs induced by the neighborhood of each vertex is done in $O(n^3)$ time.

**Coloring all vertices in $O(n^2)$ total time instead of $O(n^3)$**

However, we can do better than this. We have found that it is possible to do the coloring stage in $O(n^2)$ total time. To do this first color the entire local alignment graph. Then for each induced subgraph, we induce the color of the vertices as well. This is indeed a valid coloring for each subgraph, although the subgraphs may be divided into more color classes than if the subgraphs were individually colored rather than inferred from the global coloring. However, it has been observed that the time saved by only performing one coloring is more beneficial than achieving possibly lower upper bounds.

Determining the optimal coloring (the coloring that uses the least number of colors possible) of an arbitrary graph has been shown to be NP-Complete [23]. Therefore, heuristic
coloring algorithms are employed that can color the graph very quickly, although more colors may be used than is optimal. A few of the many available coloring algorithms are now described.

**Highest-degree first-available-color method**

We first describe a greedy coloring procedure that runs in $O(n^2)$ time, where $n$ is the number of vertices of the graph. This greedy procedure can accurately be described as the highest-degree first-available-color method. In this algorithm, the uncolored vertex of highest degree is selected and assigned to the lowest color class that does not already contain a vertex that is adjacent to the selected vertex. The procedure contains one vertex at a time until all vertices have been assigned a color class. Pseudocode is given in Algorithm 2.

**DSATUR**

Another approximate coloring method is known as DSATUR (degree of saturation largest first) [7]. The saturation degree of a vertex is defined as the number of different colors to
Algorithm 2 Highest-degree first-available-color method

\begin{algorithm}
\begin{algorithmic}
\Input{$G = (V, E)$}
$S := V$
\While{$S \neq \emptyset$}
\State Select $v \in S$ such that $|\delta_S(v)| \geq |\delta_S(w)| \ \forall \ w \in S$
\State $k = \min\{i \in N | C_i \cap \delta_{V \setminus S}(v) = \emptyset\}$
\State $C_k := C_k \cup v$
\EndWhile
\end{algorithmic}
\end{algorithm}

which it is adjacent. DSATUR selects the next vertex to be colored by choosing the available vertex of highest saturation degree. The selected vertex is then assigned the lowest number color possible. This algorithm begins by assigning different colors. It is not difficult to verify that all vertices used before a color is repeated forms a maximal clique. Thus, this algorithm simultaneously computes upper and lower bounds for the degree of the maximum clique.

It has been shown that for very sparse and very dense graphs DSATUR is an order of magnitude more expensive than the highest-degree first-available-color method described above [3]. Results have shown that the specially defined graphs used to solve our alignment problem are very dense and thus the highest-degree first-available-color method is what is used for our program.

Fractional coloring

More elaborate coloring methods, such as fractional graph coloring, can be more effective in providing tighter upper bounds and thus reducing the total search. However, preliminary experiments have shown that the fractional graph coloring method takes significantly longer to compute and thus has a negative overall effect on performance when working with local alignment graphs.

4.1.3 Neighbor Counting

Oftentimes vertices can be removed from consideration simply based on the number of adjacent vertices (neighbors). For if given vertex has less than $\sigma - 1$ adjacent vertices, then it
certainly is not part of the maximum clique, which is known to have size of at least \( \sigma \). While any vertex eliminated by this method would be eliminated by the graph coloring method described above, it is much faster to count neighbors than to color the neighbors and then count the number of colors used.

### 4.1.4 Specialized Pre-processing

As mentioned above, we are not dealing with arbitrary graphs; the local alignment graphs are formed in a specific way. Then it is possible they possess specific features which allow the maximum clique to be found more quickly than in the general case. Indeed, all the local alignment graphs are designed in the grid-like fashion which all possess the following property:

**Proposition 4.1.1.** No two vertices in the same row (or column) are connected by an edge.

**Proof.** Let \( v_1 \) and \( v_2 \) be two vertices located in the same row of the array. Let \((q,r)\) and \((q,s)\) be the locations on the grid of two vertices \( v_1 \) and \( v_2 \), respectively. Then \( v_1 \) represents an alignment of \( N^A(q) = (i^A_q, j^A_q, k^A_q, l^A_q) \) with \( N^B(r) = (i^B_r, j^B_r, k^B_r, l^B_r) \) and \( v_2 \) represents an alignment of \( N^A(q) = (i^A_q, j^A_q, k^A_q, l^A_q) \) with \( N^B(s) = (i^B_s, j^B_s, k^B_s, l^B_s) \). Because \( N^B(r) \) and \( N^B(s) \) are different neighborhoods, then either \( i^B_r \neq i^B_s \), \( j^B_r \neq j^B_s \), \( k^B_r \neq k^B_s \), or \( l^B_r \neq l^B_s \). Then an edge \((v_1, v_2)\) implies that either \( i^A_q \), \( j^A_q \), \( k^A_q \), or \( l^A_q \) is aligned with two different nucleotides, which violates the uniquely-assigned criterion. Therefore no two vertices in the same row are adjacent. The proof that no two vertices in the same column are adjacent is similar. \( \square \)

Using this result, we derive another technique for eliminating vertices:

**Corollary 4.1.2.** Let \( G = (V, E) \) be a graph and \( \sigma \) be a lower bound on the size of the maximum clique of \( G \) (\( MC_G \)). Then, when the vertices are arranged according to the array \( N^A \times N^B \) as previously described, for every \( v \in V \) such that \( v \) is not adjacent to vertices in at least \( \sigma - 1 \) different rows of the array or at least \( \sigma - 1 \) different columns of the array, \( v \notin MC_G \).
Again, this technique will not eliminate any vertices that could not be eliminated by the graph coloring technique. However, it is much faster to count the number of rows and columns the adjacent vertices are in than to color all the adjacent vertices. So it is useful to use this technique to eliminate vertices before graph coloring are found.

4.2 Controlling the Number of Neighborhoods

We have also implemented other techniques to improve runtimes. As previously described, for each RNA molecule \( A \), we derive a list \( N^A \) of neighborhoods where each neighborhood consists of four nucleotides that are in the same sphere of the pre-specified diameter. When choosing this diameter one needs to maintain the balance between having enough neighborhoods in order to represent the structure well, but not too many since the computational time increases exponentially as the number of neighborhoods increases. Because of the physical space requirements of the nucleotides, there is a limit to the number of neighborhoods any particular nucleotide can be in for a given diameter \( D \). However, due the complex 3D shape of the RNA molecule, this number can vary from nucleotide to nucleotide. For example, the nucleotides in a hairpin with no tertiary contacts typically are only close to a few other nucleotides and therefore will not be in as many neighborhoods as nucleotides found in tightly compacted regions. Figure 4.3 is a scatterplot showing the number of neighborhoods each nucleotide of the 5S rRNA molecule, H.M. (crystal structure 1s72), is a part of when a neighborhood radius of 10Å is used. The total number of neighborhoods is 873.

The figure illustrates that the approach can result in both ‘underrepresentation’ and ‘overrepresentation’ of certain nucleotides. We will describe both of these issues and then describe a method to balance the neighborhood representation.

A major problem is that certain nucleotides (such as nucleotide 23) are not in any four-nucleotide neighborhoods. That is, there is no sphere of diameter 10Å that includes nucleotide 23 and three other nucleotides. This will often be the case for certain bases that
Figure 4.3: Scatterplot displaying the number of neighborhoods each nucleotide is in using diameter $D = 10\text{Å}$.

are extruding from the structure. Not being a part of any neighborhood will necessarily prohibit the nucleotide from being a part of a neighborhood that is structurally similar to a neighborhood in the other structure. This will then prevent the nucleotide from being a part of the final alignment. This is problematic since many times these bases will indeed correspond to a base in another structure that is in a similar location. For example, two structures often have bulged bases in respectively similar locations that should be aligned. These bases need to be a part of some neighborhood in order to have the opportunity to be aligned.

One could simply expand the diameter of the neighborhood to ensure that all nucleotides are included in at least one neighborhood. However, this greatly increases the total number of neighborhoods to an unmanageable number. In the example above, to include every nucleotide in at least one neighborhood, it is necessary to increase the neighborhood radius from $10\text{Å}$ to $15\text{Å}$. However, doing so causes a jump in the total number of neighborhoods from
873 to 8532. The effect on runtime would be tremendous. Figure 4.4 gives the scatterplot when $D = 15\text{Å}$.

![Figure 4.4: Scatterplot displaying the number of neighborhoods each nucleotide is in using diameter $D = 15\text{Å}$.

While some nucleotides are not represented in any neighborhoods, other nucleotides are found in many neighborhoods, perhaps too many. For example, nucleotide 43 is a member of 91 different neighborhoods when $D = 10\text{Å}$. Each of these neighborhoods must be compared with each neighborhood of the other structure. Comparing so many neighborhoods is costly in terms of time, and because many of those neighborhoods have overlapping components, the information gathered in the comparisons is often redundant. For if two structures are very similar up to 91 different structurally similar neighborhoods would imply the correspondence of nucleotide 43, which is more than necessary. We do not need so many to confirm the alignment of this nucleotide.

A technique of expansion and refinement is used to resolve the issue of underrepresen-
tation as well as to help control the total number of neighborhoods considered. First the neighborhood diameter is found that is large enough to ensure that all nucleotides are included in at least $p$ neighborhoods. Then for each nucleotide that is included in more than $p$ neighborhoods, only the $p$ neighborhoods with the smallest diameter are retained. This approach controls the total number of neighborhoods we consider. This provides a means of controlling the size of the local alignment graph in which the maximum clique must be found to produce the alignment. It also balances the representation of diffuse and compact regions in the local alignment graph so that some regions are not more represented than others.

4.3 Seed Alignments

Computational time can also be reduced if we limit the pairs of neighborhoods for which the discrepancy is calculated and which could become vertices in the local alignment graph. One way to limit the pairs of neighborhoods for which the discrepancy is calculated is to limit which pairs of nucleotides in the two structures are compared. In the alignment of two molecules, all $m$ nucleotides are nominally compared with all $n$ nucleotides of structure B. For example, in the alignment of the 16S rRNA molecules, *E.coli* (which consists of 1530 nucleotides) and *T.thermophilus* (1513 nucleotides), all $1530 \times 1513 = 2,314,890$ combinations of nucleotides would be considered. However, in situations in which the alignment is expected to include substantially all nucleotides of structures $A$ and $B$, there is no point in superimposing neighborhoods consisting of nucleotides far apart in their respective linear sequences, say, nucleotides $(1^A, 2^A, 3^A, 4^A)$ with nucleotides $(1491^B, 1492^B, 1493^B, 1494^B)$. A bandwidth $\beta$ can be specified so that nucleotide $i^A$ is compared only with nucleotides in the set $\{\min\{1, i^B - \beta\}, \min\{1, i^B - \beta + 1\}, \ldots, \max\{i^B, n^B\}, \ldots, \max\{n^B, i^B + \beta - 1\}, \max\{n^B, i^B + \beta\}\}$. For example if $\beta = 25$, then only 76,802 combinations of nucleotides will be considered, meaning over 96% of the nucleotide combinations will not be considered by setting a bandwidth parameter.
An even better method than relying on the nucleotide numbers is to restrict the nucleotides compared based on a seed alignment that is reasonably good. Oftentimes fast alignment methods (such as sequence alignment methods) can be used to produce an alignment which is close to the optimal one, but the alignment of nucleotides may be off by several positions in certain places. Sequence alignments often are not precise at the individual nucleotide level due to sequence variability. Given a seed alignment, we define a function $\psi$ for which $\psi(a) = b$, where $b$ is the index of the nucleotide in structure B that is aligned to nucleotide $a$ in A. If $a$ is not aligned to any nucleotide in B (i.e., $a$ is aligned to a gap), then we find the closest nucleotide to $a$ that is not aligned with a gap. Let that nucleotide index be $a^*$. If two such nucleotides exist, let $a^*$ be the smaller of the two. Then we set $\psi(a) = \psi(a^*)$.

Again, a band width parameter $\beta$ can be used to indicate how many bases to the left and right of $\psi(a)$ to consider. That is, a band width $\beta$ can be specified so that nucleotide $i^A$ is compared only with nucleotides in the set \{min\{1, $\psi(i^A) - \beta\}$, min\{1, $\psi(i^A) - \beta + 1\}$, $\ldots$, $\psi(i^A)$, $\ldots$, max\{$n^B$, $\psi(i^A) + \beta - 1$\}, max\{$n^B$, $\psi(i^A) + \beta$\}\}.

### 4.4 Screening Criterion

Since we are only seeking pairs of neighborhoods whose geometric discrepancy is less than some cutoff value $d$, we can also decrease runtime by utilizing the screening criterion described in [59]. The screening algorithm can be used to rapidly identify many pairs of neighborhoods whose geometric discrepancy is higher than $d$ without the use of detailed and time-intensive calculations.

### 4.5 Greedy Cliques Produce Alignments

By the way the edges were created in the local alignment graph, every clique (whether or not it is a maximum clique) corresponds to a well-ordered and uniquely-assigned alignment.
Thus, after pre-processing the vertices as described above, the greedy clique-finding procedure may be used to find a valid alignment. R3D Align provides the user the option of using the greedy clique-finding procedure instead of the branch and bound procedure to produce a final alignment. Because the pre-processing typically produces a dense graph, this approach often produces a clique similar in size to the maximum clique. For example, in the alignment of 16S rRNA (see Chapter 5), the greedy clique procedure finds a clique of size 3748, while the maximum clique is found to consist of 3762 vertices. The alignments produced by using the greedy approach are useful for exploratory work, but are often very similar to the alignment produced when the maximum clique is found.

### 4.6 Branch and Bound

If it is desired to find the alignment that corresponds to the full maximum clique, a branch and bound algorithm can be used. However, since the problem is NP-complete, algorithms that find the maximum clique, such as the branch and bound algorithm implemented by R3D Align, may have relatively long running times in certain cases. Several different algorithms in the literature utilize a branch and bound approach to find a maximum clique of a general undirected graph, such as those found in [11] [2] [65]. These algorithms are not specifically designed for the alignment problem. The branch and bound algorithm implemented by R3D Align is now described.

As was mentioned above, the branch and bound procedure does guarantee finding the maximum clique of the graph, although it does not require explicit exploration of every possible solution, which is essentially impossible due to the exponential increase in the number of potential solutions as the problem size increases. The branch and bound technique seeks to overcome this limitation by using upper bounds as well as using the current best solution as a lower bound to search part of the solution space only implicitly by eliminating large subsets of candidates having no potential of being an optimal solution.
In a branch and bound algorithm, the original problem is divided into sets of subproblems (this is the “branching” step) in such a way that the optimal solution is contained in one of the sets of subproblems (i.e., it is contained on one of the branches). Then for each branch, an upper bound is found on the size of the maximum clique for that set of subproblems. If the upper bound is less than the size of the current largest maximum clique found, then the branch is “pruned” and none of the subproblems contained on that branch are explored any further since it is known that they will not produce the maximum clique. Thus, much of the solution space may not be searched explicitly.

There are three key issues in using branch and bound algorithms for the maximum clique problem [51]:

1. How to find a good lower bound, i.e., a clique of large size?

2. How to find a good upper bound on the size of a maximum clique?

3. How to branch, i.e., break a problem into smaller subproblems?

Similar methods for finding upper and lower bounds can be used, as were described in Section 4.1. That is, heuristic clique finding procedures can be used to find the lower bounds (the size of the largest clique found so far) and graph colorings can be used to find the upper bounds (the largest clique a particular set of vertices could possibly contain).

Now we turn our attention to branching, the process of dividing the problem into subproblems. Given graph $G = (V, E)$, with $V = \{v_1, v_2, \ldots, v_n\}$, a simple branching would be one that considers subgraphs $S_i, i = 1, 2, \ldots, n$, induced by the vertex sets $\delta(v_i) \cup v_i$. The problem of finding the largest clique in $G$ can now be found by determining the largest clique in each of the $n$ smaller (only equal if $G$ is a clique) subgraphs. It is clear that the maximum clique of $G$ is contained on one of the branches. However, given this branching, the maximum clique $MC_G$ is in fact contained on $|MC_G|$ branches. This redundancy causes every clique to be discovered $|MC_G|$ times, which increases the runtime of the algorithm.
Preferably, we assign the branches in such a way that the maximum clique of $G$ is a subgraph of one and only one branch. This is accomplished by letting each $S_i$ be the subgraph induced by the vertex sets $V_i \cup \{v_i\}$, $V_i := \delta(v_i) \cap \{v_1, v_2, \ldots, v_{i-1}\}$. It can easily be verified that regardless of the ordering of the vertices, the maximum clique of $G$ is a subgraph of exactly one $S_i, i = 1, \ldots, n$. We have found it beneficial to order the vertices according to degree so that highest ordered vertices are considered first and good lower bounds may be discovered earlier in the procedure so that later branches may be eliminated without much effort.

For each $i$, let $G_i$ be the subgraph induced by the vertex set $S_i \setminus \{v_i\}$. If a clique of size $m$ can be found in $G_i$, then obviously a clique of size $m + 1$ can be found in $S_i$. Consequently, if $G_i$ can be colored with $\chi(G_i)$ colors then $S_i$ can be colored with $\chi(G_i) + 1$ colors. So if $\chi(G_i) + 1 > \omega^*$, the size of the current best clique, we split $G_i$ into subproblems in a similar fashion as we did to $G$. This process results in the construction of a search tree with the root corresponding to the original problem.

Each node $N_t$ of the tree is comprised of a subgraph $G_t = (V_t, U_t)$ where $V_t$ is the set of vertices in the subgraph at that node and $U_t$ is the set of vertices of $G$ assigned to $G_t$ on account of the previous branchings. For each node, if $V_t$ can be colored with $\chi(V_t)$ colors and $\chi(V_t) + |U_t| > \omega^*$, then $|V_t|$ more branches are created, each corresponding to a maximum clique problem in a smaller graph. If $\chi(V_t) + |U_t| \leq \omega^*$, then none of those smaller graphs need to be considered for there is no possibility of any of them containing a maximum clique of higher cardinality than the current largest clique found. This algorithm has been adapted from the branch and bound method presented by Babel and Tinhofer [2].
CHAPTER 5

Alignment of Bacterial 16S rRNA Structures

We tested R3D Align by aligning the 3D structures of the 16S rRNAs of *E.coli* (*E.c.*) and *Thermus thermophilus* (*T.th.*), PDB files 2avy [60] and 1j5e [70], respectively. These structures are an appropriate test case because the 16S is large and highly structured, with many non-Watson-Crick basepairs and long-range interactions. Moreover, *E.c.* and *T.th.* are phylogenetically distant so there are significant differences in sequence and structure. On a practical level, it is easy to compare the two structures because the crystallographers who solved the *T.th.* structure numbered the nucleotides to correspond as closely as possible to those of the *E.c.* sequence.

In this chapter, we will compare an alignment of the two structures produced by R3D Align with the alignments produced by a number of other methods.

To align the two structures, we used an iterative approach. First we used a Needleman-Wunsch sequence alignment algorithm to obtain an initial seed alignment to feed to R3D Align. We then ran R3D Align with a band width $\beta = 30$ and $p = 3$ (for command line input, see Example 4 in the R3D Align user manual). Setting a low value of $p$ limits the number of neighborhoods and produces an alignment more quickly. The greedy clique procedure
was used to get an alignment which was much more accurate than the seed alignment. This alignment, which took a total of three minutes to produce, was then used as the seed alignment for a second iteration of R3D Align. For this second and final iteration (Example 5 in the R3D Align user manual), we were able to use a more narrow band width $\beta = 10$ since our seed alignment was more accurate, and we set $p = 9$ in order to get an accurate alignment by using more neighborhoods. We again used the greedy clique procedure to produce the final alignment. This step took ten minutes. For both iterations, the geometric discrepancy cutoff $d = 0.5$ was used. All work was done a standard laptop.

5.1 Tools and Diagnostics for Displaying and Evaluating Alignments

To evaluate the output of R3D Align and compare results with other alignment programs, we have developed new diagnostic tools for the structural evaluation of nucleotide to nucleotide alignments. We will show that the diagnostics can be meaningfully applied to alignments produced by sequence-only methods as well.

The classical tool for 3D-to-3D alignment evaluation has been the average root-mean-square deviations (RMSD) after optimal superimposition of the atomic coordinates of the set of aligned nucleotides in each structure. While the RMSD does reveal information about the global structure of the aligned nucleotides, local neighborhoods may actually differ quite substantially, as the RMSD spreads errors over the whole structure. Using the RMSD alone, it is difficult to analyze whether specific local regions are aligned sufficiently well or not. Also, the RMSD fails to give detailed information about the conservation of interactions among nucleotides between the two structures, whether they are local or long-range interactions. It is also not always sensible to simply compare the RMSD values for two different alignments because different methods often align a different number of nucleotides.

In [52], it was recognized that RNA molecules have specific structural features, and it is
necessary to have tools that evaluate methods based on these specific RNA structural features rather than on global average measurements. New metrics and diagnostics were introduced, including the deformation index and deformation profile, to aid in RNA structure comparison and to overcome the shortcomings of using the RMSD alone. These tools evaluate different modeling techniques and are useful for comparing a modeled structure with an experimentally resolved structure; that is, they are useful in the case when the two structures being compared represent the same organism (so the alignment is trivial). Because R3D Align is used to align structures of different organisms, we have developed a set of tools to aid in structure comparison and alignment evaluation when the alignment is non-trivial.

5.1.1 Displaying Alignments

Even after producing an alignment of two large RNA molecules, a challenge that remains is that it is difficult to communicate an alignment to a human reader. Typically, an alignment is displayed by listing the sequences in two rows with aligned nucleotides being listed in the same column as in Figure 5.1.7. Some of the limitations of conventional 2D matrix representations of biological sequence alignments for RNA were recently discussed in depth [9]. For larger RNA molecules, the output is typically produced in FASTA format only. Figure 5.1 is the FASTA output for the alignment of \( T.th. \) and \( E.c \) produced by R3D Align. It is difficult to gain much insight into the alignment when given this because the sequences are often so long and contain so little information in and of themselves since so much of the information contained in the 3D structure is lost in this kind of representation.

Sometimes a superposition of the two aligned 3D structures is given, which is helpful in some ways, but this does not let one look at and study local sections of the structure in detail. And because our alignment focuses on local regions superimposing well instead of just the global structure, it isn’t much use to display just a full superposition of the two structures. In any case, two full superimposed structures are often difficult to comprehend, especially when the structures are large. Figure 5.2 shows a superposition of the sets of aligned nucleotides
Figure 5.1: FASTA output for the alignment of *T.th.* and *E.c* produced by R3D Align
of two 16S rRNA molecules. The alignment consists of 1400 correspondences.

Figure 5.2: Superposition of the aligned sets of nucleotides for R3D Align alignment. The nucleotides of \textit{T.th.} (1j5e) are shown in black and the nucleotides of \textit{E.c.} are shown in gray.

We have found an intermediate between a sequence alignment in two-row form and a 3D superposition that is useful for displaying the alignment information.

Because the 3D structure is too informative for practical use, we distill the information into a set of interactions. This set of interactions could include any type of nucleotide interactions, such as basepairing, base-stacking, and base-phosphate interactions. We focus on just the basepairs. We create a basepair list (as in \[63\]) which contains every basepair
of the structure, as annotated by FR3D [59]. This is the most useful way of looking at a molecule as the 3D structure of an RNA molecule is basically determined by the interactions taking place between individual nucleotides. The secondary structure only consists of some of the interactions. Thus, what the basepair list contains is the useful information about the structure.

A 3D-to-3D alignment implies similarity in the 3D structures of the aligned regions. Then it stands to reason that if the 3D structures are said to align, then the interactions of the aligned regions should correspond as well. For example, if two nucleotides are forming a basepair in the first structure and are aligned to two other nucleotides in a second structure, then the two nucleotides in the second should be forming a basepair as well. Moreover, these two basepairs should most often be members of the same geometric basepair family.

The basepair list is organized in such a way as to include all of the alignment information in addition to all of the basepairing information for both structures. The list, which is produced in spreadsheet form, simultaneously displays all the basepairs and alignment information of the two structures. The spreadsheet is produced automatically by R3D Align upon completion of the alignment and is saved in the current Matlab working directory.

Figure 5.3 displays a portion of the spreadsheet displaying the alignment of two the 16S bacterial rRNAs, E.c. and T.th. The full spreadsheet is given in Appendix A. Each nucleotide from T.th. is listed in Column 1 by its base letter and number as in the PDB file. The nucleotide in E.c. to which it is aligned, if any, is listed in Column 4 of the same row. If there is no corresponding nucleotide in the E.c. structure, the entry in Column 4 is left blank. For example, Figure 5.3 shows that U1302 of T.th. is aligned with a gap; consequently, no nucleotide is listed in Column 4 of that row. Likewise, if a nucleotide in E.c. in Column 4 is not aligned to a nucleotide in T.th., the cell in Column 1 of the corresponding row is left blank (e.g. C1302). Blank entries such as these indicate that the nucleotide does not have a structural correspondence. The nucleotide may have a correspondence at the level of sequence homology, however. In Figure 5.3 it is likely that U1302 in T.th. is
homologous to C1302 in *E. c.*, but they do not superimpose well in 3D. Interactions classified as near-basepairs are also listed when the corresponding nucleotides in the other structure are basepaired.

To capture 3D context, the alignment also shows the basepairs made by each nucleotide in each structure. The first three columns are used to list basepairs in the first structure (*T. th.* in the example) and the next three columns, the corresponding basepairs in the second structure (*E. c.*). The type of basepair formed between the nucleotides in each structure, if any, is indicated in the second and fifth columns and is annotated according the Leontis/Westhof system. Cells are colored according to basepair types for ease of comprehension. For example Watson-Crick pairs (“cWW”) have green background.

Only the nucleotides in Columns 1 and 4 of the alignment are listed in (necessarily) increasing numerical order. When nucleotides in the first structure make more than one basepair they are listed as many times as necessary so that each basepair is in a separate row. U1301 in Figure 5.3 provides an example. If there is a nucleotide in Column 3, then Column 6 of the spreadsheet shows the nucleotide in *E. c.* to which it is aligned, if there is one. Then the basepair type of the corresponding nucleotides in Columns 4 and 6 is listed in Column 5. If the nucleotides in Columns 4 and 6 do not make a basepair interaction, ‘—’ is displayed. The spreadsheet thus shows every basepair interaction detected in each structure in addition to all of the alignment information. By indicating the basepairs made by the aligned nucleotides, we have a good indication of what is happening in the 3D structure at a very local level.

We find that for good alignments, pairs of aligned nucleotides make the same basepair types. That is, identical basepairs in Columns 2 and 5 (as above) of the same row is generally a sign of proper alignment. Because the basepair types are color coded, one can very quickly scan the spreadsheet and distinguish properly aligned regions, poorly aligned regions, and regions that perhaps should have been aligned. The spreadsheet in Figure 5.3 gives evidence that the two regions are aligned properly because of the substantial agreement of the basepair
Figure 5.3: A portion of the R3D Align alignment of \textit{T.th.} and \textit{E.c.} 16S rRNA. The spreadsheet simultaneously displays aligned nucleotides and basepairing interactions. Within each row, Columns 1 and 4 are aligned, as are Columns 3 and 6. The basepairing interaction between the nucleotides of Columns 1 and 3 (if any) is indicated in Column 2. Column 5 indicates the basepairing interaction between the nucleotides in Columns 4 and 6. Basepairs are annotated by FR3D using the Leontis/Westhof system. The last row shows a near tSW basepair in \textit{E.c.} Note that nucleotides U1302 in \textit{T.th.} and C1302 in \textit{E.c.} are not aligned to one another. This section indicates strong conservation between the two structures as well as proper alignment.
interactions. Since the basepair interactions match up, the 3D structures are then likely to be very similar.

5.1.2 Spreadsheet vs. 2D matrix and 3D Superpostion

Next, we illustrate by example the ease with which one can use the spreadsheet to distinguish between poorly and properly aligned regions, in contrast to the 2D matrix representation and the full 3D superposition. For this we focus on the alignment of the helical region of T.th. formed by nucleotides 1422-1430 and 1470-1478. Needleman-Wunsch (NW) and R3D Align produce different correspondences for this region.

Figure 5.4 gives the 2D matrix representation for the two alignments. We observe that the two methods produce different alignments. NW does not align all the nucleotides, while R3D Align does. NW, which uses the sequences only, indicates that 12 bases are conserved, while R3D Align has only 10. We cannot gain much more information from this display. We note that from these representations, we have no information regarding whether the correspondences agree structurally and thus cannot evaluate the accuracy of the two alignments.

![NW Alignment](image)

*Figure 5.4: 2D matrix alignment representations of the alignments produced by Needleman-Wunsch (NW) and R3D Align*

In Figure 5.5, we have the 3D superpositions of the aligned nucleotides according to the two alignments. Although the two methods had different nucleotides in E.c. corresponding
with some of the nucleotides in the helix of \textit{T.th.}, both of the of the 3D superpositions look pretty good because the superposition is based on global averages. It could be argued that the figures seem to suggest that the R3D Align alignment may be slightly more accurate, but even if that was known to be the case, we could not determine exactly what correspondences in the NW alignment are inaccurate.

![3D alignment superpositions](image)

\textbf{Figure 5.5:} 3D alignment superpositions of the alignments produced by Needleman-Wunsch (NW) and R3D Align

Now we turn our attention to the alignment spreadsheets for the two methods.

\textbf{Figure 5.6(a)} displays the alignment produced by the Needleman-Wunsch algorithm. Due to the format and color system of the spreadsheet, it can very quickly be determined that individual nucleotides are misaligned. We notice that in most of the rows, two nucleotides forming a cWW basepair are aligned to two nucleotides not forming any basepair at all. Therefore, it is improbable that these two regions are actually aligned properly.

\textbf{Figure 5.6(b)} displays the alignment produced by R3D Align. Even at first glance, it can be observed that the regions are more likely to be aligned correctly because cWW basepairs
Figure 5.6: Spreadsheet representations of the alignments produced by Needleman-Wunsch (NW) and R3D Align

<table>
<thead>
<tr>
<th>Needleman-Wunsch (NW)</th>
<th>T.Th. (1JSE)</th>
<th>E.c. (2AVY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1419 cWW U1481</td>
<td>G1419 cWW U1481</td>
<td></td>
</tr>
<tr>
<td>C1420 cWW G1480</td>
<td>U1420 cWW A1480</td>
<td></td>
</tr>
<tr>
<td>G1421 cWW C1479</td>
<td>G1421 cWW C1479</td>
<td></td>
</tr>
<tr>
<td>G1422 cWW C1478</td>
<td>G1422 cWW U1477</td>
<td></td>
</tr>
<tr>
<td>G1422 ---</td>
<td>G1422 cWW U1478</td>
<td></td>
</tr>
<tr>
<td>G1423 cWW C1477</td>
<td>G1423 A1476</td>
<td></td>
</tr>
<tr>
<td>G1423 C1478</td>
<td>G1423 cWW U1477</td>
<td></td>
</tr>
<tr>
<td>C1424 cWW G1476</td>
<td>U1424 G1475</td>
<td></td>
</tr>
<tr>
<td>C1424 C1477</td>
<td>U1424 cWW A1476</td>
<td></td>
</tr>
<tr>
<td>U1425 cWW G1475</td>
<td>U1425 G1474</td>
<td></td>
</tr>
<tr>
<td>U1425 G1476</td>
<td>U1425 cWW G1475</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>G1475 G1426 cWW U1474</td>
<td></td>
</tr>
<tr>
<td>C1426 cWW G1474</td>
<td>C1427 cWW G1473</td>
<td></td>
</tr>
<tr>
<td>U1427 cWW A1473</td>
<td>A1428 ---</td>
<td></td>
</tr>
<tr>
<td>U1427 A1428 cWW U1472</td>
<td>U1425 cWW G1475</td>
<td></td>
</tr>
<tr>
<td>A1428 cWW U1472</td>
<td>A1429 U1472</td>
<td></td>
</tr>
<tr>
<td>A1428 G1471 A1429 cWW U1471</td>
<td>U1427 cWW A1473</td>
<td></td>
</tr>
<tr>
<td>C1429 cWW G1471 A1430 U1471</td>
<td>A1428 cWW U1472</td>
<td></td>
</tr>
<tr>
<td>C1429 G1470 A1430 cWW U1470</td>
<td>C1429 cWW G1471 A1429 cWW U1471</td>
<td></td>
</tr>
<tr>
<td>C1430 cWW G1470 A1431 U1470</td>
<td>C1430 cWW G1470 A1430 cWW U1470</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R3D Align</th>
<th>T.Th. (1JSE)</th>
<th>E.c. (2AVY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1419 cWW U1481</td>
<td>G1419 cWW U1481</td>
<td></td>
</tr>
<tr>
<td>C1420 cWW G1480</td>
<td>U1420 cWW A1480</td>
<td></td>
</tr>
<tr>
<td>G1421 cWW C1479</td>
<td>G1421 cWW C1479</td>
<td></td>
</tr>
<tr>
<td>G1422 cWW C1478</td>
<td>G1422 cWW U1478</td>
<td></td>
</tr>
<tr>
<td>G1422 ---</td>
<td>G1422 cWW U1478</td>
<td></td>
</tr>
<tr>
<td>G1423 cWW C1477</td>
<td>G1423 A1476</td>
<td></td>
</tr>
<tr>
<td>G1423 C1478</td>
<td>G1423 cWW U1477</td>
<td></td>
</tr>
<tr>
<td>G1424 cWW G1476</td>
<td>U1424 G1475</td>
<td></td>
</tr>
<tr>
<td>C1426 cWW G1474</td>
<td>C1427 cWW G1473</td>
<td></td>
</tr>
<tr>
<td>U1427 cWW A1473</td>
<td>A1428 ---</td>
<td></td>
</tr>
<tr>
<td>U1427 A1428 cWW U1472</td>
<td>U1425 cWW G1475</td>
<td></td>
</tr>
<tr>
<td>U1425 cWW G1475</td>
<td>U1425 cWW G1475</td>
<td></td>
</tr>
<tr>
<td>A1428 cWW U1472</td>
<td>A1429 U1472</td>
<td></td>
</tr>
<tr>
<td>A1428 G1471 A1429 cWW U1471</td>
<td>U1427 cWW A1473</td>
<td></td>
</tr>
<tr>
<td>C1429 cWW G1471 A1430 U1471</td>
<td>A1428 cWW U1472</td>
<td></td>
</tr>
<tr>
<td>C1429 G1470 A1430 cWW U1470</td>
<td>C1429 cWW G1471 A1429 cWW U1471</td>
<td></td>
</tr>
<tr>
<td>C1430 cWW G1470 A1431 U1470</td>
<td>C1430 cWW G1470 A1430 cWW U1470</td>
<td></td>
</tr>
</tbody>
</table>
are aligned throughout the region. This illustrates how the spreadsheet provides valuable information for the evaluation of an alignment.

On a side note, this example also illustrates a case when the Needleman-Wunsch algorithm produces inaccuracies due to high sequence variation in the two regions being aligned. In the proper alignment of this region, there is only 41.7% sequence conservation between the two molecules. It is no surprise then that a method that relies solely on the sequential information may produce inaccurate results. This is one of the main reasons there is such a high demand for structural alignment tools such as R3D Align. Even so, the NW alignment is only off by a few nucleotides and would serve as a good seed alignment for R3D Align as discussed in Section 4.3.

5.2 Comparison of 16S Alignments in Spreadsheet Form

For comparison purposes, we used various methods to create alignments of the two 16S rRNA structures; these are named in bold in this paragraph. First, we used two sequence-based alignment methods. We made an alignment using a Needleman-Wunsch (NW) algorithm in which the parameters were set to give the most favorable alignment in this case. Next, we obtained an alignment using NAST [14], which is a multiple sequence alignment server specifically designed for 16S rRNA sequences. The first 3D-to-3D alignment of these structures that we consider is the “crystallographic alignment”, which is implicit in the numbering scheme used by the crystallographers when the second of the two PDB files was released (called Crystallographer below). In earlier work, a manually-derived alignment of the basepaired nucleotides was constructed [63]. This Basepair alignment is a reliable standard against which we can compare other alignment methods. This manual alignment only aligned basepairs; unpaired nucleotides do not appear in that alignment; consequently, we have made a Composite alignment by adding to the correspondences in the basepair alignment all correspondences from the R3D Align alignment that are consistent with the
basepair alignment. Alignments by SARA and DIAL were provided by the authors of the respective programs. The ARTS alignment was obtained using the ARTS web server [16]. The results obtained using SARSA were not comparable, so we do not present a comparison with PARTS.

We use the spreadsheet display to compare several alignments side by side. Figure 5.7 compares a portion of the R3D Align alignment (using the greedy clique procedure only) to six alignments, the manually-produced Basepair alignment from [63], the sequence alignments produced by NAST and Needleman-Wunsch (NW), and the 3D-to-3D alignments produced by ARTS, SARA, and DIAL. The full comparison with all methods appears in Appendix B. We list all of the basepair interactions of the T.th. structure in Columns 1-3 as above. Then the aligned nucleotides are placed in the subsequent columns, three columns per alignment. Aligned nucleotides are again located in the same row, but no extra rows are added to display all nucleotides from E.c. Figure 5.7 shows that, in the indicated region, R3D Align agrees almost entirely with the manually derived alignment. NAST aligns this region very well. ARTS agrees with R3D Align except that it aligns very few of the nucleotides in the internal loop, helix, and hairpin extending from C149 to U172 in T.th. to the corresponding stem in E.c. The SARA alignment generally agrees with the basepair alignment, but fails to correctly align two nucleotides in the hairpin extending from G159 to A162. The DIAL alignment is off by one or two nucleotides in each position and consequently doesn’t align any basepairs. The Needleman-Wunsch alignment, which allows inexact matches, also struggles in places because the two structures have relatively low (69%) sequence similarity within this region. However, NW performs better in certain places where the sequence conservation is higher, such as the helix and internal loop of G144 to C153 and G168 to C178. There the sequence conservation is 86%. Note the long-range interactions between A151 and G102, between A152 and G68, and between A160 and G347. C142-A196 is a basepair across a junction. R3D Align does well with such long-range interactions.
Figure 5.7: A section of the comparison of the 16S rRNA alignment produced by different methods. To the left of the black line are the nucleotides and basepairs of *T. th.* To the right of the black line are the corresponding nucleotides of *E. c.* as aligned by the different methods. R3D Align is the first method displayed to the three other 3D-to-3D alignments, and then two sequence based alignments.
5.3 Alignment Diagnostics

The spreadsheet displays the details of the alignments including annotations of base interactions, but is over 1800 lines long and does not report on the quality of the geometric superposition of aligned neighborhoods. Therefore, we also computed several global diagnostics to compare the alignments and present these in Figure 5.8.

<table>
<thead>
<tr>
<th>Number of nucleotides aligned</th>
<th>Composite</th>
<th>Manual</th>
<th>Automated 3D-to-3D</th>
<th>Sequence-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1414</td>
<td>1208</td>
<td>1400</td>
<td>1482</td>
</tr>
<tr>
<td>Number of exact base matches</td>
<td>1098</td>
<td>924</td>
<td>1089</td>
<td>1129</td>
</tr>
<tr>
<td>Nested cWW aligned</td>
<td>431</td>
<td>431</td>
<td>428</td>
<td>422</td>
</tr>
<tr>
<td>Nested non-cWW aligned</td>
<td>88</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>Non-nested cWW aligned</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Non-nested non-cWW aligned</td>
<td>75</td>
<td>75</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Stacking interactions aligned</td>
<td>1375</td>
<td>1169</td>
<td>1372</td>
<td>1375</td>
</tr>
<tr>
<td>Base-phosphate interactions aligned</td>
<td>77</td>
<td>32</td>
<td>76</td>
<td>78</td>
</tr>
<tr>
<td>Correspondences agreeing with Composite</td>
<td>1414</td>
<td>1208</td>
<td>1362</td>
<td>1356</td>
</tr>
<tr>
<td>Correspondences missing, compared to Composite</td>
<td>0</td>
<td>206</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td>Correspondences extra, compared to Composite</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>126</td>
</tr>
<tr>
<td>Mean local neighborhood discrepancy</td>
<td>0.261</td>
<td>0.251</td>
<td>0.247</td>
<td>0.391</td>
</tr>
<tr>
<td>Median local neighborhood discrepancy</td>
<td>0.197</td>
<td>0.198</td>
<td>0.193</td>
<td>0.208</td>
</tr>
<tr>
<td>Global discrepancy of all aligned nucleotides</td>
<td>0.081</td>
<td>0.087</td>
<td>0.099</td>
<td>0.118</td>
</tr>
</tbody>
</table>

As mentioned above, R3D Align aligns nearly as many nucleotides as the composite alignment. Sequence alignment methods like NW and NAST and the crystallographer alignment suffer from the problem that they basically need to align things even if they don’t correspond very well in 3D. They have little reason not to align nucleotides that are sandwiched between aligned nucleotides. Thus, they tend to align more nucleotides than the other methods. It is also evident that NW tends to find the alignment with a greater number of identical base correspondences than the other methods.

Because of the high degree of structural conservation between these two structures, a
good performance characteristic is the number of basepairs aligned by each method. If an alignment has many of the basepairs aligned correctly, it is usually indicative of corresponding regions of the two structures being aligned properly. The total number of nucleotides aligned could be another performance characteristic. Also interesting to compare are the number of pairs of nucleotides forming stacking or base-phosphate interactions that are aligned. These are summary statistics, so one alignment method aligning more interactions than another is not necessarily a proof that one method is better than another, but it does help us compare.

In Figure 5.8 the aligned basepairs are broken down according to whether they are nested cis Watson-Crick/Watson-Crick basepairs (as in helices), non-nested Watson-Crick basepairs (as in pseudoknots), nested non-Watson-Crick (as in internal loops), or non-nested non-Watson-Crick basepairs (as in A-minor and other tertiary interactions). In these four categories, R3D Align performs nearly as well as it can, and outperforms the other alignment techniques.

Also given in the figure are the stacking and base-phosphate interactions that are preserved by each method. The backbone-based methods do better on stacking interactions, which are often between adjacent nucleotides, than on basepairs. Base-phosphate interactions, which were described in [73], are most often found in highly conserved regions, which explains why the Needleman-Wunsch alignment preserved so many of them.

While these statistics are indicators of a good alignment, we need more comprehensive measures of how well the aligned nucleotides correspond structurally. For each method we superimposed the entire sets of aligned nucleotides and calculated the geometric discrepancy between them. However, this measure does not give much indication of how the two structures correspond locally as errors are spread out across all aligned nucleotides.

To measure how the aligned structures correspond locally, for each nucleotide in T.th., we found the nearest four nucleotides in T.th. which were aligned to something in E.c., and then found the geometric discrepancy between the five nucleotides from T.th. and the corresponding five nucleotides from E.c. The mean and median of these discrepancies is given
in the table. This diagnostic is somewhat independent of the method used by R3D Align, as it uses 5-nucleotide neighborhoods rather than 4, and as it may superimpose nucleotides that are never part of the same 4-nucleotide neighborhood. In particular, it will give large error readings in places where nucleotides that are nearby in *T.th.* are aligned to nucleotides that are farther apart in *E.c.*, which happens at the boundaries of element expansion regions.

Figure 5.8 shows that the R3D Align alignment outperforms the other automated 3D-to-3D alignment methods according to the most relevant metrics. R3D Align correctly aligns more nucleotides engaged in basepair, base-stacking, and base-phosphate interactions than do SARA, ARTS, and DIAL, and approaches the correct counts of the manually derived alignment. R3D Align also achieves the lowest mean local discrepancy value. The fact that ARTS achieves a lower median local discrepancy value than R3D Align is not necessarily surprising as the main problem with the global alignments produced by ARTS is not nucleotides that are misaligned, but nucleotides that should be aligned but are not. So what ARTS does align, it aligns well, but it aligns significantly fewer nucleotides than the other methods.

Using the global geometric discrepancy value alone would not allow one to accurately evaluate the alignment methods. ARTS achieves the lowest value in this category because it is a method based on one rigid superposition and then selects only nucleotides which are very close together, while the global geometric discrepancy value is also based on one rigid superposition. Among the three methods that use reduced representations of structures and then apply a quadratic time algorithm, SARA outperforms both DIAL and PARTS (results not displayed) in all categories.

### 5.4 “Bar Diagram” Display and Evaluation Tool

Given an alignment of two RNA structures, it is useful to know which aligned nucleotides superimpose well and which do not. Calculating discrepancies of local neighborhoods as
described above and then analyzing that list of numbers is one method of achieving this. However, sorting through and analyzing a long list of numbers takes time. Analyzing the colored interaction list described above is another way. However, for large structures that list can also get quite long and it is desirable to identify certain regions with just a quick glance at a single diagram. Therefore, we have developed a new graphical display that not only shows the aligned nucleotides, but it also indicates the quality of local aligned regions.

Given an alignment, we create a “bar diagram” in which the nucleotides of the first structure are labeled uniformly across the upper horizontal line and the nucleotides of the second structure are labeled across the lower horizontal line. Straight line segments connect the corresponding nucleotides in the two structures (as in Figure 1.8). The lines are colored to indicate the structural similarity of the local neighborhoods of the nucleotides connected to the line. As above, for each aligned nucleotide of the first structure, we then find its four nearest neighbors that have a corresponidences and then superimpose those five nucleotides onto the five corresponding nucleotides. The geometric discrepancy of that superimposition is then calculated and used to determine the color of the connecting line. Blues to reds are used to uniformly represent the geometric discrepancy values from 0 to 1.2. Any line corresponding to a geometric discrepancy over 1.2 is colored red.

In Figure 5.9, we display the bar diagrams for each of the different 16S alignments. A color bar is also given that indicates exactly how the colors of the lines correspond with the geometric discrepancies. Figure 5.9 makes it clear that R3D Align performs nearly the same as the manually-derived composite alignment. It should be noted that one possible reason red lines may appear in the composite alignment is because of errors in the crystal structures, as described in [63]. Briefly, some nucleotides were modeled wrong, the crystallographers agreed, and the pairs were aligned manually. SARA does very well, but does have some poorly aligned localized regions, as indicated by the sections with many red lines. ARTS does well where it make correspondences, but it is clear from the figure that it leaves out some large regions entirely. DIAL does well in some regions, but poorly in others. The Needleman-
Figure 5.9: 16S rRNA alignments produced by the different methods with the nucleotides of \textit{T. th.} (1j5e) listed along the top of each bar and the nucleotides of \textit{E. c.} (2avy) listed along the bottom. For each nucleotide in 1j5e with a correspondence, it and the nearest 4 nucleotides in 1j5e with a correspondence are superimposed with the corresponding 5 nucleotides of 2avy and the line is colored according to the geometric discrepancy, with values as indicated by the color bar.
Wunsch alignment is good in some places, poor in others. This is no surprise; any Needleman-Wunsch algorithm requires sequence conservation to perform well, and some regions are more highly conserved than others. The NAST alignment, which is a method optimized for 16S rRNA alignment, generally does well along the whole length of the sequence, performing similarly to the crystallographer alignment.
CHAPTER 6

Applications of R3D Align

6.1 Building Consensus Interaction Lists

Several programs are available that take a 3D RNA crystal structure and list which nucleotides of the structure are forming basepairs with each other. FR3D is one such program. However, classification of base interactions necessarily results in a dichotomy - two bases interact or they don’t. Near interaction categories can be used to soften that, but they do not eliminate the problem that a pair of bases might happen to have been caught just outside the classification limits in the crystal structure, or were modeled wrong, or the electron density was just not clear enough. We must remember that RNA crystal structures are snapshots of dynamic molecules, which are then fit by humans. This dichotomy sometimes creates problems in applications which rely on basepair classification systems.

Because R3D Align produces alignments that are accurate at the individual nucleotide level, it can be used to help alleviate this problem. In Figure 6.1 we see a region where one or the other structure has several near basepairs, but the other has good basepairs. Also, the geometric discrepancy between these two regions is pretty good. We can then infer that the near basepairs are true basepairs. This could be used in a program like FR3D, to assist with the annotation of basepairs. It could be useful for programs which search the
annotations rather than the 3D structures. It can be useful to crystallographers when they are trying to decide how to model a basepair. Finally, it will be useful for programs like JAR3D (manuscript in preparation), which use 3D structure to make alignments.

<table>
<thead>
<tr>
<th>T.Th. (1JSE)</th>
<th>E.c. (2AVY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G685 tSH</td>
<td>G685 tSH</td>
</tr>
<tr>
<td>U686 tWH</td>
<td>U686 tWH</td>
</tr>
<tr>
<td>A687 tSS</td>
<td>A687 ntSS</td>
</tr>
<tr>
<td>A687 tHS</td>
<td>A687 tHS</td>
</tr>
<tr>
<td>G688 cWW</td>
<td>G688 cWW</td>
</tr>
<tr>
<td>G688 tSS</td>
<td>A687 ntSS</td>
</tr>
<tr>
<td>C689 cWW</td>
<td>C689 cWW</td>
</tr>
<tr>
<td>U690 tSH</td>
<td>G690 ntSH</td>
</tr>
<tr>
<td>A691 tSH</td>
<td>G691 tSH</td>
</tr>
<tr>
<td>U692 ---</td>
<td>U692 ---</td>
</tr>
<tr>
<td>A693 ---</td>
<td>A693 ---</td>
</tr>
<tr>
<td>A694 ---</td>
<td>A694 ---</td>
</tr>
<tr>
<td>A695 ---</td>
<td>A695 ---</td>
</tr>
<tr>
<td>A696 tHS</td>
<td>A696 tHS</td>
</tr>
<tr>
<td>A696 tSS</td>
<td>A696 tSS</td>
</tr>
<tr>
<td>A696 ncSS</td>
<td>A696 cSS</td>
</tr>
<tr>
<td>U697 tHS</td>
<td>U697 ntHS</td>
</tr>
<tr>
<td>U697 ncSS</td>
<td>U697 cSS</td>
</tr>
<tr>
<td>G698 cWW</td>
<td>G698 cWW</td>
</tr>
<tr>
<td>G699 cWW</td>
<td>G699 cWW</td>
</tr>
<tr>
<td>G700 tSS</td>
<td>A687 ntSS</td>
</tr>
<tr>
<td>C701 ---</td>
<td>U701 ---</td>
</tr>
<tr>
<td>A702 ---</td>
<td>A702 ---</td>
</tr>
<tr>
<td>G703 tSH</td>
<td>A687 tSH</td>
</tr>
<tr>
<td>A704 tWH</td>
<td>A704 tWH</td>
</tr>
<tr>
<td>A704 tSS</td>
<td>A704 ntSS</td>
</tr>
<tr>
<td>U705 tHS</td>
<td>G685 tHS</td>
</tr>
<tr>
<td>A706 cWW</td>
<td>A706 cWW</td>
</tr>
<tr>
<td>C707 cWW</td>
<td>G683 cWW</td>
</tr>
</tbody>
</table>

Figure 6.1: A region of the 16S alignment that shows several pairs of nucleotides forming basepairs that are aligned with nucleotides forming near basepairs.

JAR3D is a 1D-to-3D alignment program that is based on a SCFG/MRF (Stochastic Context-Free Grammar/Markov Random Field) model built from a 3D crystal structure. JAR3D uses the 3D structure of an entire RNA molecule or of an individual motif to construct an SCFG/MRF model for likely sequence variability for that RNA or individual motif. The model incorporates the interaction information contained within the 3D crystal structure. For example, if JAR3D were to perform an alignment of 16S rRNA molecules, a model would
first be created using a 16S rRNA whose 3D information is available, such as 2avy. The 
interactions of the bases of 2avy would be analyzed and classified (by FR3D) and then used 
to create a model for all other 16S rRNA. FR3D analysis of 2avy concludes that the molecule 
consists 680 basepairs and so the SCFG model would consequently model the 16S rRNA as 
having 680 basepairs. However, as described above, classification of base interactions results 
in a dichotomy and depends on a variety of factors besides the actual 3D structure.

Because the SCFG/MRF model is to represent all 16S rRNA, it is important to be able 
to distinguish bases that actually are forming basepairs versus those that are not. As was 
mentioned above, in the analysis of an RNA molecule, many bases are not considered to be 
forming basepairs, but instead are identified as nearly forming basepairs. Some of these near-
basepairs may actually be forming real basepairs or the corresponding bases in the majority 
of other 16S rRNA might be. In either case, it would be useful to identify those places in 
2avy where the SCFG model should model an interaction as a true basepair instead of a 
near-basepair. So given a list of near-basepairs, how do we determine which should actually 
be modeled as forming real basepairs? In the analysis of the structure 2avy by FR3D, 
1,101 base combinations have been classified as forming near-basepairs. Of these 1,101 near 
basepairs, which should actually be modeled as true basepairs in the SCFG model? In some 
cases, the near basepairs are actually true basepairs while other times this is not true.

We can use the 3D-to-3D alignment to help solve this problem since we have the 3D 
information for a second 16S rRNA molecule, 1j5e. We can use the results of our 3D-to-3D 
alignment of 2avy and 1j5e to create a consensus model for 16S rRNA.

Using the 3D-to-3D alignment of 2avy and 1j5e obtained by R3D Align, it is observed that 
40 of the 1,101 near basepairs of 2avy are aligned with nucleotides in 1j5e that are forming 
true basepairs of the same geometric family. If we infer that 2avy has a true basepair at 
each of these positions, it would increase the number of identified basepairs. Since FR3D 
classifies 2avy as having 680 real basepairs, identifying another 40 increases the total by 
only 6%, which significantly adds to the amount of information that is incorporated into
the SCFG/MRF model.

If a consensus model is to represent all 16S rRNA, one may argue that the above approach really does not create a consensus model because it incorporates some basepairs that are truly only found in one structure. Instead of creating a consensus model, we have created an interaction list for one structure that is augmented by additional basepairs found in the other structure.

It may be more appropriate to create a consensus model by including only those basepairs truly found in both structures. Once again, our 3D-to-3D alignment can be used to create this model. While FR3D classifies 2avy as having 680 basepairs and 1j5e as having 690, our 3D-to-3D alignment shows that 598 of these basepairs are conserved across the two structures. For these 598 basepairs, we have a greater confidence that they are representative of all 16S bacterial rRNA since they are found in the two known 3D crystal structures of different organisms.

What we have done is use the information provided in the 3D-to-3D alignment to build a consensus basepairing interaction list. In a similar way, we can include those stacking and base-phosphate interactions which are aligned to produce an even more comprehensive consensus interaction list. Such a list provides much more information than a consensus secondary structure, which only contains information regarding the Watson-Crick basepairs found in helices. Because this consensus interaction list will contain all of the information contained within a consensus secondary structure and more, it can be useful in any situation where a consensus secondary structure has been shown to be useful.

6.2 Improving Existing Alignments

Several of the other 3D to 3D alignment programs currently available, such as DIAL, SARA, and PARTS, are able to execute an alignment in quadratic time. These programs are often able to compute an alignment of two structures more quickly than R3D Align. However, in
order to achieve a faster runtime, these programs often sacrifice the quality of the final alignment. As noted in [20], due to the complexity of RNA 3D structural alignment, one cannot hope that a quadratic time algorithm such as DIAL would be highly accurate. Because of this, they believe the primary application of quadratic time algorithms such as DIAL will be to serve as primary alignment mechanisms which are then followed by a very accurate, but computationally intensive alignment algorithm.

Because R3D Align is an accurate 3D to 3D alignment program with the capability of accepting seed alignments, we believe that a beneficial use of the program will be to first apply any of the other 3D-to-3D methods and then use the alignment produced by that program as a seed alignment for R3D Align. Because we believe the alignments produced by the other programs are at least somewhat accurate, we can use a smaller bandwidth when running R3D Align. This should have the effect of producing an alignment that is more accurate than that produced by the initial method but one that is achieved in less time than executing R3D Align alone.

As an example, we consider the 3D-to-3D alignment of the RNA molecules with pdb identification numbers, 1u6b (chain B) and 1y0q (chain A) [1, 26]. 1u6b is the crystal structure of a self-splicing group I intron with both exons and 1y0q is the crystal structure of an active group I ribozyme-product complex.

To get an idea about the structure of these two molecules, we take a look at the pairwise nucleotide interactions within each molecule. An even more concise method of displaying all the interactions for the structures than using a spreadsheet is to create a circular diagram in which nucleotides are ordered around a circle and lines are drawn between nucleotides having a either a basepairing, stacking, or base-phosphate interaction. In Figures 6.2 and 6.3 the circular diagrams for 1u6b and 1y0q are given.

The circular diagrams for the two structures show that the two structures have many similar interactions, but also quite a few differences. While both have approximately the same number of interactions, several of the helical regions differ in length and location
Figure 6.2: 1u6b:B interactions
Dark blue chords indicate the 50 nested Watson-Crick basepairs
Cyan chords indicate the 6 nested non-Watson-Crick basepairs
Red chords indicate the 9 non-nested Watson-Crick basepairs
Green chords indicate the 18 non-nested non-Watson-Crick basepairs
Yellow chords indicate the 213 stacking interactions
Magenta chords indicate the 5 base-phosphate interactions

Figure 6.3: 1y0q:A interactions
Dark blue chords indicate the 55 nested Watson-Crick basepairs
Cyan chords indicate the 8 nested non-Watson-Crick basepairs
Red chords indicate the 9 non-nested Watson-Crick basepairs
Green chords indicate the 13 non-nested non-Watson-Crick basepairs
Yellow chords indicate the 250 stacking interactions
Magenta chords indicate the 12 base-phosphate interactions
within the structure. So we expect that many interactions will not be aligned in the proper alignment. However, a 3D-to-3D alignment will be used to investigate if there is a common core set of interactions that are conserved across the two structures.

We first analyze the improvement in the accuracy of the alignment of the two structures when R3D Align is applied to the alignment produced by the cubic time algorithm ARTS. The alignments will once again be evaluated based on the number of aligned basepairs in addition to using the measures involving geometric discrepancy calculations. When aligning the 195 nucleotides of 1u6b with the 229 nucleotides of 1y0q, ARTS aligns 116 nucleotides. When R3D Align is applied to this alignment, the result is an alignment featuring 124 correspondences, including 25% more aligned nested cWW basepairs (30 compared to 24), and 50% more aligned non-nested non-cWW basepairs (9 vs. 6). Also, the geometric discrepancy diagnostics indicate the alignment is much improved after R3D Align. The mean geometric discrepancy of the 5 nucleotide local neighborhood superpositions is 0.585 for the ARTS alignment, but improves to 0.331 after the application of R3D Align, nearly a 43.4% decrease. If we calculate the geometric discrepancy between the 116 aligned nucleotides of 1u6b with the corresponding 116 nucleotides of 1y0q, the geometric discrepancy is 0.249; however the global discrepancy between the aligned sets of nucleotides produced after R3D Align is run is 0.170, a 31.7% decrease.

<table>
<thead>
<tr>
<th>Number of nucleotides aligned</th>
<th>ARTS</th>
<th>ARTS/R3D</th>
<th>SARA</th>
<th>SARA/R3D</th>
<th>PARTS</th>
<th>PARTS/R3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested cWW aligned</td>
<td>116</td>
<td>124</td>
<td>147</td>
<td>125</td>
<td>182</td>
<td>125</td>
</tr>
<tr>
<td>Nested non-cWW aligned</td>
<td>24</td>
<td>30</td>
<td>28</td>
<td>30</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Non-nested cWW aligned</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Non-nested non-cWW aligned</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Stacking interactions aligned</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Base-phosphate interactions aligned</td>
<td>99</td>
<td>120</td>
<td>120</td>
<td>121</td>
<td>88</td>
<td>121</td>
</tr>
<tr>
<td>Mean local neighborhood discrepancy</td>
<td>0.585</td>
<td>0.331</td>
<td>0.787</td>
<td>0.233</td>
<td>4.754</td>
<td>0.539</td>
</tr>
<tr>
<td>Median local neighborhood discrepancy</td>
<td>0.354</td>
<td>0.318</td>
<td>0.436</td>
<td>0.239</td>
<td>4.866</td>
<td>0.299</td>
</tr>
<tr>
<td>Global discrepancy of all aligned nucleotides</td>
<td>0.249</td>
<td>0.170</td>
<td>0.329</td>
<td>0.169</td>
<td>1.904</td>
<td>0.789</td>
</tr>
<tr>
<td>Number of exact base matches</td>
<td>57</td>
<td>68</td>
<td>73</td>
<td>69</td>
<td>30</td>
<td>66</td>
</tr>
</tbody>
</table>

Figure 6.4: Summary statistics for intron alignments by different methods
It should be noted that the improved alignment produced by R3D Align only required 8 seconds to produce. The parameters used were $d = 0.5$, $p = 7$, and $\beta = 10$.

Because applying R3D Align to the existing ARTS alignment produced an alignment that increased the number of correspondences and the number of aligned basepairs, all while lowering the geometric discrepancy calculations, we have very strong evidence that R3D Align was effective in improving the ARTS alignment. The results for the alignments are given in Figure 6.4.

To get a better idea about the conserved nucleotides and the conserved interactions they are involved in, we have created a circular diagram that displays just those interactions in 1u6b that are aligned with a corresponding interaction in 1y0q, which is given in Figure 6.5.

![Figure 6.5: Interactions in 1u6b:B that are aligned with a corresponding interaction in 1y0q:A](image)

Figure 6.5 also gives the results for the improvement of the alignment produced by the quadratic time algorithm SARA. SARA produces an alignment of 147 correspondences compared to only 116 for the ARTS alignment. SARA aligns more basepairs than ARTS does but does have higher geometric discrepancy values, as might be expected since it aligns more nucleotides.
In Figure 6.4 we see that the alignment produced after applying R3D Align to the SARA alignment aligns fewer nucleotides but actually aligns more basepairs. For example, the improved alignment features 30 nested cWW basepairs and 9 non-nested non-cWW basepairs that are aligned, while the SARA alignment alone only has 28 and 9, respectively. Once again, we see that applying R3D Align lowers the geometric discrepancy values. The mean geometric discrepancy of local neighborhoods dropped a substantial 57.7% from 0.787 to 0.333. The global discrepancy value also dropped, going from 0.329 for the SARA alignment to 0.169 for the alignment produced by SARA and then improved by R3D Align, a drop of nearly 50%.

Again, the improved SARA alignment only required 8 seconds to produce. The parameters used were the same as those used to improve the ARTS alignment, \( d = 0.5, p = 7, \) and \( \beta = 10. \)

As a final illustration, R3D Align was used to improve the alignment produced by PARTS in order to show that it is able to improve alignments that are much more inaccurate. As the figure shows, PARTS did not align any basepairs at all. However, the alignment produced after using R3D Align again aligns many more basepairs and reduces the geometric discrepancy values significantly. It should be noted that a larger bandwidth had to be used (\( \beta = 30 \)) because the PARTS alignment was so inaccurate.

These improvement of these three alignments exhibits the versatility and proficiency of using R3D Align to increase the accuracy of an existing alignment. In the case of ARTS, R3D Align was able to add additional correspondences to improve the alignment while for SARA, R3D Align was able to remove incorrect correspondences to improve the alignment.

### 6.3 RNase P Alignment

R3D Align was used to align the specificity (S) domains of the two types of bacterial Ribonuclease (RNase) P. The two major distinct types of RNase P, A and B, are classified on
the basis of the sequence of the RNA component. While these two types of RNase P have identical function, there are notable differences in their secondary structures, suggesting alternative tertiary folds as well [39]. As reported in [39], although the specificity domain of an A-type RNase P differs in secondary and tertiary structure from the specificity domain of a B-type RNase P, the S domains have similar 3D structure. We will see that R3D Align identifies this similarity.

A structural alignment of an A-type S domain contained in PDB file 1u9s [39] and a B-type S domain contained in 1nbs:B [40] was obtained. The interaction of 1u9s and 1nbs are displayed in Figures 6.6 and 6.7

![Figure 6.6: 1u9s interactions](image)

Dark blue chords indicate the 51 nested Watson-Crick basepairs  
Cyan chords indicate the 0 nested non-Watson-Crick basepairs  
Red chords indicate the 12 non-nested Watson-Crick basepairs  
Green chords indicate the 8 non-nested non-Watson-Crick basepairs  
Yellow chords indicate the 174 stacking interactions  
Magenta chords indicate the 17 base-phosphate interactions

This alignment provides another good test case for an automated procedure such as R3D Align as the structures have much more notable differences than the 16S rRNA structures do. The R3D Align alignment was found to be in agreement with the findings described in [39]. That is, elements that were described as being present in both structures were aligned by
Figure 6.7: 1nbs:B interactions
Dark blue chords indicate the 42 nested Watson-Crick basepairs
Cyan chords indicate the 0 nested non-Watson-Crick basepairs
Red chords indicate the 19 non-nested Watson-Crick basepairs
Green chords indicate the 6 non-nested non-Watson-Crick basepairs
Yellow chords indicate the 162 stacking interactions
Magenta chords indicate the 12 base-phosphate interactions

R3D Align, while all elements that exist in only one of the structures were not assigned any correspondences. To obtain the alignment, we used $d=0.3$, $p=12$, $\beta=60$, and the full branch and bound maximum clique method was used (Example 1 in the Section 8.7). Figure 6.8 displays the interactions in 1u9s that are aligned with a matching interaction in 1nbs:B. The full alignment spreadsheet is given in Appendix C. The nucleotides have also been labeled according to which elements of the structure they belong. The same numbering scheme was used as in [39] for ease of comparison.

6.4 Alignment of 23S rRNA Structures

As discussed in Section 1.7.2, the 23S rRNA molecule has been crystallized in four organisms: *Haloarcula marismortui* (H.m.), *Thermus thermophilus* (T.th.), *Escherichia coli* (E.c.), and *Deinococcus radiodurans* (D.r.). H.m. is an archaeon, the rest are bacteria.
Alignments of the 23S molecules were obtained using R3D Align. PDB file 2aw4 contains a crystal structure of E. c. PDB file 2j01 is a T. th. crystal structure, 2zjr is D. r. and 1s72 is H. m. Three pairwise alignments were obtained using R3D Align. The E. c. crystal structure 2aw4 was first aligned with 2j01, then 2zjr, and lastly 1s72.

The same R3D Align parameters were used as in the case of the 16S alignment presented in Chapter 5. To briefly summarize, an iterative approach was used in which $\beta = 30$ and $p = 3$ for the first iteration. The alignment produced in the first iterations was used as the seed alignment in the second iteration, which used $\beta = 10$ and $p = 9$. For both iterations, the geometric discrepancy cutoff $d = 0.5$ was used.

In Figure 6.9(a), the R3D Align alignment of 2aw4 and 2j01 is displayed using the corresponding bar diagrams. For comparison purposes, the ARTS alignment of 2aw4 and 2j01 was also found and is presented in Figure 6.9(b). Figures 6.10(a) and 6.10(b) display the alignments of 2aw4 and 2zjr as produced by R3D Align and ARTS, respectively. Finally, in Figure 6.11 the alignments of 2aw4 and 1s72 are displayed.
Figure 6.9: *T.*th. (2j01) and *E.*c. (2aw4) 23S Alignments by R3D Align and ARTS. The nucleotides of 2j01 are listed on the left of each bar, and those of 2aw4 are on the right.
Figure 6.10: *D. r.* (2zjr) and *E. c.* (2aw4) 23S Alignments by R3D Align and ARTS. The nucleotides of 2zjr are listed on the left of each bar, and those of 2aw4 are on the right.
Figure 6.11: *H. m. (1s72) and E. c. (2aw4) 23S Alignments by R3D Align and ARTS. The nucleotides of 2j01 are listed on the left of each bar, and those of 2aw4 are on the right.
CHAPTER 7

Detecting Conformational Changes Among Structures

7.1 Introduction/Motivation

In this chapter, we describe methods to detect local conformational changes between two 3D structures of the same RNA. These structures may be crystallizations from the same molecule from the same organism, or they may be crystallizations of molecules from different, yet related, organisms. In our approach, we study the variability that exists among the translation and rotation that are needed to superimpose local neighborhoods after global superposition. Each translation is represented by a three-dimensional vector as is each rotation. Thus, we investigate the variability that exists among sets of multivariate data.

As discussed in Section 1.2 over 1700 RNA 3D structures have been deposited within the Protein Data Bank (PDB) and its partner, the Nucleic Acid Data Bank (NDB). Most of the 3D structures deposited have been determined to atomic precision or near-atomic precision by a process known as x-ray crystallography, but some have been determined by NMR (nuclear magnetic resonance) or electron microscopy. In every case, one is attempting to determine the atomic coordinates of hundreds or thousands of atoms in a particular type
of molecule. But molecules themselves are dynamic entities; the positions of atoms fluctuate due to thermal noise about their equilibrium positions and because the molecule gets hit often by water molecules in the solution. Moreover, every crystal will contain many copies of the same molecule, and these copies will all be in slightly different geometric conformations.

Crystallography reduces the variability somewhat, because the molecules align to form a periodic crystal structure (with the unit cell as the repetitive element), and so are locked into place, but they can still jiggle about internally. Some parts of the molecule might be loose enough to move about quite a bit, which will make it more difficult to determine the locations of the atoms in these parts.

In x-ray crystallography, one diffracts x-rays from the crystal, measures the x-ray intensity and scattering as a function of scattering angle, and infers the density of electrons at spatial points throughout a single copy of the molecule by 3D Fourier transformation. This is inherently imprecise due to variability in detection of x-rays and the limited resolution of the detectors. Then a crystallographer will attempt to fit atoms and RNA nucleotides to the electron density. To the extent that parts of the molecule are mobile, the electron density may be too diffuse to fit.

Thus, 3D structures as determined by x-ray crystallography are subject to four kinds of variability. There is variability in the conformation of the molecule itself, as well as thermal fluctuations which make parts of the electron density difficult to fit, detection errors in determining the electron density, and also human error in fitting atoms to the electron density.

Each structure can result in a 3D crystal structure that is unique and differs from the determination of other crystal structures. We have seen above that 1j5e and 2avy are crystal structures of the 16S ribosomal RNA molecule from two different organisms, *T. th.* and *E. coli.* The molecules themselves are different, so it is no surprise that the 3D structures differ, but they are homologous, so it is also no surprise that in many places, the 3D structures are quite similar. But not all of the crystal structures in the PDB represent distinct molecules.
The same RNA molecules have been crystallized by multiple research groups or have been crystallized multiple times by the same research group (seeking improved versions of the crystal structure or examining the effects of binding drugs or other molecules). Also, some molecules have been crystallized in complex with functional ligands such as mRNA or tRNA (in the case of ribosomes). Other times small molecules like antibiotics can be diffused into existing crystals to bind to specific sites, and then the same crystal can be subjected to x-ray crystallography again.

It is necessary to develop methods capable of detecting and assessing the variability among these structures since there are a variety of instances in which it is useful to detect the conformational changes between two structures. For example, research groups may want to analyze which regions of two crystal structures deviate from one another to determine any errors that may have been made during the modeling process. Or it may be beneficial to learn what conformational changes take place within the 3D structure of a molecule when an antibiotic or other ligand is bound. Some crystallizations catch the two subunits of the ribosome in different stages of ratcheting [72]. In all cases, we need to determine what variability between RNA 3D structures is typical, and what variability is unusual.

Fortunately, we have many instances of duplicate structures, and so are able to characterize the variability by comparing two (or more) 3D structures. We note that when we compare two different crystal structures of the same molecule from the same organism, we know exactly what nucleotides correspond between the two structures; thus, no alignment of the structures is required.

The goal of this chapter is to characterize the variability in local neighborhood locations and orientations in RNA 3D crystal structures, and between RNA crystal structures of different organisms.

As discussed in Section 5.1, the classical tool for comparing two different models of a molecule is the average root-mean-square deviations (RMSD) after optimal superposition of the two structures. However, while RMSDs do capture the general shapes of the molecules,
they do not provide any information regarding the variability that exists between the structures at a more local level. Finding local conformational changes can help improve the modeling process by determining local areas of deviation.

7.2 Method

We analyze the variability that exists in the deviations among the optimal superpositions of local neighborhoods, in comparison with the optimal superposition of the global structures. An optimal superposition of nucleotides $c_1, \ldots, c_n$ onto $l_1, \ldots, l_n$, is one that consists of a translation vector $t$ and a rotation matrix $R$ that minimize the squared error,

$$L^2 = \min_R \min_t \|l_i - R(c_i - t)\|^2.$$  \hspace{1cm} (7.1)

First, the optimal global rotation matrix $R^*$ and translation vector $T^*$ are obtained using the technique described in [35]. These are used to superimpose the two structures globally. Then for each nucleotide $i$ in the first structure, its five-nucleotide local neighborhood is found (as in Sections 5.3 and 5.4), and then the translation vector $t_i$ and rotation matrix $R_i$ of the optimal superposition onto the corresponding five nucleotides in the other structure are found.

We note that careful attention must be paid to the calculation of each $R_i$ and $t_i$ so that each neighborhood’s rotation and translation corresponds to the same reference orientation. The calculation of $t_i$ and $R_i$ is now described in more detail.

Suppose nucleotides $\alpha_1, \alpha_2, \alpha_3, \alpha_4,$ and $\alpha_5$ are to be superimposed onto the neighborhood consisting of nucleotides $\beta_1, \beta_2, \beta_3, \beta_4,$ and $\beta_5$. We let $\overline{\alpha}$ be the center of mass for the neighborhood in the first structure and $\overline{\beta}$ be the center of mass for the neighborhood in the second structure. That is,

$$\overline{\alpha} = \frac{\sum_{j=1}^{5} \alpha_j}{5} \quad \text{and} \quad \overline{\beta} = \frac{\sum_{j=1}^{5} \beta_j}{5}.$$
Then we have

\[ t_i = \beta - \alpha \]

and

\[ R_i \text{ determined as in [35].} \]

We note that while a rotation is typically represented as a $3 \times 3$ matrix, it can be expressed as a 3-dimensional vector using the axis of rotation and making its length equal to the angle of rotation in degrees. For our purposes, we will represent each $R_i$ in this vector form.

It is important to point out once again that the translation and rotation vectors only indicate deviations from the global superposition. Therefore, the translation vectors $t_i$ should be small and centered around 0, since they only indicate deviations from the global superposition. The rotation vectors should also be short (small angle of rotation).

### 7.3 Examples Using Translations and Rotations to Detect Conformational Changes

As an example we compare the crystal structures found in the PDB files 2uub and 2uuc. We select these two structures because we expect them to be very similar since they are both crystallizations of the same organism with the same antibiotic bound. Also, the crystal structures were determined by the same research group and were deposited in the PDB only one week apart. Both files contain crystal structures of *T.th.* 16S RNA bound to an mRNA with a codon in the A-site and the antibiotic paromomycin. 2uub is bound with a GUU-codon, 2uuc is bound with a GUA-codon [68].

First, the translation vectors are analyzed. For each nucleotide $i$ in the structure, the translation vector $t_i$ is calculated as described above. Figure 7.1 displays a scatterplot of the translation vectors. It can be seen that the three-dimensional data points are centered at zero. What is of interest is the variability of the points about the mean as that provides a measure
of the similarity (or dissimilarity) of the two structures. The more a point deviates from the mean, the greater the conformational difference among the neighborhood of nucleotides represented by that point.

Figure 7.1: View of 3D scatterplot of translation vectors for files 2uub and 2uuc. Units are Angstroms (1Å=10^{-10}m). Points are centered about the origin and colored according to the corresponding nucleotide’s distance from the center of the structure. The colors range from blues to reds according to the color bar shown in Figure 7.2. Extreme data points are labeled using the method outlined in Section 7.4 using a cutoff value of 25 for $D_i^2$.

For example, in Figure 7.1 nucleotide 1257 certainly appears to be an extreme data point as it is set apart from the rest of the data points. Figure 7.3 shows a portion of the global superposition showing nucleotide 1257 and its local neighborhood. Nucleotides 1257 from the two structures do not superimpose well, which makes it clear why $t_{1257}$ is farther removed from the origin.
Figure 7.2: The data points in the 3D scatterplots of this chapter are colored according to the distance to the center of the molecule, as given by this color bar. The color bar ranges from 0 Ångstroms to 130 Ångstroms since nucleotides of 16s T.th. rRNA molecules typically are no more than 130 Ångstroms from the center of the structure.

Figure 7.3: Portion of the global superposition showing the neighborhoods of nucleotide 1257 in 2uub and 2uuc. The neighborhood includes nucleotide 1257 along with the nearest four nucleotides in 3D space, which are nucleotides 1256, 1258, 1359, and 1360.

We are also able to detect larger regions of conformational difference by discovering clusters of data points that are scattered further from the mean. In Figure 7.1 there is a cluster of data points representing nucleotides 1026-1039 that have relatively extreme values. These nucleotides comprise a hairpin loop and basepairs in the helical region adjacent to the hairpin loop. Figure 7.4 displays the portion of the global superposition with these nucleotides along with nucleotides 1019-1025 and 1040-1045. It can be seen that nucleotides 1026-1039 do not superimpose as well in 3D space as the other nucleotides, which explains why the corresponding $t_i$ values are extreme.

The rotation vectors can be analyzed in a similar way as the translation vectors, although the interpretation is different. The rotation vectors provide additional information
Figure 7.4: Superposition of nucleotides 1019-1045 of 2uub and 2uuc. Nucleotide in the hairpin and near the hairpin do not superimpose as well as the others which illustrates why the corresponding points on the scatterplot are relatively extreme.

by indicating a different type of conformational change.

Figure 7.5 shows a view of the 3D scatterplot of the points representing the rotation deviations for the nucleotides of the crystal structures 1n33 and 1n34. Again, these structures were selected because they were produced by the same lab so we expect there to be many similarities. However, they have different complexes bound so we expect there to be some variability between the two structure. The 1n33 file contains the crystal structure of the T.th. 30S ribosomal subunit bound to codon and near-cognate transfer RNA anticodon stem-loop mismatched at the second codon position at the a site with paromomycin, and the 1n34 file contains the crystal structure of the T.th. 30S ribosomal subunit in the presence of codon and crystallographically disordered near-cognate transfer rna anticodon stem-loop mismatched at the first codon position [48].

The point labeled 1141 appears to be an extreme value. That is, there is a greater
variability between the optimal rotation of the neighborhood of nucleotide 1141 and the
global optimal rotation than the optimal rotations of other neighborhoods. Figure 7.6 shows
the superposition of the neighborhood taken from the global superposition.

Figure 7.5: View of 3D scatterplot of rotation vectors for files 1n33 and 1n34. Units are
degrees. Extreme data points are labeled using the method outlined in Section 7.4, using a
cutoff value of 25 for $D_i^2$.

Although we have just seen that extreme data points can be found by simply observing
the scatterplot, a less subjective, more formal statistical process is needed. In the next
section, we provide a description of the statistical methods that will be employed in order
to measure variability and detect extreme values within multivariate data of this type.
7.4 Detecting Outliers in Multivariate Data

We are essentially dealing with the problem of identifying outliers and other extreme data points within a sample of three-dimensional data. Detecting outliers is more difficult when dealing with multivariate data than with univariate data. This is because multivariate data cannot be ordered in the same way that a univariate sample can. Thus, the data cannot simply be ordered in such a way that the extreme values show up on either end. With multivariate data, an observation vector may have a large error in one of its components or smaller errors in several components. However, if the multivariate data is two-dimensional or three-dimensional, we may still perceive an observation to be particularly extreme when the data is displayed as a scatterplot as in Figures 7.1 and 7.5.

Despite no obvious ordering of the data being present, it is necessary to adopt some notion of ordering in order to determine extremeness. The most common method of ordering multivariate data is to reduce each multivariate \( \mathbf{x} \) to a scalar quantity \( D(\mathbf{x}) \) (often some type of distance measure), thereby creating a univariate data set from which extreme values can be detected \cite{4}.

We will represent each observation vector \( \mathbf{x}_i \) (which may either represent \( t_i \) or \( R_i \)) by its
standardized distance from the mean. We denote $D(x_i)$ as $D_i$. Then we have,

$$D_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}),$$

where $S$ is the sample variance-covariance matrix:

$$S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'.$$

The higher the value of $D_i^2$, the greater the distance of the point from the mean. Thus the task of finding conformational changes among structures amounts to finding nucleotides whose corresponding data points have larger values for $D_i^2$.

When the data is normally distributed, a formal outlier test exists that is based on $D_i^2$. This is because the Wilks’ statistic [69] can be expressed in terms of $D_{i(n)}^2$, where

$$D_{i(n)}^2 = \max_i D_i^2.$$  

The Wilks’ statistic is

$$w = \max_i \frac{|(n-2)S_{-i}|}{|(n-1)S|},$$

where $S_{-i}$ is the sample variance-covariance matrix of the data with the $i$th observation deleted. The Wilks’ statistic can also be written as

$$w = 1 - \frac{nD_{i(n)}^2}{(n-1)^2}.$$  

Thus a test for an outlier can be based on $D_{i(n)}^2$. Tables are available containing the upper 5% and 1% critical values of $D_{i(n)}^2$ [55]. However, in our case no normality assumption is made since we are not particularly interested in the formal declaration of outliers. We are more interested in which points are more extreme relative to the others, for which an ordering of the $D_i^2$ values is sufficient. It is left to a future study to determine whether the
data points $t_i$ and $R_i$ can be modeled as normally distributed.

In addition to discovering local regions of conformational change, we are also interested in determining the overall similarity of two structures. We know that the more similar two structures are, the closer the data points should be centered about the mean. Thus, if we study the overall variability of the data points, we can learn more about the overall similarity of the two structures. While this information is contained within the sample variance-covariance matrix $S$, it is desirable to have a single numerical value for the overall multivariate scatter. For that we use the total sample variance, which is simply the trace of $S$. Since we are working with three dimensional data, we have

$$\text{Total sample variance} = s_{11} + s_{22} + s_{33}$$

### 7.5 Same Molecule, No Ligands Bound, Same Lab Analysis

First, we want to understand the variability in local conformations between two RNA 3D crystal structures of the same molecule from the same organism and produced by the same research group.

As our first example, we use two crystal structures of the *T.thermophilus* 16S rRNA that were produced by the same research group. We use the crystal structure found in PDB file 1fjf and a more recent structure found in PDB file 1j5e. 1fjf was released in September of 2000 while 1j5e was released in April of 2002. Each of these crystal structures were determined with no ligands or other complexes bound to the molecule.

The scatterplots of the translation and rotation vectors are given in Figures 7.7 and 7.8 respectively. Points on the plot have been labeled with the corresponding nucleotide number if the value for $D_i^2$ is greater than 25. 25 was used as a cutoff value since that is approximately the critical value for the outlier test described above at the .05 level for a
sample of this size (with normal data). Figure 7.9(a) gives the largest 20 values of $D_i^2$ along with the corresponding nucleotide numbers.

As described above, we can use the scatterplots and $D_i^2$ values to find regions of conformational change. This would indicate which regions of the structure were modeled differently from one experiment to the next. However, we are also interested in the overall variability of the data points since that provides an indication of how similar the two structures are. The variance-covariance matrix for the translation data and rotation data ($S^T$ and $S^R$, respectively) are

$$
S^T = \begin{bmatrix}
0.0025 & -0.0026 & 0.0016 \\
-0.0026 & 0.0034 & -0.0020 \\
0.0016 & -0.0020 & 0.0014 \\
\end{bmatrix}
$$

$$
S^R = \begin{bmatrix}
1.6857 & 0.1350 & -1.1464 \\
0.1350 & 0.4843 & 0.2426 \\
-1.1464 & 0.2426 & 1.4301 \\
\end{bmatrix}
$$

Thus the total sample variance for the translation vectors is 0.0073, and the total sample variance for the rotation vectors is 3.6001. While it is difficult to analyze these numbers in and of themselves, we can see that there is very little variability in the data points, implying the two structures are very similar. These variances will also be useful in the following sections where structures are compared that have been determined under other situations than here where the same lab crystallized the same molecule from the same organism under the same conditions. For example in the next section, we again compare two structures of the same molecule from the same organism; however, they have been crystallized by different research labs. We expect there to be greater variability in the data points than in that case since there is additional variability in the modeling processes.
Figure 7.7: View of 3D scatterplot of translation vectors for files 1j5e and 1fjf. Units are Ångstroms. Points are colored according to the distance to the center of the structure (See Figure 7.2). Points are labeled if the corresponding value for $D_i^2$ is greater than 25 (the critical value of $D_i^2$ for normal data).
Figure 7.8: View of 3D scatterplot of rotation vectors for files 1j5e and 1ff. Units are degrees. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.
7.6 Same Molecule, No Ligands Bound, Different Labs

Analysis

We again consider two crystal structures of the same molecule of the same organism, but where the crystal structures have been determined independently by different research groups.

We again consider two crystal structures of 16S T.th., with no ligands or other complexes bound to the molecule, as before. We use the 1j5e crystal structure again, but instead of using 1fjf which was determined by the same group that determined 1j5e, we use 2zm6 which was solved more recently (2009) by an independent research lab [36].

The two variance-covariance matrices are

$$S^T = \begin{bmatrix} 0.3265 & 0.0298 & -0.0945 \\ 0.0298 & 0.3412 & -0.0473 \\ -0.0945 & -0.0473 & 0.2320 \end{bmatrix} \quad S^R = \begin{bmatrix} 11.1536 & -3.0655 & -2.5901 \\ -3.0655 & 9.2062 & 3.4793 \\ -2.5901 & 3.4793 & 17.4081 \end{bmatrix}$$
As expected, the variances of the translation and rotation vectors are larger in the case when the structures were crystallized by different groups than by the same group. The total sample variance for the translation and rotation data is 0.8997 and 37.7679, respectively, compared to 0.0073 and 3.6001 before.

The scatterplots are given in Figures 7.10 and 7.11. They also show that the data is much more spread out in this case. The extreme values are listed in Figure 7.12.

Figure 7.10: View of 3D scatterplot of translation vectors for files 1j5e and 2zm6. Units are Ångstroms. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.
Figure 7.11: View of 3D scatterplot of rotation vectors for files 1j5e and 2zm6. Units are degrees. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.
In this section, we consider crystal structures of the same molecule of the same organism that have different ligands bound. The 1j5e crystal structure, which has no ligands bound, will again be considered. It will be compared with two different crystal structures of \( T.th. \) that do have ligands bound, 1hnw and 1hnx. 1hnw is a crystal structure of \( T.th. \) in complex with the antibiotic tetracycline and 1hnx is a crystal structure of \( T.th. \) in complex with the antibiotic pactamycin [8].

Both of these crystal structures were obtained by soaking the \( T.th. \) structure in the respective antibiotic (post-crystallization) at a concentration at which they were known to bind to the molecule. These two antibiotics, tetracycline and pactamycin, are known to bind to different regions of the molecule [8]. Most ribosomal antibiotics interfere with the function of the ribosome during protein synthesis by binding to specific sites of the ribosomal RNA.
Understanding the structural effects these antibiotics have on the ribosome is important in understanding the effects on the ribosome function.

We first consider the effects of tetracycline by comparing 1hnw with 1j5e. Tetracycline has been used since the 1940s, but has there has been a widespread microbial resistance to it in recent years. This emergence of new antibiotic resistant bacteria further necessitates understanding how these antibiotics interact with the ribosome.

The scatterplots of the translation and rotation vectors are given in Figures 7.13 and 7.14, respectively. It is observed that there is very little variability among the data points which makes sense since 1hnw is not in reality a totally new crystal structure; it was formed post-crystallization of 1fjf (which was superseded by 1j5e). 1fjf was also shown to be very similar to 1j5e in Section 7.5.

The two variance-covariance matrices are

\[
S^T = \begin{bmatrix}
0.0168 & -0.0016 & 0.0036 \\
-0.0016 & 0.0199 & -0.0007 \\
0.0036 & -0.0007 & 0.0205
\end{bmatrix}
\]

\[
S^R = \begin{bmatrix}
4.0278 & 0.1036 & -1.2611 \\
0.1036 & 2.2147 & 0.2744 \\
-1.2611 & 0.2744 & 2.9815
\end{bmatrix}
\]

We next consider the effects of pactamycin and display the corresponding scatterplots. Pactamycin has been found to have an effect on translation in all three domains, eukarya, bacteria, and archaea. Because of this, pactamycin is expected to interact with highly conserved regions of the 16S RNA structure.

The scatterplots of the translation and rotation vectors are given in Figures 7.15 and 7.16, respectively. Since 1hnw and 1hnx are very similar, Figures 7.15 and 7.16 are very similar to Figures 7.13 and 7.14, respectively.
Figure 7.13: View of 3D scatterplot of translation vectors for files 1j5e and 1hnw. Units are Angstroms. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.

The two variance-covariance matrices are

$$S_T = \begin{bmatrix} 0.0195 & -0.0024 & 0.0026 \\ -0.0024 & 0.0214 & -0.0012 \\ 0.0026 & -0.0012 & 0.0199 \end{bmatrix} \quad S_R = \begin{bmatrix} 4.8004 & 0.0926 & -1.4297 \\ 0.0926 & 2.2627 & 0.3770 \\ -1.4297 & 0.3770 & 3.3778 \end{bmatrix}$$

Since the scatterplots for the comparisons with the two different structures are so similar, we next investigate the $D_i^2$ which measure the variability of the conformational changes. The nucleotides with the largest values of $D_i^2$, along with the corresponding $D_i^2$ values, are listed in Figure 7.17. The values were calculated based on the translation vectors.

While nucleotides 1491-1493 are different in 1hnw from 1j5e, the nucleotides in 1hn2
Figure 7.14: View of 3D scatterplot of rotation vectors for files 1j5e and 1hnw. Units are degrees. Points are labeled if the corresponding value for $D_2^i$ is greater than 25.

making contact with tetracycline are 242, 906-908, 892-894, 965-966, 1054, and 1195. This indicates that binding tetracycline does not significantly affect the local translations or orientations.

7.8 Different Organism, No Ligands Bound, Different Lab Analysis

In this section, we analyze two 16S rRNAs of different organisms crystallized by different research labs. When we compare two molecules from different organisms, we do not know
Figure 7.15: View of 3D scatterplot of translation vectors for files 1j5e and 1hnx. Units are Angstroms. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.
Figure 7.16: View of 3D scatterplot of rotation vectors for files 1j5e and 1lnx. Units are degrees. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.
Figure 7.17: Largest values of $D_i^2$ for 1j5e and 1hnx based on the translation vector data

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>$D_i^2$</th>
<th>Nucleotide</th>
<th>$D_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1492</td>
<td>107.814967</td>
<td>A1492</td>
<td>80.59159</td>
</tr>
<tr>
<td>A1493</td>
<td>107.723438</td>
<td>A1493</td>
<td>79.42878</td>
</tr>
<tr>
<td>G1491</td>
<td>95.8887878</td>
<td>G1491</td>
<td>74.26600</td>
</tr>
<tr>
<td>G1443</td>
<td>55.9833746</td>
<td>G1443</td>
<td>41.78761</td>
</tr>
<tr>
<td>C1163</td>
<td>35.9115094</td>
<td>G1031</td>
<td>35.57215</td>
</tr>
<tr>
<td>G1442</td>
<td>34.2379494</td>
<td>G1030A</td>
<td>32.25652</td>
</tr>
<tr>
<td>A1446</td>
<td>34.2379494</td>
<td>G1032</td>
<td>31.05792</td>
</tr>
<tr>
<td>G1494</td>
<td>31.2087025</td>
<td>C1030</td>
<td>29.03551</td>
</tr>
<tr>
<td>C1030B</td>
<td>28.8509619</td>
<td>C1163</td>
<td>28.75354</td>
</tr>
<tr>
<td>G1030C</td>
<td>28.8509619</td>
<td>G1442</td>
<td>28.58864</td>
</tr>
<tr>
<td>A1030D</td>
<td>28.8509619</td>
<td>A1446</td>
<td>28.58864</td>
</tr>
<tr>
<td>G1030A</td>
<td>28.5442812</td>
<td>G1164</td>
<td>25.40779</td>
</tr>
<tr>
<td>G1173</td>
<td>28.2811427</td>
<td>G1173</td>
<td>24.79179</td>
</tr>
<tr>
<td>G1174</td>
<td>26.1137649</td>
<td>C1030B</td>
<td>24.61511</td>
</tr>
<tr>
<td>G1009</td>
<td>25.4722673</td>
<td>C1030C</td>
<td>24.61511</td>
</tr>
<tr>
<td>G1164</td>
<td>25.4165626</td>
<td>G1030C</td>
<td>24.61511</td>
</tr>
<tr>
<td>C1162</td>
<td>24.6929828</td>
<td>A1483</td>
<td>23.24143</td>
</tr>
<tr>
<td>G1031</td>
<td>24.001683</td>
<td>G1174</td>
<td>21.57507</td>
</tr>
<tr>
<td>A1483</td>
<td>22.1599167</td>
<td>C1029</td>
<td>21.01771</td>
</tr>
<tr>
<td>C1409</td>
<td>20.101754</td>
<td>C1162</td>
<td>20.74588</td>
</tr>
<tr>
<td>C1008</td>
<td>19.8480511</td>
<td>G1482</td>
<td>20.30800</td>
</tr>
<tr>
<td>C1030</td>
<td>18.0871149</td>
<td>G1419</td>
<td>19.42480</td>
</tr>
<tr>
<td>U1020</td>
<td>18.0740233</td>
<td>G1494</td>
<td>19.35968</td>
</tr>
<tr>
<td>G1021</td>
<td>17.4807439</td>
<td>G1033</td>
<td>18.67407</td>
</tr>
<tr>
<td>C1484</td>
<td>16.6883441</td>
<td>C1420</td>
<td>18.54142</td>
</tr>
<tr>
<td>G1482</td>
<td>16.0398023</td>
<td>U1481</td>
<td>17.76545</td>
</tr>
<tr>
<td>G1175</td>
<td>15.917222</td>
<td>G1009</td>
<td>17.50402</td>
</tr>
<tr>
<td>G1036</td>
<td>15.7557395</td>
<td>C1484</td>
<td>17.45590</td>
</tr>
<tr>
<td>C1172</td>
<td>15.7020192</td>
<td>C1426</td>
<td>16.57629</td>
</tr>
<tr>
<td>C1424</td>
<td>14.8883778</td>
<td>U204</td>
<td>16.48960</td>
</tr>
<tr>
<td>G1010</td>
<td>14.8054778</td>
<td>G1475</td>
<td>16.22346</td>
</tr>
<tr>
<td>G1417</td>
<td>14.7367651</td>
<td>U1427</td>
<td>15.93851</td>
</tr>
<tr>
<td>G1032</td>
<td>14.7298641</td>
<td>A1418</td>
<td>15.74116</td>
</tr>
<tr>
<td>G1475</td>
<td>14.2948461</td>
<td>C1456</td>
<td>15.38063</td>
</tr>
<tr>
<td>G1476</td>
<td>14.2503195</td>
<td>A1005</td>
<td>15.34237</td>
</tr>
<tr>
<td>C1165</td>
<td>13.9601546</td>
<td>G476</td>
<td>15.00616</td>
</tr>
<tr>
<td>U1425</td>
<td>13.3906322</td>
<td>C1172</td>
<td>15.00089</td>
</tr>
<tr>
<td>C1477</td>
<td>12.7980446</td>
<td>C1008</td>
<td>14.84953</td>
</tr>
</tbody>
</table>
exactly which nucleotides correspond, so an alignment of the two structures is first required. We want to analyze the conformational changes between the *T.th.* 1j5e and *E.c.* 2avy files. A structural alignment was obtained using R3D Align (Chapter 5).

Because 1j5e and 2avy are crystal structures of different molecules, we expect there to be greater variability in the translation and rotation vectors than in the previous cases where the crystal structures are of the same molecule. The scatterplots are shown for the translation and rotation vectors are shown in Figures 7.18 and 7.19 respectively.

![3D scatterplot of translation vectors for files 1j5e and 2avy](image)

**Figure 7.18:** View of 3D scatterplot of translation vectors for files 1j5e and 2avy. Points are centered about the origin and colored according to the corresponding nucleotides distance from the center of the structure. Units are Ångstroms. Points are labeled if the corresponding value for $D_i^2$ is greater than 25.

It is evident from the scatterplots that there is indeed more variability in this case. This
confirmed by analyzing the variance-covariance matrices which are

\[
S^T = \begin{bmatrix}
2.2310 & -0.0160 & -0.4510 \\
-0.0160 & 2.5597 & 1.2031 \\
-0.4510 & 1.2031 & 5.1068
\end{bmatrix}
\quad S^R = \begin{bmatrix}
375.9167 & 230.1330 & -124.0879 \\
230.1330 & 283.5737 & -171.2514 \\
-124.0879 & -171.2514 & 265.5654
\end{bmatrix}
\]

The total sample variances are then 9.8974 and 925.0557, much higher than in the other cases where the structures were from the same molecule.

Finally, Figure 7.20 lists the nucleotides that have the most variability in the translation
and rotation vectors according to the $D_i^2$ values.

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>$D_i^2$</th>
<th>Nucleotide</th>
<th>$D_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>U81</td>
<td>105.161</td>
<td>A1460</td>
<td>153.649</td>
</tr>
<tr>
<td>G80</td>
<td>105.161</td>
<td>C1448</td>
<td>138.283</td>
</tr>
<tr>
<td>C89</td>
<td>105.161</td>
<td>C1459</td>
<td>138.283</td>
</tr>
<tr>
<td>U90</td>
<td>105.161</td>
<td>C1449</td>
<td>132.306</td>
</tr>
<tr>
<td>G70</td>
<td>75.513</td>
<td>G1455</td>
<td>127.770</td>
</tr>
<tr>
<td>C73</td>
<td>75.513</td>
<td>U1450</td>
<td>124.634</td>
</tr>
<tr>
<td>U98</td>
<td>75.513</td>
<td>C1452</td>
<td>124.634</td>
</tr>
<tr>
<td>C91</td>
<td>71.163</td>
<td>G1454</td>
<td>123.011</td>
</tr>
<tr>
<td>C1028</td>
<td>51.119</td>
<td>C99</td>
<td>116.355</td>
</tr>
<tr>
<td>G1034</td>
<td>49.412</td>
<td>G70</td>
<td>115.369</td>
</tr>
<tr>
<td>G1033</td>
<td>49.412</td>
<td>C73</td>
<td>115.369</td>
</tr>
<tr>
<td>C1029</td>
<td>48.264</td>
<td>U98</td>
<td>115.369</td>
</tr>
<tr>
<td>C99</td>
<td>47.863</td>
<td>G1453</td>
<td>106.676</td>
</tr>
<tr>
<td>C1030</td>
<td>43.613</td>
<td>C89</td>
<td>83.635</td>
</tr>
<tr>
<td>G1031</td>
<td>43.613</td>
<td>U81</td>
<td>83.635</td>
</tr>
<tr>
<td>G1032</td>
<td>43.613</td>
<td>U90</td>
<td>83.635</td>
</tr>
<tr>
<td>G1453</td>
<td>42.474</td>
<td>G80</td>
<td>83.635</td>
</tr>
<tr>
<td>C92</td>
<td>40.397</td>
<td>U95</td>
<td>82.062</td>
</tr>
<tr>
<td>G78</td>
<td>40.397</td>
<td>G75</td>
<td>82.062</td>
</tr>
<tr>
<td>C1448</td>
<td>35.208</td>
<td>C76</td>
<td>82.062</td>
</tr>
</tbody>
</table>

Figure 7.20: Largest values of $D_i^2$ for 1j5e and 2avy data, indicating extreme data points

There are clear clusters of local neighborhoods that vary all in the same way. Most of these are at some distance from the center of the molecule, where more flexibility is allowed.

### 7.9 All-Against-All Comparison of *T.th.* 16S Structures

As we’ve discussed, not all of the crystal structures in the PDB represent distinct organisms. In fact, there are nearly sixty PDB files that include a crystal structure of *T.th.* 16S rRNA. Most of these crystal structures include the 16S molecule in complex with mRNA, tRNA, one or more antibiotics, or some other type of ligand.

We have already shown that it is of interest to compare pairs of structures and analyze the similarities and differences. Also important is to do an all-against-all comparison of the structures and group them according to the overall similarities of their structures. As a
measure of similarity between two structures, we can use the total sample variance of the translation or rotation vectors, as described above.

Here we focus on the translation vectors to analyze a subset of the *T. th.* 16S rRNA structures. The PDB files were selected that had identical base sequences to the 1j5e crystal structure in the first 1500 nucleotides, so that it was not necessary to determine an alignment. There procedure yielded 28 PDB files (including 1j5e). PDB file 2zm6 is also equivalent when disregarding an extra nucleotide at the beginning of the sequence, so that structure is considered as well, giving a set of 29 PDB files.

Figure 7.21 displays the mutual total sample variances among the 29 crystal structures. The variances are represented by colors as indicated by the color bars. The figure illustrates which structures and groups are more similar and which are not.

![Figure 7.21: Mutual translation vector total sample variance matrix](image-url)
A complete linkage cluster analysis was also performed on the structures. The dendrogram is provided in Figure 7.22. As expected many structures crystallized by the same group are in the same cluster. For example, the crystal structures 2uua, 2uub, 2uuc, and 2uu9, which were all crystallized by the same group under the same conditions, are clustered together.

![Dendrogram](image)

Figure 7.22: Dendrogram of 1j5e and the 28 structures with equivalent sequences.

In some cases, common structures are related in that they are both in complex with the same type of molecule. For example, the dendrogram illustrates that the 1IBL and 1N32 crystal structures are closely related. While 1ibl and 1n32 were deposited in the PDB 18 months apart, they are both crystal structures of the 16S rRNA in complex with the antibiotic paromomycin.
CHAPTER 8

R3D Align User Manual

8.1 Installation

R3D Align was written in Matlab version 7.7 and has been incorporated into the FR3D suite of Matlab programs. For more information on FR3D, go to http://rna.bgsu.edu/FR3D. The easiest way to install R3D Align is to download the latest version of the zip file from http://r3dalign.googlecode.com. This will download a folder named FR3D and several subfolders. The R3D Align subfolder contains the files that are specific to its use. The first time R3D Align is asked to align a given PDB file, it reads the text, analyzes it, and saves a data file in the subfolder named PrecomputedData. After that, it will not need to re-read the original PDB file. PrecomputedData contains data for four large PDB files upon download. The subfolder PDBfiles is where the user should place the PDB files of the structures to be aligned by R3D Align. If you have another folder on your computer with PDB files, add that folder to Matlab’s path (File, Set Path, Add Folder, Save).

8.2 Initialization

Start Matlab and change the working directory to FR3D (you can use the cd command to change the directory or the “Set Path...” command under the File menu).
8.3 Calling R3DAlign.m

To align two 3D structures, the user should call the function contained in R3DAlign.m, which has the following syntax:

\[
[\text{AlignedNTs1, AlignedNTs2}] = \text{R3DAlign}(\text{File1, NTList1, File2, NTList2, discCut, numNeigh, bandwidth, cliqueMethod, seed1, seed2})
\]

The function R3DAlign accepts 9 parameters, of which the first 7 are required. A short description and the proper syntax for each is provided next.

8.4 Input Parameters

\text{File1} is the pdb id for the first RNA molecule. The parameter must be entered as a string variable, so the PDB id is to be placed in quotes (eg. ’1s72’)

\text{NTList1} specifies which nucleotides of the first RNA molecule are to be aligned. The user can specify the nucleotides in a variety of ways:

- To align all the nucleotides contained in \text{File1}, simply enter the string ’all’ for this parameter.

- To align all nucleotides contained in a specific chain of \text{File1}, enter the chain id placed in quotes (eg. ’A’).

- To align a specific set of nucleotides, a string of nucleotide numbers separated by commas can be entered. Any entry of the string can use the colon notation for a range of nucleotide numbers (eg. 121:125 ). Nucleotide numbers can be followed by (A) or _A to indicate chain A. For example to align nucleotides 1 to 5, 7, and 8-10 of chain B in \text{File1}, enter ’1:5(B), 7(B), 8:10(B)’. The quotes must be entered.

\text{File2} is the PDB id for the second molecule.
**NTList2** specifies which nucleotides of the second RNA molecule are to be aligned. The same options exist for **NTList2** as for **NTList1**. **NTList2** does not need to contain the same number of nucleotides as **NTList1**.

**discCut** is the geometric discrepancy cutoff value for which two neighborhoods are considered to be structurally similar. **discCut** is a positive real-valued number (e.g., 0.5). If **discCut** = 0.5, then two sets of nucleotides (neighborhoods) with a geometric discrepancy of less than or equal 0.5 are considered to be similar and any geometric discrepancy above 0.5 implies dissimilarity.

**numNeigh** is positive, integer-valued, and gives the number of neighborhoods to retain for each nucleotide. If **numNeigh** = 7, then the seven four-nucleotide neighborhoods with the smallest diameter will be retained for each nucleotide.

**bandwidth** is also positive and integer-valued. If **bandwidth** = b, then nucleotide i of the first structure will be compared with nucleotide k only if j − b < k < j + b, where j is the nucleotide to which i is aligned in the seed alignment.

**cliqueMethod** specifies whether the alignment will be produced by finding a clique using the greedy clique procedure or the full branch-and-bound maximum clique procedure. **cliqueMethod** is a string variable with two possible values: ‘greedy’ and ‘full’.

**seed1** and **seed2** are optional parameters which can be used to enter a seed alignment. They are cell arrays of the same size containing the nucleotide numbers of aligned nucleotides in corresponding positions. That is, for each i, **seed1**{i, 1} contains the number of the nucleotide in **File1** that is aligned with the nucleotide in **File2** with the number contained in **seed2**{i, 1}. The chain for each nucleotide can be identified in **seed1**{i, 2} for each i. If **seed1** and **seed2** are omitted, then R3DAlign will internally compute a seed alignment using a sequence alignment algorithm. The output arguments described next can directly be used as **seed1** and **seed2**.
8.5 Output Arguments

The function returns two arguments:

AlignedNTs1 is an $M \times 3$ cell array, where $M$ is the number of nucleotides aligned. The first column of the cell array contains the numbers of the aligned nucleotides. The second column lists the chain for each aligned nucleotide. The third column contains the base type of each aligned nucleotide. Nucleotides in File1 aligned with a gap are not included in the cell array.

AlignedNTs2 is also an $M \times 3$ cell array where the nucleotide represented in AlignedNTs2{$j$} is aligned with the nucleotide represented in AlignedNTs1{$j$}

AlignedNTs1 and AlignedNTs2 can be used in a subsequent call to R3DAlign.m as seed1 and seed2.

8.6 Output Files

R3DAlign.m also produces three additional files, each containing the alignment information in different formats. Each of the three files is produced automatically and saved in the current working directory. These three files include:

1. A FASTA file containing the alignment
2. An EXCEL spreadsheet that simultaneously displays the alignment and basepairing interactions of the structures
3. A ‘bar diagram’ that displays the alignment and indicates the structural similarity of aligned local neighborhoods. This ‘bar diagram’ is saved as a PDF file.

All three files are automatically saved with the filename in the following format:

‘‘R3D Align Name_of_File1 Name_of_File2 Current_Date Current_Time''
The FASTA file is saved with the extension ‘.FASTA’, the Excel file with ‘.xlsx’, and the bar diagram with ‘.pdf’.

8.7 Examples

Example 1:
Aligning all the nucleotides of 1U9S with chain B of 1NBS; a discrepancy cutoff of 0.3 is to be used; 12 neighborhoods are to be retained for each nucleotide; a bandwidth of 60 is to be used on the internally computed seed alignment; the full maximum clique algorithm will be employed:

\[
\text{[AlNTs1u9s, AlNTs1nbs]} = \text{R3DAlign('1u9s', 'all', '1nbs', 'B', 0.3, 12, 60, 'full')};
\]

The produced bar diagram looks like the following:

The FASTA file contains the following:

> 1U9S
-GGUGCCAGGUAACG------------CCUGGGCGGGUAAC------CCGACGGAAA----------------------
-----------------------------------GUGCCACAGA-GAAGAGACCGCCAGCGGC-----CGGGG--------CCCCGGUGCG
GGCAAGG-GUGAAACGCGGGGUAAGAGCCCACCACCGCGCCUGGCAACAGGCCGGGGCAC---GGCAA-ACCCCACC--
> 1NBS B
GCCAGCCUAGC-----CGAAGUCAUAAGCUAGGCCAG-------UCUUGGCUAGACGCGC---AGGAAAAAAGCCUACGUCUUCG
GAAUAGGCUAGUAUCUUAGAAAGUGCCACAG-UGAC---------------------GAAGUCUCACUAGAAUGGUGAG------
----------AGUGGAA----------------------CGCGGUA-AACCCUC-GC
**Example 2:**

Aligning the 5S rRNAs of *E. Coli* (chain A of PDB 2aw4) and *T. thermophilus* (chain B of PDB 2J01); discrepancy cutoff 0.5; 7 neighborhoods; bandwidth = 50; full clique procedure used for final alignment; no seed alignment manually entered:

\[
\text{[A1NTs2aw4,A1NTs2j01]} = \text{R3DAlign('2aw4', 'A', '2j01', 'B', 0.5, 7, 50, 'full')};
\]

Bar diagram:

![Bar diagram](image)

FASTA output:

> 2aw4 A
-GCCUGGCGGCC-GUAGCGCGG--UGGU-CCCACCUGAC-CCCAUGCCGAACUCAGA-AGUGAAACGCCG-UA-GCGCCG
AUGGUAGUG-UUGGGUCU----CCCCAUGCGAGAUGGAACUGCCAGGC

> 2J01 B
UCCCCCGUGCC-CAUAGCGG--CGUGG-AACCACCCGU-UCCCAUUCCGAACACGG-AAGUGAAACCGC-CC-AGCGCCG
AUGGUACU-GGGCGGG--CGACCCGCCUGGAGAGUAGGUGCGGCGG

**Example 3:**

Same as Example 2 but only aligning nucleotides 2-20, 62-69, 109-118 of 2aw4 and nucleotides 2-20, 62-69, 110-119 of 2J01:

\[
\text{[A1NTs2aw4,A1NTs2j01]} = \text{R3DAlign('2aw4', '2:20(A),62:69(A),109:118(A)', '2j01', '2:20(B),62:69(B),110:119(B)', 0.5, 7, 50, 'full')};
\]

Bar diagram:
FASTA output:

> 2aw4 2:20(A),62:69(A),109:118(A)
GCCUGCCGCCC-GU-AGCGCGCCGUA-GCGACUUGCAGGC
> 2J01 2:20(B),62:69(B),110:119(B)
CCCCCGUGCC-CA-UAGCGGCCGCC-AGCGGGUGCGGGG

Example 4:
Align all nucleotides in 1j5e and all nucleotides in 2avy; discrepancy cutoff=0.5; 3 neighborhoods; bandwidth=30; greedy clique procedure used for final alignment; no seed alignment manually entered

[AlNTs1j5e1, AlNTs2avy1] = R3DAlign('1j5e','all','2avy','all',0.5,3,30,'greedy');

Bar diagram:

FASTA output:

> 1J5E
UGGAGAGUUUGAUCGCGCCAGGGUGAAGCUGCGCGCGCGUAGACUAAGCAUGCGU-GCG-----GGCCGCGGG--
Example 5 (entering a seed alignment):

Align all nucleotides in 1j5e and all nucleotides in 2avy; discrepancy cutoff=0.5; 9 neighborhoods; bandwidth=10; greedy clique procedure used for final alignment; use output from Example 4 as the seed alignment:

```
[AlNTs1j5e2, AlNTs2avy2] = R3DAlign('1j5e','all','2avy','all',0.5,9,10,
                                 'greedy', AlNTs1j5e1, AlNTs2avy1);
```

Bar diagram:

```
1J5E

1.1

2AVY
```

FASTA output omitted.
BIBLIOGRAPHY


[53] A. Pavesi, R. Percudani, and F. Conterio. A novel algorithm for the search of 5S rRNA genes in DNA databases: comparison with other methods and identification of


# Appendix A—R3D Align 16S

## Alignment Spreadsheet

<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U5</td>
<td>---</td>
<td>G39 cWW</td>
<td>G403 cWW</td>
</tr>
<tr>
<td>G6</td>
<td>---</td>
<td>G40 cWW</td>
<td>G402 cWW</td>
</tr>
<tr>
<td>G7</td>
<td>---</td>
<td>C41 cWW</td>
<td>G401 cWW</td>
</tr>
<tr>
<td>A8</td>
<td>---</td>
<td>G42 cWW</td>
<td>G42 cWW</td>
</tr>
<tr>
<td>G9 cWW</td>
<td>C25 G9 cWW</td>
<td>G42 tSS</td>
<td>A622 cWW</td>
</tr>
<tr>
<td>A10 cWW</td>
<td>U24 A10 cWW</td>
<td>G43 cWW</td>
<td>G399 cWW</td>
</tr>
<tr>
<td>A10 cSS</td>
<td>A10 cSS</td>
<td>G44 cWW</td>
<td>A44 cWW</td>
</tr>
<tr>
<td>G11 cWW</td>
<td>C23 G11 cWW</td>
<td>G45 cWW</td>
<td>G396 cWW</td>
</tr>
<tr>
<td>U12 cWW</td>
<td>G22 U12 cWW</td>
<td>G46 cWW</td>
<td>G395 cWW</td>
</tr>
<tr>
<td>U13 cWW</td>
<td>U20 U13 cWH</td>
<td>C47 cWW</td>
<td>G361 cWW</td>
</tr>
<tr>
<td>U13 cWH</td>
<td>A19 U13 cWH</td>
<td>C48 cWW</td>
<td>---</td>
</tr>
<tr>
<td>U14 ---</td>
<td>U14 ---</td>
<td>U49 tSS</td>
<td>A632 cWW</td>
</tr>
<tr>
<td>G15 cWW</td>
<td>U920 G15 cWW</td>
<td>U49 tWW</td>
<td>U365 cWW</td>
</tr>
<tr>
<td>A16 cSS</td>
<td>A1080 A16 cSS</td>
<td>A50 ntSS</td>
<td>A360 cWW</td>
</tr>
<tr>
<td>U17 cWW</td>
<td>A918 U17 cWW</td>
<td>A51 cWS</td>
<td>U114 cWS</td>
</tr>
<tr>
<td>U17 cSS</td>
<td>G1079 U17 cSS</td>
<td>A51 tSS</td>
<td>U114 tSS</td>
</tr>
<tr>
<td>C18 cSS</td>
<td>A854 A18 ncSS</td>
<td>A52 cWW</td>
<td>A360 cWW</td>
</tr>
<tr>
<td>C18 cWW</td>
<td>G917 C18 cWW</td>
<td>A53 cWW</td>
<td>U358 cWW</td>
</tr>
<tr>
<td>C19 tSS</td>
<td>A572 A19 tSS</td>
<td>C54 cWW</td>
<td>G357 cWW</td>
</tr>
<tr>
<td>C19 cWW</td>
<td>G916 A19 cWW</td>
<td>C55 tHS</td>
<td>G357 tHS</td>
</tr>
<tr>
<td>U20 tWH</td>
<td>U13 U20 tWH</td>
<td>C55 tWW</td>
<td>U368 cWW</td>
</tr>
<tr>
<td>G21 tSH</td>
<td>A914 G21 tSH</td>
<td>U56 cWW</td>
<td>A356 cWW</td>
</tr>
<tr>
<td>G22 cWW</td>
<td>U12 G22 cWW</td>
<td>G57 cWW</td>
<td>G355 cWW</td>
</tr>
<tr>
<td>G22 cSS</td>
<td>A913 G22 cSS</td>
<td>C58 cWW</td>
<td>G354 cWW</td>
</tr>
<tr>
<td>C23 cWW</td>
<td>G11 C23 cWW</td>
<td>A59 cWW</td>
<td>---</td>
</tr>
<tr>
<td>C23 cWW</td>
<td>U20 U24 cWW</td>
<td>A60 tSS</td>
<td>G107 cWW</td>
</tr>
<tr>
<td>C25 cWW</td>
<td>G9 C25 cWW</td>
<td>G61 cWW</td>
<td>C106 cWW</td>
</tr>
<tr>
<td>A26 tHS</td>
<td>G558 A26 tHS</td>
<td>U62 cWW</td>
<td>G105 cWW</td>
</tr>
<tr>
<td>G27 cWW</td>
<td>G556 G27 cWW</td>
<td>C63 cWW</td>
<td>G104 cWW</td>
</tr>
<tr>
<td>G28 cWW</td>
<td>C555 A28 cWW</td>
<td>G64 cWW</td>
<td>G68 cWW</td>
</tr>
<tr>
<td>G29 cWW</td>
<td>C554 U28 cWW</td>
<td>U65 cWW</td>
<td>---</td>
</tr>
<tr>
<td>G30 cWW</td>
<td>A553 U30 cWW</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G31 ---</td>
<td>G11 ---</td>
<td>G66 cWW</td>
<td>C101 cWW</td>
</tr>
<tr>
<td>A32 cWW</td>
<td>U552 A32 cWW</td>
<td>G66 ncSS</td>
<td>A172 cSS</td>
</tr>
<tr>
<td>A33 cWW</td>
<td>U551 A33 cWW</td>
<td>C67 cWW</td>
<td>G102 cWW</td>
</tr>
<tr>
<td>C34 cWW</td>
<td>G550 C34 cWW</td>
<td>C67 cSS</td>
<td>A171 cSS</td>
</tr>
<tr>
<td>G35 cWW</td>
<td>C549 G35 cWW</td>
<td>G68 cWW</td>
<td>G68 cWH</td>
</tr>
<tr>
<td>G36 cWW</td>
<td>G548 C36 cWW</td>
<td>G68 cWW</td>
<td>G68 cWW</td>
</tr>
<tr>
<td>G37 tWW</td>
<td>A397 U37 tWW</td>
<td>G68 cWW</td>
<td>A101 cWW</td>
</tr>
<tr>
<td>G38 ---</td>
<td>G38 ---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

---

155
<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A171</td>
<td>tSS</td>
<td>C103</td>
<td>A171</td>
</tr>
<tr>
<td>A171</td>
<td>tSH</td>
<td>C150</td>
<td>A171</td>
</tr>
<tr>
<td>A172</td>
<td>nSS</td>
<td>C103</td>
<td>A172</td>
</tr>
<tr>
<td>A172</td>
<td>tHW</td>
<td>A149</td>
<td>A172</td>
</tr>
<tr>
<td>A172</td>
<td>nSS</td>
<td>G66</td>
<td>A172</td>
</tr>
<tr>
<td>U173</td>
<td>---</td>
<td>U173</td>
<td>---</td>
</tr>
<tr>
<td>C174</td>
<td>cWW</td>
<td>G148</td>
<td>C174</td>
</tr>
<tr>
<td>C175</td>
<td>cWW</td>
<td>G147</td>
<td>C175</td>
</tr>
<tr>
<td>C175</td>
<td>---</td>
<td>C175</td>
<td>nSS</td>
</tr>
<tr>
<td>C176</td>
<td>cWW</td>
<td>G146</td>
<td>C176</td>
</tr>
<tr>
<td>C177</td>
<td>cWW</td>
<td>G145</td>
<td>C177</td>
</tr>
<tr>
<td>C178</td>
<td>cWW</td>
<td>G144</td>
<td>C178</td>
</tr>
<tr>
<td>A179</td>
<td>tWW</td>
<td>A196</td>
<td>A179</td>
</tr>
<tr>
<td>U180</td>
<td>tWW</td>
<td>A195</td>
<td>U180</td>
</tr>
<tr>
<td>G181</td>
<td>cSH</td>
<td>U182</td>
<td>A181</td>
</tr>
<tr>
<td>U182</td>
<td>cSH</td>
<td>G181</td>
<td>A181</td>
</tr>
<tr>
<td>A183</td>
<td>cWW</td>
<td>C194</td>
<td>A183</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>A184</td>
<td>---</td>
</tr>
<tr>
<td>G184</td>
<td>cWW</td>
<td>C193</td>
<td>G184</td>
</tr>
<tr>
<td>A185</td>
<td>cWW</td>
<td>U192</td>
<td>A185</td>
</tr>
<tr>
<td>C186</td>
<td>cWW</td>
<td>G191</td>
<td>C186</td>
</tr>
<tr>
<td>C187</td>
<td>cWW</td>
<td>U190L</td>
<td>C187</td>
</tr>
<tr>
<td>C187</td>
<td>---</td>
<td>G187</td>
<td>---</td>
</tr>
<tr>
<td>C188</td>
<td>cWW</td>
<td>G190K</td>
<td>---</td>
</tr>
<tr>
<td>A190</td>
<td>cWW</td>
<td>U1901</td>
<td>A190</td>
</tr>
<tr>
<td>C190</td>
<td>cWW</td>
<td>G190L</td>
<td>---</td>
</tr>
<tr>
<td>C190A</td>
<td>cWW</td>
<td>G190H</td>
<td>C190A</td>
</tr>
<tr>
<td>C190B</td>
<td>cWW</td>
<td>G190G</td>
<td>---</td>
</tr>
<tr>
<td>C190C</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U190D</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U190E</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G190F</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G190G</td>
<td>cWW</td>
<td>C190B</td>
<td>---</td>
</tr>
<tr>
<td>G190H</td>
<td>cWW</td>
<td>C190A</td>
<td>---</td>
</tr>
<tr>
<td>C190I</td>
<td>cWW</td>
<td>C190</td>
<td>---</td>
</tr>
<tr>
<td>U190J</td>
<td>cWW</td>
<td>G189</td>
<td>---</td>
</tr>
<tr>
<td>G190K</td>
<td>cWW</td>
<td>C188</td>
<td>---</td>
</tr>
<tr>
<td>U190L</td>
<td>cWW</td>
<td>C187</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>C187</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>A189</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>C187</td>
<td>A190</td>
</tr>
<tr>
<td>G191</td>
<td>cWW</td>
<td>C186</td>
<td>G191</td>
</tr>
<tr>
<td>U192</td>
<td>cWW</td>
<td>A185</td>
<td>A192</td>
</tr>
<tr>
<td>A192</td>
<td>cWW</td>
<td>G184</td>
<td>C193</td>
</tr>
<tr>
<td>C193</td>
<td>cWW</td>
<td>G183</td>
<td>C194</td>
</tr>
<tr>
<td>C194</td>
<td>cWW</td>
<td>G183</td>
<td>C194</td>
</tr>
<tr>
<td>A195</td>
<td>tWW</td>
<td>U190</td>
<td>A195</td>
</tr>
<tr>
<td>A195</td>
<td>cWW</td>
<td>U282</td>
<td>A195</td>
</tr>
<tr>
<td>C196</td>
<td>cSS</td>
<td>G142</td>
<td>A196</td>
</tr>
<tr>
<td>C196</td>
<td>tSS</td>
<td>G142</td>
<td>A196</td>
</tr>
<tr>
<td>A196</td>
<td>tWW</td>
<td>A179</td>
<td>A196</td>
</tr>
<tr>
<td>A196</td>
<td>cSS</td>
<td>C221</td>
<td>A196</td>
</tr>
<tr>
<td>A196</td>
<td>ncSS</td>
<td>U222</td>
<td>A196</td>
</tr>
<tr>
<td>A197</td>
<td>ntSS</td>
<td>G220</td>
<td>A197</td>
</tr>
<tr>
<td>G198</td>
<td>cWW</td>
<td>C219</td>
<td>G198</td>
</tr>
<tr>
<td>A199</td>
<td>cWW</td>
<td>C218</td>
<td>A199</td>
</tr>
<tr>
<td>G200</td>
<td>cWW</td>
<td>C217</td>
<td>G200</td>
</tr>
<tr>
<td>C201</td>
<td>cWW</td>
<td>G216</td>
<td>C201</td>
</tr>
<tr>
<td>C202</td>
<td>---</td>
<td>G201</td>
<td>---</td>
</tr>
<tr>
<td>U202</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U203</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U204</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>G202</td>
<td>cWW</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>U256 cWW</td>
<td>A270 U256 cWW</td>
<td>C311 cWW</td>
<td>G289 C311 cWW</td>
</tr>
<tr>
<td>G257 cWW</td>
<td>G269 G257 cWW</td>
<td>C312 cWW</td>
<td>G135 C312 cWW</td>
</tr>
<tr>
<td>G258 cWW</td>
<td>G268 G258 cWW</td>
<td>A313 cWW</td>
<td>U114 A313 cWW</td>
</tr>
<tr>
<td>G259 cWW</td>
<td>G267 G259 cWW</td>
<td>C314 cSW</td>
<td>A51 C314 cSW</td>
</tr>
<tr>
<td>G260 ---</td>
<td>---</td>
<td>C314 cWW</td>
<td>G113 C314 cWW</td>
</tr>
<tr>
<td>U261 ---</td>
<td>U261 ---</td>
<td>A315 cWW</td>
<td>G112 A315 cWW</td>
</tr>
<tr>
<td>A262 cSS</td>
<td>C131 A262 cSS</td>
<td>G316 cWW</td>
<td>C337 C316 cWW</td>
</tr>
<tr>
<td>A263 cSS</td>
<td>A130 A263 cSS</td>
<td>G317 cWW</td>
<td>C336 U317 cWW</td>
</tr>
<tr>
<td>A266 cSS</td>
<td>G332 A263 cSS</td>
<td>G318 cWW</td>
<td>C335 G318 cWW</td>
</tr>
<tr>
<td>G265 ---</td>
<td>---</td>
<td>G319 cWW</td>
<td>C334 G319 cWW</td>
</tr>
<tr>
<td>G266 ---</td>
<td>---</td>
<td>A319 cSW</td>
<td>A1434 A319 cSW</td>
</tr>
<tr>
<td>--- ---</td>
<td>---</td>
<td>C320 cWW</td>
<td>A329 cWW</td>
</tr>
<tr>
<td>C267 cWW</td>
<td>G259 G267 cWW</td>
<td>A321 cWW</td>
<td>G332 A321 cWW</td>
</tr>
<tr>
<td>C268 cWW</td>
<td>G258 U268 cWW</td>
<td>C322 tWW</td>
<td>A329 tWW</td>
</tr>
<tr>
<td>C269 cWW</td>
<td>G257 G269 cWW</td>
<td>U233 tWW</td>
<td>A327 tWW</td>
</tr>
<tr>
<td>A270 cWW</td>
<td>U256 A270 cWW</td>
<td>A273 tSH</td>
<td>U317 tSH</td>
</tr>
<tr>
<td>C271 cWW</td>
<td>G255 C271 cWW</td>
<td>G324 tSH</td>
<td>A109 G324 tSH</td>
</tr>
<tr>
<td>C272 cWW</td>
<td>G254 C272 cWW</td>
<td>A325 ---</td>
<td>---</td>
</tr>
<tr>
<td>C273 cWW</td>
<td>U253 G273 cWW</td>
<td>A327 tWW</td>
<td>U323 tWW</td>
</tr>
<tr>
<td>C274 cWW</td>
<td>U252 A274 cWW</td>
<td>A329 tSH</td>
<td>U322 tSH</td>
</tr>
<tr>
<td>C275 cWW</td>
<td>G249 G275 cWW</td>
<td>C330 ---</td>
<td>---</td>
</tr>
<tr>
<td>C276 cWW</td>
<td>G248 G276 cWW</td>
<td>A331 bff</td>
<td>G111 bff</td>
</tr>
<tr>
<td>C277 cWW</td>
<td>G247 G277 cWW</td>
<td>G332 cWW</td>
<td>G321 A332 cWW</td>
</tr>
<tr>
<td>C278 cSS</td>
<td>A282 C277 cSS</td>
<td>G333 cWW</td>
<td>C320 U333 cWW</td>
</tr>
<tr>
<td>C279 cSS</td>
<td>A246 G278 cSS</td>
<td>G334 cWW</td>
<td>G319 C334 cWW</td>
</tr>
<tr>
<td>C280 ---</td>
<td>---</td>
<td>G334 cSS</td>
<td>A1434 A334 cSS</td>
</tr>
<tr>
<td>G281 tSH</td>
<td>A246 G281 tSH</td>
<td>C335 cWW</td>
<td>G318 C335 cWW</td>
</tr>
<tr>
<td>A282 tWW</td>
<td>A243 G282 tWW</td>
<td>C335 cSS</td>
<td>A1433 C335 cSS</td>
</tr>
<tr>
<td>A283 cSS</td>
<td>C277 G283 cSS</td>
<td>C336 cWW</td>
<td>G317 C336 cWW</td>
</tr>
<tr>
<td>A284 cSS</td>
<td>G247 A282 cSS</td>
<td>C337 cWW</td>
<td>G316 C337 cWW</td>
</tr>
<tr>
<td>A285 cWW</td>
<td>U253 G285 cWW</td>
<td>A338 cWW</td>
<td>G351 A338 cWW</td>
</tr>
<tr>
<td>A286 cWW</td>
<td>C241 G286 cWW</td>
<td>A339 cWW</td>
<td>G350 A339 cWW</td>
</tr>
<tr>
<td>A287 tWH</td>
<td>U119 U287 tWH</td>
<td>A340 cWW</td>
<td>G349 A340 cWW</td>
</tr>
<tr>
<td>A288 ---</td>
<td>---</td>
<td>C341 cWW</td>
<td>G348 A341 cWW</td>
</tr>
<tr>
<td>G289 cWW</td>
<td>C311 G289 cWW</td>
<td>C342 cWW</td>
<td>G347 A342 cWW</td>
</tr>
<tr>
<td>C290 cWW</td>
<td>G310 C290 cWW</td>
<td>U343 tSW</td>
<td>G346 U343 tSW</td>
</tr>
<tr>
<td>C291 cWW</td>
<td>G309 C291 cWW</td>
<td>A344 ---</td>
<td>A344 ---</td>
</tr>
<tr>
<td>G292 cWW</td>
<td>C308 G292 cWW</td>
<td>C345 ---</td>
<td>---</td>
</tr>
<tr>
<td>G293 cWW</td>
<td>G304 G293 cWW</td>
<td>A345 tWW</td>
<td>U343 A345 tWW</td>
</tr>
<tr>
<td>U294 cWW</td>
<td>A303 U294 cWW</td>
<td>G346 tWW</td>
<td>G346 A346 tWW</td>
</tr>
<tr>
<td>C295 cWW</td>
<td>G302 C295 cWW</td>
<td>G347 tWW</td>
<td>G347 A347 tWW</td>
</tr>
<tr>
<td>U296 cWW</td>
<td>G301 U296 cWW</td>
<td>A348 cWW</td>
<td>G348 A348 cWW</td>
</tr>
<tr>
<td>A297 tSH</td>
<td>A300 G297 tSH</td>
<td>A351 cWW</td>
<td>G351 A351 cWW</td>
</tr>
<tr>
<td>A298 ---</td>
<td>---</td>
<td>C352 ---</td>
<td>---</td>
</tr>
<tr>
<td>G299 cWW</td>
<td>G566 G299 cWW</td>
<td>A353 cSS</td>
<td>G113 A353 cSS</td>
</tr>
<tr>
<td>A300 tHS</td>
<td>G297 A300 tHS</td>
<td>G354 cWW</td>
<td>G58 C354 cWW</td>
</tr>
<tr>
<td>A300 cSW</td>
<td>U565 A300 cSW</td>
<td>G355 cWW</td>
<td>G57 C355 cWW</td>
</tr>
<tr>
<td>G301 cWW</td>
<td>U296 G301 cWW</td>
<td>A356 cWW</td>
<td>U56 A356 cWW</td>
</tr>
<tr>
<td>G302 cWW</td>
<td>C295 G302 cWW</td>
<td>G356 cWW</td>
<td>U296 G356 cWW</td>
</tr>
<tr>
<td>A303 cWW</td>
<td>U294 A303 cWW</td>
<td>G357 cWW</td>
<td>C54 G357 cWW</td>
</tr>
<tr>
<td>G304 cWW</td>
<td>G293 A304 cWW</td>
<td>G358 cWW</td>
<td>A55 G358 cWW</td>
</tr>
<tr>
<td>A305 ---</td>
<td>---</td>
<td>G358 cWW</td>
<td>A55 G358 cWW</td>
</tr>
<tr>
<td>A306 ---</td>
<td>---</td>
<td>U360 ntSS</td>
<td>A50 U360 ntSS</td>
</tr>
<tr>
<td>C307 ---</td>
<td>---</td>
<td>A361 tWW</td>
<td>U361 A361 tWW</td>
</tr>
<tr>
<td>A308 cWW</td>
<td>G292 C308 cWW</td>
<td>G362 cWW</td>
<td>U49 G362 cWW</td>
</tr>
<tr>
<td>G309 cWW</td>
<td>C291 A309 cWW</td>
<td>A363 ---</td>
<td>---</td>
</tr>
<tr>
<td>G310 cWW</td>
<td>A607 A309 cWW</td>
<td>A364 ---</td>
<td>---</td>
</tr>
<tr>
<td>U365 tWW</td>
<td>U49 U365 tWW</td>
<td>U49</td>
<td>U49</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G474 cWW</td>
<td>G458 cWW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G475 cWW</td>
<td>G457 cWW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G476 cWW</td>
<td>G456 cWW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G477 cWW</td>
<td>A478 ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C479 tWW</td>
<td>U443 A543</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C479 tWW</td>
<td>U479 cWW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U480 tWW</td>
<td>A452 U465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A481 ---</td>
<td>--- A481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A482 cSS</td>
<td>C370 A482</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C483 cWW</td>
<td>G460 C483</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G484 cWW</td>
<td>C449 U484</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C485 cSH</td>
<td>U486 ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U486 tWW</td>
<td>A448 U486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A497 tHS</td>
<td>G438 G496</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U498 tWW</td>
<td>U404 ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U498 tWW</td>
<td>A547 ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A499 cWW</td>
<td>U405 A499</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G500 cWW</td>
<td>C545 G500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G501 cWW</td>
<td>G544 C501</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G502 cWW</td>
<td>C543 A502</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C503 cWW</td>
<td>A510 C503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C504 cWW</td>
<td>G542 C504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G505 cWW</td>
<td>C526 G505</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G506 cWW</td>
<td>C525 G506</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C507 cSS</td>
<td>A10 C507</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C507 cWW</td>
<td>G524 C507</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C508 cSH</td>
<td>U508 U508</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A509 cSH</td>
<td>C508 A509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A509 cSS</td>
<td>C543 A509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A510 tSS</td>
<td>C503 A510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A510 cSS</td>
<td>G542 A510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A510 cSS</td>
<td>C543 A510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C511 cWW</td>
<td>G540 C511</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U512 cWW</td>
<td>A539 U512</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G513 cWW</td>
<td>G538 C513</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C514 cWW</td>
<td>G537 C514</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G515 cWW</td>
<td>C516 G515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U516 cSW</td>
<td>C519 U516</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U516 tWW</td>
<td>A533 U516</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C517 ---</td>
<td>--- C517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C518 cWS</td>
<td>U516 C519</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A520 tHS</td>
<td>G529 A520</td>
</tr>
<tr>
<td>G521 cWW</td>
<td>G528 G521</td>
</tr>
<tr>
<td>C522 cWW</td>
<td>G527 C522</td>
</tr>
<tr>
<td>A523 ---</td>
<td>A523 ---</td>
</tr>
<tr>
<td>G524 cWW</td>
<td>G507 C524</td>
</tr>
<tr>
<td>C525 cWW</td>
<td>G506 C525</td>
</tr>
<tr>
<td>G526 cWW</td>
<td>G505 C526</td>
</tr>
<tr>
<td>C527 cWW</td>
<td>G527 C527</td>
</tr>
<tr>
<td>A533 tSS</td>
<td>A535 A535</td>
</tr>
<tr>
<td>C528 cWW</td>
<td>G521 G528</td>
</tr>
<tr>
<td>G529 tHS</td>
<td>A520 G529</td>
</tr>
<tr>
<td>U530 ---</td>
<td>U530 ---</td>
</tr>
<tr>
<td>U531 ---</td>
<td>U531 ---</td>
</tr>
<tr>
<td>A532 ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>---</td>
<td>--- A532</td>
</tr>
<tr>
<td>A533 tWW</td>
<td>U516 A533</td>
</tr>
<tr>
<td>U534 ---</td>
<td>U534 ---</td>
</tr>
<tr>
<td>A535 tSS</td>
<td>G527 A535</td>
</tr>
<tr>
<td>C536 cWW</td>
<td>G513 C536</td>
</tr>
<tr>
<td>G537 cWW</td>
<td>C517 G537</td>
</tr>
<tr>
<td>G538 cWW</td>
<td>C513 G538</td>
</tr>
<tr>
<td>A539 ---</td>
<td>C519 U512</td>
</tr>
<tr>
<td>G540 cWW</td>
<td>C510 G540</td>
</tr>
<tr>
<td>G541 cWW</td>
<td>C504 G541</td>
</tr>
<tr>
<td>G542 cWW</td>
<td>C503 G542</td>
</tr>
<tr>
<td>A542 cWW</td>
<td>G543 cWW</td>
</tr>
<tr>
<td>A543 cWW</td>
<td>G542 cWW</td>
</tr>
<tr>
<td>A544 ---</td>
<td>--- A543</td>
</tr>
<tr>
<td>---</td>
<td>--- A493</td>
</tr>
<tr>
<td>A496 tWH</td>
<td>C449 U494</td>
</tr>
<tr>
<td>A497 tHS</td>
<td>G438 G496</td>
</tr>
<tr>
<td>U498 cWW</td>
<td>U404 ---</td>
</tr>
<tr>
<td>U498 cWW</td>
<td>A547 ---</td>
</tr>
<tr>
<td>---</td>
<td>--- G497</td>
</tr>
<tr>
<td>---</td>
<td>--- A498</td>
</tr>
<tr>
<td>---</td>
<td>--- A499</td>
</tr>
<tr>
<td>G540 cWW</td>
<td>C540 cWW</td>
</tr>
<tr>
<td>C541 cWW</td>
<td>C504 cWW</td>
</tr>
<tr>
<td>G542 cWW</td>
<td>C526 cWW</td>
</tr>
<tr>
<td>G546 cWW</td>
<td>C525 cWW</td>
</tr>
<tr>
<td>---</td>
<td>--- A10</td>
</tr>
<tr>
<td>---</td>
<td>--- C507</td>
</tr>
<tr>
<td>---</td>
<td>--- A507</td>
</tr>
<tr>
<td>---</td>
<td>--- C510</td>
</tr>
<tr>
<td>---</td>
<td>--- C510</td>
</tr>
<tr>
<td>---</td>
<td>--- A509</td>
</tr>
<tr>
<td>---</td>
<td>--- C509</td>
</tr>
<tr>
<td>---</td>
<td>--- A509</td>
</tr>
<tr>
<td>---</td>
<td>--- C503</td>
</tr>
<tr>
<td>---</td>
<td>--- A510</td>
</tr>
<tr>
<td>---</td>
<td>--- C510</td>
</tr>
<tr>
<td>C545 cWW</td>
<td>U884 A563</td>
</tr>
<tr>
<td>C546 cWW</td>
<td>U884 ---</td>
</tr>
<tr>
<td>---</td>
<td>--- C564</td>
</tr>
<tr>
<td>---</td>
<td>--- C564</td>
</tr>
<tr>
<td>---</td>
<td>--- C564</td>
</tr>
<tr>
<td>U565 cWS</td>
<td>A300 U565</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>---</td>
<td>--- U561</td>
</tr>
<tr>
<td>C566 cWW</td>
<td>G299 G566</td>
</tr>
<tr>
<td>G567 cWW</td>
<td>C883 G567</td>
</tr>
<tr>
<td>U571 cWW</td>
<td>A865 U571</td>
</tr>
<tr>
<td>A572 tSS</td>
<td>C19 A572</td>
</tr>
<tr>
<td>A573 cSS</td>
<td>C883 A573</td>
</tr>
<tr>
<td>A574 cSS</td>
<td>C883 A574</td>
</tr>
<tr>
<td>A575 cSS</td>
<td>A575 cSS</td>
</tr>
<tr>
<td>A576 cSS</td>
<td>A576 cSS</td>
</tr>
<tr>
<td>A577 cSS</td>
<td>A577 cSS</td>
</tr>
<tr>
<td>A578 cSS</td>
<td>A578 cSS</td>
</tr>
<tr>
<td>A579 cSS</td>
<td>A579 cSS</td>
</tr>
<tr>
<td>A580 cSS</td>
<td>A580 cSS</td>
</tr>
<tr>
<td>A581 cSS</td>
<td>A581 cSS</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>C470 cWW</td>
</tr>
<tr>
<td>---</td>
<td>C462 cWW</td>
</tr>
<tr>
<td>---</td>
<td>C459 cWW</td>
</tr>
<tr>
<td>---</td>
<td>U471 cWW</td>
</tr>
<tr>
<td>G474 cWW</td>
<td>C458 cWW</td>
</tr>
<tr>
<td>G475 cWW</td>
<td>C457 cWW</td>
</tr>
<tr>
<td>G476 cWW</td>
<td>C456 cWW</td>
</tr>
<tr>
<td>G477 cWW</td>
<td>C455 cWW</td>
</tr>
<tr>
<td>A478</td>
<td>---</td>
</tr>
<tr>
<td>C479 lWW</td>
<td>A453 cWW</td>
</tr>
<tr>
<td>C479 lWW</td>
<td>U479 cWW</td>
</tr>
<tr>
<td>U480 lWW</td>
<td>A452 cWW</td>
</tr>
<tr>
<td>U481 cHS</td>
<td>---</td>
</tr>
<tr>
<td>A482 cSS</td>
<td>C370 cWW</td>
</tr>
<tr>
<td>A482 cSS</td>
<td>G391 cWW</td>
</tr>
<tr>
<td>C483 cWW</td>
<td>G460 cWW</td>
</tr>
<tr>
<td>C484 cWW</td>
<td>G464 cWW</td>
</tr>
<tr>
<td>U485 cSH</td>
<td>U486 cWW</td>
</tr>
<tr>
<td>U486 lWW</td>
<td>A448 cWW</td>
</tr>
<tr>
<td>U486 cHS</td>
<td>G485 cWW</td>
</tr>
<tr>
<td>A487 tHS</td>
<td>A447 cWW</td>
</tr>
<tr>
<td>C488 cWW</td>
<td>G446 cWW</td>
</tr>
<tr>
<td>C489 cWW</td>
<td>G445 cWW</td>
</tr>
<tr>
<td>C490 cWW</td>
<td>C444 cWW</td>
</tr>
<tr>
<td>C491 cWW</td>
<td>G443 cWW</td>
</tr>
<tr>
<td>C492 cWW</td>
<td>G442 cWW</td>
</tr>
<tr>
<td>C494 tSH</td>
<td>A440 cWW</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U495 lWH</td>
<td>A439 cWH</td>
</tr>
<tr>
<td>U496 tHW</td>
<td>A437 cWW</td>
</tr>
<tr>
<td>A497 tHS</td>
<td>G438 cWH</td>
</tr>
<tr>
<td>U498 cWH</td>
<td>U404 cWH</td>
</tr>
<tr>
<td>U498 lWH</td>
<td>A547 cWH</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A499 cWH</td>
<td>U405 cWH</td>
</tr>
<tr>
<td>C500 cWW</td>
<td>C545 cWW</td>
</tr>
<tr>
<td>C501 cWW</td>
<td>G501 cWW</td>
</tr>
<tr>
<td>C502 cWW</td>
<td>C544 cWW</td>
</tr>
<tr>
<td>C503 cWW</td>
<td>A510 cWH</td>
</tr>
<tr>
<td>C504 cWW</td>
<td>G542 cWW</td>
</tr>
<tr>
<td>C505 cWW</td>
<td>C526 cWW</td>
</tr>
<tr>
<td>C506 cWW</td>
<td>C525 cWW</td>
</tr>
<tr>
<td>C507 cSS</td>
<td>A10 cWH</td>
</tr>
<tr>
<td>C507 cWW</td>
<td>G524 cWW</td>
</tr>
<tr>
<td>C508 cSH</td>
<td>A509 cSS</td>
</tr>
<tr>
<td>A509 cHS</td>
<td>C508 cWH</td>
</tr>
<tr>
<td>A509 cSS</td>
<td>C543 cWH</td>
</tr>
<tr>
<td>A510 tSS</td>
<td>C503 cWH</td>
</tr>
<tr>
<td>A510 cSS</td>
<td>C542 cWH</td>
</tr>
<tr>
<td>A510 cSS</td>
<td>C541 cWH</td>
</tr>
<tr>
<td>C515 cWW</td>
<td>C516 cWW</td>
</tr>
<tr>
<td>U516 cSH</td>
<td>U516 cWH</td>
</tr>
<tr>
<td>U516 lWH</td>
<td>A533 cWW</td>
</tr>
<tr>
<td>G517</td>
<td>---</td>
</tr>
<tr>
<td>C518</td>
<td>---</td>
</tr>
<tr>
<td>C519 cWS</td>
<td>U516 cWH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A520 tHS</td>
<td>G529 cWW</td>
</tr>
<tr>
<td>G521 cWW</td>
<td>G528 cWW</td>
</tr>
<tr>
<td>C522 cWW</td>
<td>G527 cWW</td>
</tr>
<tr>
<td>A523 cWW</td>
<td>---</td>
</tr>
<tr>
<td>G524 cWW</td>
<td>C507 cWW</td>
</tr>
<tr>
<td>G525 cWW</td>
<td>C506 cWW</td>
</tr>
<tr>
<td>C526 cWW</td>
<td>C505 cWW</td>
</tr>
<tr>
<td>G527 cWW</td>
<td>C522 cWW</td>
</tr>
<tr>
<td>A530 cWW</td>
<td>U521 cWW</td>
</tr>
<tr>
<td>A532</td>
<td>---</td>
</tr>
<tr>
<td>A532</td>
<td>---</td>
</tr>
<tr>
<td>U533 tWH</td>
<td>U516 cWH</td>
</tr>
<tr>
<td>U534</td>
<td>---</td>
</tr>
<tr>
<td>A535 cSS</td>
<td>G527 cWW</td>
</tr>
<tr>
<td>C536 cWW</td>
<td>C536 cWW</td>
</tr>
<tr>
<td>C537 cWW</td>
<td>C537 cWW</td>
</tr>
<tr>
<td>G538 cWW</td>
<td>C538 cWW</td>
</tr>
<tr>
<td>A539 cWW</td>
<td>U512 cWW</td>
</tr>
<tr>
<td>G540 cWW</td>
<td>G540 cWW</td>
</tr>
<tr>
<td>G541 cWW</td>
<td>G541 cWW</td>
</tr>
<tr>
<td>G542 cWW</td>
<td>G542 cWW</td>
</tr>
<tr>
<td>G543 cWW</td>
<td>G543 cWW</td>
</tr>
<tr>
<td>C544 cWW</td>
<td>G544 cWW</td>
</tr>
<tr>
<td>C545 cWW</td>
<td>G545 cWW</td>
</tr>
<tr>
<td>C546 cWW</td>
<td>G546 cWW</td>
</tr>
<tr>
<td>C547 cWW</td>
<td>G547 cWW</td>
</tr>
<tr>
<td>C548 cWW</td>
<td>C548 cWW</td>
</tr>
<tr>
<td>C549 cWW</td>
<td>C549 cWW</td>
</tr>
<tr>
<td>G550 cWW</td>
<td>C550 cWW</td>
</tr>
<tr>
<td>A551 cWW</td>
<td>U513 cWW</td>
</tr>
<tr>
<td>A552 cWW</td>
<td>U512 cWW</td>
</tr>
<tr>
<td>A553 cWW</td>
<td>U30 cWW</td>
</tr>
<tr>
<td>C554 cWW</td>
<td>G54 cWW</td>
</tr>
<tr>
<td>C555 cWW</td>
<td>G555 cWW</td>
</tr>
<tr>
<td>G556 cWW</td>
<td>G527 cWW</td>
</tr>
<tr>
<td>G557</td>
<td>---</td>
</tr>
<tr>
<td>G558 tSH</td>
<td>A26 cWW</td>
</tr>
<tr>
<td>A559</td>
<td>---</td>
</tr>
<tr>
<td>U560</td>
<td>---</td>
</tr>
<tr>
<td>U561</td>
<td>---</td>
</tr>
<tr>
<td>C562</td>
<td>U562 cWH</td>
</tr>
<tr>
<td>A563 cWW</td>
<td>U884 cWW</td>
</tr>
<tr>
<td>A564 cWW</td>
<td>U885 cWW</td>
</tr>
<tr>
<td>U565 cWS</td>
<td>A30 cWH</td>
</tr>
<tr>
<td>G566 cWW</td>
<td>G299 cWW</td>
</tr>
<tr>
<td>C567 cWW</td>
<td>C883 cWH</td>
</tr>
<tr>
<td>G568 cWW</td>
<td>A574 cWW</td>
</tr>
<tr>
<td>U571 cWW</td>
<td>A865 cWW</td>
</tr>
<tr>
<td>A572 cSS</td>
<td>C19 cWH</td>
</tr>
<tr>
<td>A573 cSS</td>
<td>C883 cWW</td>
</tr>
<tr>
<td>A574 cSS</td>
<td>G568 cWW</td>
</tr>
<tr>
<td>A574 cSS</td>
<td>A574 cWW</td>
</tr>
<tr>
<td>A574 cSS</td>
<td>C882 cWH</td>
</tr>
<tr>
<td>A574 cSS</td>
<td>C883 cWW</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>G575</td>
<td>cWW c880</td>
</tr>
<tr>
<td>G576</td>
<td>cWW c876</td>
</tr>
<tr>
<td>G577</td>
<td>cWW c764</td>
</tr>
<tr>
<td>C578</td>
<td>cWW A728</td>
</tr>
<tr>
<td>G579</td>
<td>cWW G763</td>
</tr>
<tr>
<td>U580</td>
<td>cWW G761</td>
</tr>
<tr>
<td>G581</td>
<td>tSH G760</td>
</tr>
<tr>
<td>U582</td>
<td>cWW A759</td>
</tr>
<tr>
<td>A583</td>
<td>tHS G758</td>
</tr>
<tr>
<td>G584</td>
<td>cWW U757</td>
</tr>
<tr>
<td>C585</td>
<td>cWW A616</td>
</tr>
<tr>
<td>C586</td>
<td>cWW G765</td>
</tr>
<tr>
<td>G587</td>
<td>---</td>
</tr>
<tr>
<td>G588</td>
<td>cWW C651</td>
</tr>
<tr>
<td>C589</td>
<td>cWW G660</td>
</tr>
<tr>
<td>C590</td>
<td>cWW G649</td>
</tr>
<tr>
<td>U591</td>
<td>cWW A648</td>
</tr>
<tr>
<td>G592</td>
<td>cWW G647</td>
</tr>
<tr>
<td>C593</td>
<td>cWW G636</td>
</tr>
<tr>
<td>G594</td>
<td>cWW G645</td>
</tr>
<tr>
<td>G595</td>
<td>ncHS C596</td>
</tr>
<tr>
<td>C596</td>
<td>cWW G646</td>
</tr>
<tr>
<td>C597</td>
<td>ncSH G595</td>
</tr>
<tr>
<td>G598</td>
<td>cWW C643</td>
</tr>
<tr>
<td>C599</td>
<td>cWW A640</td>
</tr>
<tr>
<td>G600</td>
<td>cWW G639</td>
</tr>
<tr>
<td>C600</td>
<td>cWW G638</td>
</tr>
<tr>
<td>C601</td>
<td>cWW G637</td>
</tr>
<tr>
<td>A602</td>
<td>cWW U636</td>
</tr>
<tr>
<td>U603</td>
<td>cWW G635</td>
</tr>
<tr>
<td>G604</td>
<td>cWW G634</td>
</tr>
<tr>
<td>G605</td>
<td>cWW G633</td>
</tr>
<tr>
<td>G606</td>
<td>cWW G632</td>
</tr>
<tr>
<td>G607</td>
<td>cWW G309</td>
</tr>
<tr>
<td>A608</td>
<td>tWS G292</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>C748</td>
<td>---</td>
</tr>
<tr>
<td>G750</td>
<td>cww</td>
</tr>
<tr>
<td>U751</td>
<td>cWW</td>
</tr>
<tr>
<td>G752</td>
<td>tSH</td>
</tr>
<tr>
<td>A753</td>
<td>tHS</td>
</tr>
<tr>
<td>C754</td>
<td>cWW</td>
</tr>
<tr>
<td>G755</td>
<td>cWW</td>
</tr>
<tr>
<td>C756</td>
<td>cWW</td>
</tr>
<tr>
<td>U757</td>
<td>cWW</td>
</tr>
<tr>
<td>G758</td>
<td>tSH</td>
</tr>
<tr>
<td>A759</td>
<td>tHS</td>
</tr>
<tr>
<td>C760</td>
<td>tHS</td>
</tr>
<tr>
<td>G761</td>
<td>cWW</td>
</tr>
<tr>
<td>C762</td>
<td>cWW</td>
</tr>
<tr>
<td>G763</td>
<td>cWW</td>
</tr>
<tr>
<td>C764</td>
<td>cWW</td>
</tr>
<tr>
<td>U765</td>
<td>---</td>
</tr>
<tr>
<td>G766</td>
<td>tWW</td>
</tr>
<tr>
<td>A767</td>
<td>---</td>
</tr>
<tr>
<td>C767</td>
<td>tWW</td>
</tr>
<tr>
<td>A768</td>
<td>---</td>
</tr>
<tr>
<td>G769</td>
<td>cWW</td>
</tr>
<tr>
<td>C770</td>
<td>tHS</td>
</tr>
<tr>
<td>G771</td>
<td>cWW</td>
</tr>
<tr>
<td>U772</td>
<td>cWW</td>
</tr>
<tr>
<td>G773</td>
<td>cWW</td>
</tr>
<tr>
<td>C774</td>
<td>cWW</td>
</tr>
<tr>
<td>U775</td>
<td>---</td>
</tr>
<tr>
<td>G776</td>
<td>---</td>
</tr>
<tr>
<td>A777</td>
<td>---</td>
</tr>
<tr>
<td>G778</td>
<td>---</td>
</tr>
<tr>
<td>C779</td>
<td>cWW</td>
</tr>
<tr>
<td>A780</td>
<td>tSH</td>
</tr>
<tr>
<td>C782</td>
<td>tHS</td>
</tr>
<tr>
<td>A783</td>
<td>---</td>
</tr>
<tr>
<td>G784</td>
<td>cWW</td>
</tr>
<tr>
<td>C785</td>
<td>tWW</td>
</tr>
<tr>
<td>A786</td>
<td>---</td>
</tr>
<tr>
<td>G787</td>
<td>---</td>
</tr>
<tr>
<td>C788</td>
<td>tSS</td>
</tr>
<tr>
<td>A789</td>
<td>---</td>
</tr>
<tr>
<td>G789</td>
<td>---</td>
</tr>
<tr>
<td>C790</td>
<td>---</td>
</tr>
<tr>
<td>A791</td>
<td>---</td>
</tr>
<tr>
<td>G792</td>
<td>---</td>
</tr>
<tr>
<td>C793</td>
<td>---</td>
</tr>
<tr>
<td>A794</td>
<td>---</td>
</tr>
<tr>
<td>G795</td>
<td>---</td>
</tr>
<tr>
<td>C796</td>
<td>cWW</td>
</tr>
<tr>
<td>A797</td>
<td>---</td>
</tr>
<tr>
<td>G798</td>
<td>cWW</td>
</tr>
<tr>
<td>C799</td>
<td>cWW</td>
</tr>
<tr>
<td>A799</td>
<td>---</td>
</tr>
<tr>
<td>G800</td>
<td>tSH</td>
</tr>
<tr>
<td>C801</td>
<td>---</td>
</tr>
<tr>
<td>A802</td>
<td>---</td>
</tr>
<tr>
<td>G803</td>
<td>cWW</td>
</tr>
<tr>
<td>C804</td>
<td>---</td>
</tr>
<tr>
<td>A805</td>
<td>---</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>A914 chW G21</td>
<td>A914 chW G21</td>
</tr>
<tr>
<td>G916 cWW U19</td>
<td>G916 cWW A19</td>
</tr>
<tr>
<td>G917 cWW A864</td>
<td>A918 cWW U17</td>
</tr>
<tr>
<td>A919 cWW A1080</td>
<td>A919 cWW A1080</td>
</tr>
<tr>
<td>A919 cWW U19</td>
<td>A919 cWW U19</td>
</tr>
<tr>
<td>G922 cWW C1395</td>
<td>G922 cWW C1395</td>
</tr>
<tr>
<td>G922 cWW A1398</td>
<td>G922 cWW A1398</td>
</tr>
<tr>
<td>G924 cWW A1502</td>
<td>G924 cWW A1502</td>
</tr>
<tr>
<td>G926 cWW U1390</td>
<td>G926 cWW U1390</td>
</tr>
<tr>
<td>G927 cWW C1389</td>
<td>G927 cWW C1389</td>
</tr>
<tr>
<td>G928 cWW C1388</td>
<td>G928 cWW C1388</td>
</tr>
<tr>
<td>G930 cWW C1387</td>
<td>G930 cWW C1387</td>
</tr>
<tr>
<td>G931 cWW C1386</td>
<td>G931 cWW C1386</td>
</tr>
<tr>
<td>G932 cWW C1385</td>
<td>G932 cWW C1385</td>
</tr>
<tr>
<td>A934 cWW U1380</td>
<td>A934 cWW U1380</td>
</tr>
<tr>
<td>A935 cWW C1383</td>
<td>A935 cWW C1383</td>
</tr>
<tr>
<td>A935 cWW C1383</td>
<td>A935 cWW C1383</td>
</tr>
<tr>
<td>A936 cWW G1379</td>
<td>A936 cWW G1379</td>
</tr>
<tr>
<td>A936 cWW G1379</td>
<td>A936 cWW G1379</td>
</tr>
<tr>
<td>A937 cWW C1378</td>
<td>A937 cWW C1378</td>
</tr>
<tr>
<td>A938 cWW C1376</td>
<td>A938 cWW C1376</td>
</tr>
<tr>
<td>A939 cWW C1344</td>
<td>A939 cWW C1344</td>
</tr>
<tr>
<td>A940 cWW C1343</td>
<td>A940 cWW C1343</td>
</tr>
<tr>
<td>A941 cWW C1342</td>
<td>A941 cWW C1342</td>
</tr>
<tr>
<td>A942 cWW C1341</td>
<td>A942 cWW C1341</td>
</tr>
<tr>
<td>A943 cWW C1340</td>
<td>A943 cWW C1340</td>
</tr>
<tr>
<td>A944 cWW A1339</td>
<td>A944 cWW A1339</td>
</tr>
<tr>
<td>A945 cWW A1236</td>
<td>A945 cWW A1236</td>
</tr>
<tr>
<td>A946 cWW A1235</td>
<td>A946 cWW A1235</td>
</tr>
<tr>
<td>A947 cWW A1234</td>
<td>A947 cWW A1234</td>
</tr>
<tr>
<td>A948 cWW A1233</td>
<td>A948 cWW A1233</td>
</tr>
<tr>
<td>A949 cWW A1232</td>
<td>A949 cWW A1232</td>
</tr>
<tr>
<td>A950 cWW A1231</td>
<td>A950 cWW A1231</td>
</tr>
<tr>
<td>A951 cSS C970</td>
<td>A951 cSS C970</td>
</tr>
<tr>
<td>A952 cWW C1230</td>
<td>A952 cWW C1230</td>
</tr>
<tr>
<td>A953 cWW U1229</td>
<td>A953 cWW U1229</td>
</tr>
<tr>
<td>A954 cWW C1228</td>
<td>A954 cWW C1228</td>
</tr>
<tr>
<td>A955 cWW A1225</td>
<td>A955 cWW A1225</td>
</tr>
<tr>
<td>A956 cWW U1224</td>
<td>A956 cWW U1224</td>
</tr>
<tr>
<td>A957 cWW C959</td>
<td>A957 cWW C959</td>
</tr>
<tr>
<td>A958 cSS C985</td>
<td>A958 cSS C985</td>
</tr>
<tr>
<td>A959 cSS C985</td>
<td>A959 cSS C985</td>
</tr>
<tr>
<td>A960 cSS G1597</td>
<td>A960 cSS G1597</td>
</tr>
<tr>
<td>A961 cWW A1201</td>
<td>A961 cWW A1201</td>
</tr>
<tr>
<td>A962 cWW G973</td>
<td>A962 cWW G973</td>
</tr>
<tr>
<td>A963 cWW G972</td>
<td>A963 cWW G972</td>
</tr>
<tr>
<td>A964 cSS A969</td>
<td>A964 cSS A969</td>
</tr>
<tr>
<td>A965 cSS A969</td>
<td>A965 cSS A969</td>
</tr>
<tr>
<td>A966 cSS A969</td>
<td>A966 cSS A969</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>T. th. (1j5e)</td>
<td>E. c. (2avy)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>A1015 cSS</td>
<td>C1218</td>
</tr>
<tr>
<td>A1016 cSS</td>
<td>G1088</td>
</tr>
<tr>
<td>A1016 tHS</td>
<td>G1013</td>
</tr>
<tr>
<td>A1016 cWW</td>
<td>C1217</td>
</tr>
<tr>
<td>A1017 cWW</td>
<td>G1012</td>
</tr>
<tr>
<td>A1018 cWW</td>
<td>G1011</td>
</tr>
<tr>
<td>A1019 cWW</td>
<td>G1010</td>
</tr>
<tr>
<td>U1020 cWW</td>
<td>G1009</td>
</tr>
<tr>
<td>G1021 cWW</td>
<td>C1008</td>
</tr>
<tr>
<td>G1022 cWW</td>
<td>A1007</td>
</tr>
<tr>
<td>G1024</td>
<td>---</td>
</tr>
<tr>
<td>G1026</td>
<td>---</td>
</tr>
<tr>
<td>C1027 cWW</td>
<td>G1034</td>
</tr>
<tr>
<td>---</td>
<td>G1024</td>
</tr>
<tr>
<td>C1028 cWW</td>
<td>G1032</td>
</tr>
<tr>
<td>C1029 cWW</td>
<td>G1030</td>
</tr>
<tr>
<td>C1030 cWW</td>
<td>G1030</td>
</tr>
<tr>
<td>C1030A cWW</td>
<td>---</td>
</tr>
<tr>
<td>C1030B cWW</td>
<td>---</td>
</tr>
<tr>
<td>G1031 cWW</td>
<td>C1030</td>
</tr>
<tr>
<td>G1032 cWW</td>
<td>C1030</td>
</tr>
<tr>
<td>G1033 cWW</td>
<td>C1028</td>
</tr>
<tr>
<td>A1035 cWW</td>
<td>---</td>
</tr>
<tr>
<td>C1037 cWW</td>
<td>---</td>
</tr>
<tr>
<td>C1037 cWW</td>
<td>C1031</td>
</tr>
<tr>
<td>C1038 cWW</td>
<td>G1033A</td>
</tr>
<tr>
<td>C1039 cWW</td>
<td>G1002</td>
</tr>
<tr>
<td>U1040 cWW</td>
<td>A1001</td>
</tr>
<tr>
<td>A1041 cWW</td>
<td>U1000</td>
</tr>
<tr>
<td>G1042 cWW</td>
<td>C999</td>
</tr>
<tr>
<td>C1043 cWW</td>
<td>G998</td>
</tr>
<tr>
<td>C1043 cWW</td>
<td>---</td>
</tr>
<tr>
<td>---</td>
<td>A1042</td>
</tr>
</tbody>
</table>

**Notes:**
- The table entries are placeholders for specific values or symbols that are not clearly visible in the image.
- The table structure and content are designed to represent a scientific or technical table.

**Page Reference:**
- Page 166
<p>| T.th. (1|5e) | E.c. (2avy) | T.th. (1|5e) | E.c. (2avy) |
|-------|--------|-------|--------|
| C1107 | cWW   | G1068 | C1107 | cWW   | G1068 | U1159 | --- | --- |
| G1108 | ---    | G1108 | ---    | G1108 | ---    |
| C1109 | ---    | C1109 | ---    | C1109 | ---    |
| A1110 | ---    | A1110 | ---    | A1110 | ---    |
| A1111 | ---    | A1111 | ---    | A1111 | ---    |
| C1112 | ---    | C1112 | ---    | C1112 | ---    |
| C1113 | cWW   | G1187 | C1113 | cWW   | G1187 | U1159 | --- | --- |
| C1114 | cWW   | G1186 | C1114 | cWW   | G1186 | G1160 | cWW | A1176 |
| G1115 | cWW   | G1185 | U1115 | cWW   | G1185 | G1160 | cWW | A1176 |
| G1116 | cWW   | G1184 | U1116 | cWW   | G1184 | G1160 | cWW | A1176 |
| C1117 | cWW   | G1155 | U1118 | cWW   | G1155 | A1169 | 1cWW | C1097 |
| C1118 | cWW   | G1154 | U1119 | cWW   | G1154 | A1169 | 1cWW | C1096 |
| G1119 | cWW   | G1153 | U1117 | cWW   | G1153 | A1169 | 1cWW | C1096 |
| G1120 | cWW   | A1121 | U1121 | cWW   | A1121 | A1172 | cWW | C1097 |
| C1121 | cWW   | A1121 | U1121 | cWW   | A1121 | A1172 | cWW | C1097 |
| U1121 | cWW   | A1121 | U1121 | cWW   | A1121 | A1172 | cWW | C1097 |
| G1123 | cWW   | C1129 | C1129 | cWW   | C1129 | C1129 | cWW | C1097 |
| G1124 | cWW   | C1130 | C1130 | cWW   | C1130 | C1130 | cWW | C1097 |
| G1125 | ---    | ---    | ---    | ---    | ---    | ---    | ---    | ---    |
| ---   | G1125 | cWW   | G1125 | cWW   | G1125 | U1120 | tWW | C1098 |
| ---   | A1280 | U1280 | U1280 | U1280 | U1280 | ---    | ---    | ---    |
| U1126 | cWW   | U1126 | U1126 | cWW   | U1126 | A1179 | tWW | U1126 |
| G1127 | cWW   | U1127 | G1127 | cWW   | U1127 | G1127 | cWW | U1127 |
| U1128 | cWW   | U1128 | U1128 | cWW   | U1128 | G1128 | cWW | U1128 |
| G1129 | cWW   | G1129 | G1129 | cWW   | G1129 | G1129 | cWW | U1128 |
| G1130 | cWW   | G1130 | G1130 | cWW   | G1130 | G1130 | cWW | U1128 |
| G1131 | cWW   | G1131 | G1131 | cWW   | G1131 | G1131 | cWW | U1128 |
| G1132 | cWW   | G1132 | G1132 | cWW   | G1132 | G1132 | cWW | U1128 |
| C1133 | cWW   | C1133 | C1133 | cWW   | C1133 | C1133 | cWW | U1128 |
| C1134 | cWW   | C1134 | C1134 | cWW   | C1134 | C1134 | cWW | U1128 |
| C1135 | cWW   | C1135 | C1135 | cWW   | C1135 | C1135 | cWW | U1128 |
| C1136 | cWW   | C1136 | C1136 | cWW   | C1136 | C1136 | cWW | U1128 |
| G1137 | cWW   | G1137 | G1137 | cWW   | G1137 | G1137 | cWW | U1128 |
| G1138 | cWW   | G1138 | G1138 | cWW   | G1138 | G1138 | cWW | U1128 |
| G1139 | cWW   | G1139 | G1139 | cWW   | G1139 | G1139 | cWW | U1128 |
| G1140 | cWW   | G1140 | G1140 | cWW   | G1140 | G1140 | cWW | U1128 |
| C1141 | cWW   | C1141 | C1141 | cWW   | C1141 | C1141 | cWW | U1128 |
| C1142 | cWW   | C1142 | C1142 | cWW   | C1142 | C1142 | cWW | U1128 |
| G1142 | cWW   | G1142 | G1142 | cWW   | G1142 | G1142 | cWW | U1128 |
| C1143 | cWW   | C1143 | C1143 | cWW   | C1143 | C1143 | cWW | U1128 |
| C1144 | cWW   | C1144 | C1144 | cWW   | C1144 | C1144 | cWW | U1128 |
| A1146 | cWW   | A1146 | A1146 | cWW   | A1146 | A1146 | cWW | U1128 |
| A1147 | ---    | ---    | ---    | ---    | ---    | ---    | ---    | ---    |
| U1148 | cWW   | U1148 | U1148 | cWW   | U1148 | U1150 | cWW | U1150 |
| G1149 | cWW   | G1149 | G1149 | cWW   | G1149 | U1150 | cWW | U1150 |
| C1150 | cWW   | C1150 | C1150 | cWW   | C1150 | C1150 | cWW | U1150 |
| A1151 | cWW   | A1151 | A1151 | cWW   | A1151 | A1151 | cWW | U1150 |
| A1152 | cWW   | A1152 | A1152 | cWW   | A1152 | A1152 | cWW | U1150 |
| G1153 | cWW   | G1153 | G1153 | cWW   | G1153 | C1153 | cWW | C1153 |
| C1154 | cWW   | C1154 | C1154 | cWW   | C1154 | C1154 | cWW | C1153 |
| G1155 | cWW   | G1155 | G1155 | cWW   | G1155 | G1155 | cWW | U1150 |
| G1156 | cWW   | G1156 | G1156 | cWW   | G1156 | G1156 | cWW | U1150 |
| A1157 | tWW   | G1157 | A1157 | tWW   | G1157 | A1157 | tWW | U1150 |
| C1158 | cWW   | C1158 | C1158 | cWW   | G1158 | A1158 | tWW | U1150 |
| A1159 | 1cWW | A1159 | A1159 | 1cWW | A1159 | A1159 | 1cWW | A1159 |</p>
<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
<th>T.th. (1j5e)</th>
<th>E.c. (2avy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1424 cWW</td>
<td>G1476 U1424 cWW A1476</td>
<td>G1476 cWW C1424 A1476 cWW U1424</td>
<td></td>
</tr>
<tr>
<td>U1425 cWW</td>
<td>G1475 U1425 cWW G1475</td>
<td>G1477 cWW G1423 U1477 cWW G1423</td>
<td></td>
</tr>
<tr>
<td>C1426 cWW</td>
<td>G1474 G1426 cWW U1474</td>
<td>G1478 cWW G1422 U1478 cWW G1422</td>
<td></td>
</tr>
<tr>
<td>U1427 cWW</td>
<td>A1473 C1427 cWW G1473</td>
<td>G1479 cWW G1421 C1479 cWW G1421</td>
<td></td>
</tr>
<tr>
<td>A1428 cWW</td>
<td>U1472 G1428 cWW U1472</td>
<td>G1480 cWW C1420 A1480 cWW U1420</td>
<td></td>
</tr>
<tr>
<td>C1429 cWW</td>
<td>G1471 A1429 cWW U1471</td>
<td>A1481 cWW G1419 U1481 cWW G1419</td>
<td></td>
</tr>
<tr>
<td>C1430 cWW</td>
<td>G1470 A1430 cWW U1470</td>
<td>G1482 tSH A1418 G1482 tSH A1418</td>
<td></td>
</tr>
<tr>
<td>C1431 cWW</td>
<td>G1469 A1431 tSH C1469</td>
<td>A1483 tSH C1475 A1483 tSH C1475</td>
<td></td>
</tr>
<tr>
<td>G1432 tSH</td>
<td>A1468 G1432 tSH A1468</td>
<td>C1484 cWW G1416 C1484 cWW G1416</td>
<td></td>
</tr>
<tr>
<td>A1433 cSS</td>
<td>C135 A1433 ncSS C135</td>
<td>U1485 cWW G1415 U1485 cWW G1415</td>
<td></td>
</tr>
<tr>
<td>A1433 mtHS</td>
<td>G1467 A1433 tHS C1467</td>
<td>G1486 cWW U1414 G1486 cWW U1414</td>
<td></td>
</tr>
<tr>
<td>A1434 cSS</td>
<td>C134 A1434 ncSS C134</td>
<td>G1487 cWW A1413 G1487 cWW A1413</td>
<td></td>
</tr>
<tr>
<td>A1434 tHS</td>
<td>G1467 A1434 nbf G1467</td>
<td>G1488 cWW G1412 G1488 cWW G1412</td>
<td></td>
</tr>
<tr>
<td>G1435 cWW</td>
<td>G1466 G1435 cWW C1466</td>
<td>G1489 cWW C1411 G1489 cWW C1411</td>
<td></td>
</tr>
<tr>
<td>U1436 cWW</td>
<td>C1465 U1436 cWW A1465</td>
<td>C1490 cWW G1410 U1490 cWW A1410</td>
<td></td>
</tr>
<tr>
<td>C1437 cWW</td>
<td>G1464 A1437 cWW U1464</td>
<td>G1491 cWW C1409 G1491 ncWW C1409</td>
<td></td>
</tr>
<tr>
<td>A1438 cSS</td>
<td>G1457 G1438 cSS U1457</td>
<td>A1492 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>A1439 cWW</td>
<td>G1458 A1439 cWW C1458</td>
<td>A1493 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>G1440 cWW</td>
<td>G1461 U1460 cSS G1461</td>
<td>G1494 cWW C1407 G1494 cWW C1407</td>
<td></td>
</tr>
<tr>
<td>A1441 tSH</td>
<td>A1460 --- --- --- ---</td>
<td>U1495 cWW U1406 U1495 tWH U1406</td>
<td></td>
</tr>
<tr>
<td>A1442 ---</td>
<td>--- --- --- ---</td>
<td>G1496 cWW C1405 G1496 cWW C1405</td>
<td></td>
</tr>
<tr>
<td>A1443 ---</td>
<td>--- --- --- ---</td>
<td>G1497 cWW C1404 G1497 cWW C1404</td>
<td></td>
</tr>
<tr>
<td>A1446 ---</td>
<td>--- --- --- ---</td>
<td>U1498 --- U1498 --- ---</td>
<td></td>
</tr>
<tr>
<td>G1447 cWW</td>
<td>C1459 --- --- A1456</td>
<td>A1499 cSS C1403 A1499 cSS C1403</td>
<td></td>
</tr>
<tr>
<td>--- G1461 A1441 tHS G1461</td>
<td>A1500 cWW C1402 A1500 ncWW C1402</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--- G1462 cWW C1462</td>
<td>C1501 cWW G1403 C1501 cWW G1403</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--- --- --- --- ---</td>
<td>--- --- --- --- ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1448 cWW</td>
<td>G1455 G1448 cWW G1455</td>
<td>G1505 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>G1450 tSS</td>
<td>G1453 G1453 tSS G1453</td>
<td>U1506 --- U1506</td>
<td></td>
</tr>
<tr>
<td>A1501 ---</td>
<td>U1450 G1450 tSS U1450</td>
<td>U1506 --- U1506</td>
<td></td>
</tr>
<tr>
<td>C1502 ---</td>
<td>C1452 --- --- --- ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1453 tWS</td>
<td>U1450 G1453 tWS U1450</td>
<td>G1511 cSS A767 G1511 cSS A767</td>
<td></td>
</tr>
<tr>
<td>G1454 cWW</td>
<td>C1449 G1454 cWW C1449</td>
<td>G1511 cSS U813 G1511 cSS U813</td>
<td></td>
</tr>
<tr>
<td>G1455 cWW</td>
<td>G1455 G1455 cWW C1448</td>
<td>G1511 cWW C1524 G1511 cWW C1524</td>
<td></td>
</tr>
<tr>
<td>G1459 cWW</td>
<td>G1457 A1457 cWW G1457</td>
<td>U1512 cWW G1523 U1512 cWW G1523</td>
<td></td>
</tr>
<tr>
<td>A1460 ---</td>
<td>G1457 cWW U1445</td>
<td>A1513 cWW U1523 A1513 cWW U1523</td>
<td></td>
</tr>
<tr>
<td>--- --- --- --- ---</td>
<td>--- --- --- --- ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1461 cWW</td>
<td>C1440 G1461 cWW U1440</td>
<td>G1514 cWW G1521 G1514 cWW G1521</td>
<td></td>
</tr>
<tr>
<td>A1462 cWW</td>
<td>C1439 G1462 cWW G1439</td>
<td>C1515 cWW G1520 G1515 cWW G1520</td>
<td></td>
</tr>
<tr>
<td>G1463 cWW</td>
<td>G1463 cWW U1444</td>
<td>C1516 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>G1464 cWW</td>
<td>U1436 A1456 cWW U1436</td>
<td>G1517 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>G1465 cWW</td>
<td>G1435 C1466 cWW G1435</td>
<td>G1518 ncSS G1405 A1518 ncSS G1405</td>
<td></td>
</tr>
<tr>
<td>G1466 cWW</td>
<td>G1466 cWW U1442</td>
<td>C1519 cWW G1409 A1519 ncSS G1409</td>
<td></td>
</tr>
<tr>
<td>G1467 tSH</td>
<td>A1434 C1467 nbf A1434</td>
<td>C1520 cWW C1505 C1520 cWW C1505</td>
<td></td>
</tr>
<tr>
<td>G1468 mtHS</td>
<td>A1433 C1467 tSH A1433</td>
<td>C1521 cWW C1514 C1521 cWW G1514</td>
<td></td>
</tr>
<tr>
<td>A1469 cWW</td>
<td>G1441 A1469 cWW G1441</td>
<td>U1522 cWW A1513 U1522 cWW A1513</td>
<td></td>
</tr>
<tr>
<td>G1470 cWW</td>
<td>G1430 U1470 cWW A1430</td>
<td>C1523 cWW A1512 U1523 cWW U1512</td>
<td></td>
</tr>
<tr>
<td>G1471 cWW</td>
<td>C1429 U1471 cWW A1429</td>
<td>G1524 cWW C1508 A1524 cWW A1508</td>
<td></td>
</tr>
<tr>
<td>U1472 cWW</td>
<td>A1428 U1472 cWW A1428</td>
<td>G1524 cWW U1520 C1524 cWW U1520</td>
<td></td>
</tr>
<tr>
<td>A1473 cWW</td>
<td>U1427 G1473 cWW C1427</td>
<td>A1531 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>G1474 cWW</td>
<td>C1426 U1474 cWW G1426</td>
<td>U1532 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>C1475 cWW</td>
<td>G1425 C1475 cWW U1425</td>
<td>C1533 --- --- --- ---</td>
<td></td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>E.c. (2avy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1534</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1539</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1540</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1541</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1542</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1543</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1544</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>C1533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>A1534</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix B—Comparison

**Spreadsheet for 16S Alignment Methods**

<table>
<thead>
<tr>
<th>T.th. (15e)</th>
<th>R3D Align (E.c. (2avy))</th>
<th>Basepair (E.c. (2avy))</th>
<th>ARTS (E.c. (2avy))</th>
<th>SARA (E.c. (2avy))</th>
<th>DIAL (E.c. (2avy))</th>
<th>NW (E.c. (2avy))</th>
</tr>
</thead>
<tbody>
<tr>
<td>U5</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>U5</td>
<td>U5</td>
<td>U5</td>
</tr>
<tr>
<td>G6</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>G6</td>
<td>G6</td>
<td>G6</td>
</tr>
<tr>
<td>U7</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>U7</td>
<td>U7</td>
<td>U7</td>
</tr>
<tr>
<td>A8</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>A8</td>
<td>A8</td>
<td>A8</td>
</tr>
<tr>
<td>G9</td>
<td>G9</td>
<td>G9</td>
<td>G9</td>
<td>G9</td>
<td>G9</td>
<td>G9</td>
</tr>
<tr>
<td>A10</td>
<td>A10</td>
<td>A10</td>
<td>A10</td>
<td>A10</td>
<td>A10</td>
<td>A10</td>
</tr>
<tr>
<td>A12</td>
<td>A12</td>
<td>A12</td>
<td>A12</td>
<td>A12</td>
<td>A12</td>
<td>A12</td>
</tr>
<tr>
<td>G11</td>
<td>G11</td>
<td>G11</td>
<td>G11</td>
<td>G11</td>
<td>G11</td>
<td>G11</td>
</tr>
<tr>
<td>U12</td>
<td>U12</td>
<td>U12</td>
<td>U12</td>
<td>U12</td>
<td>U12</td>
<td>U12</td>
</tr>
<tr>
<td>U13</td>
<td>U13</td>
<td>U13</td>
<td>U13</td>
<td>U13</td>
<td>U13</td>
<td>U13</td>
</tr>
<tr>
<td>A16</td>
<td>A16</td>
<td>A16</td>
<td>A16</td>
<td>A16</td>
<td>A16</td>
<td>A16</td>
</tr>
<tr>
<td>U17</td>
<td>U17</td>
<td>U17</td>
<td>U17</td>
<td>U17</td>
<td>U17</td>
<td>U17</td>
</tr>
<tr>
<td>C18</td>
<td>C18</td>
<td>C18</td>
<td>C18</td>
<td>C18</td>
<td>C18</td>
<td>C18</td>
</tr>
<tr>
<td>U20</td>
<td>U20</td>
<td>U20</td>
<td>U20</td>
<td>U20</td>
<td>U20</td>
<td>U20</td>
</tr>
<tr>
<td>A21</td>
<td>A21</td>
<td>A21</td>
<td>A21</td>
<td>A21</td>
<td>A21</td>
<td>A21</td>
</tr>
<tr>
<td>G22</td>
<td>G22</td>
<td>G22</td>
<td>G22</td>
<td>G22</td>
<td>G22</td>
<td>G22</td>
</tr>
<tr>
<td>C23</td>
<td>C23</td>
<td>C23</td>
<td>C23</td>
<td>C23</td>
<td>C23</td>
<td>C23</td>
</tr>
<tr>
<td>U24</td>
<td>U24</td>
<td>U24</td>
<td>U24</td>
<td>U24</td>
<td>U24</td>
<td>U24</td>
</tr>
<tr>
<td>A26</td>
<td>A26</td>
<td>A26</td>
<td>A26</td>
<td>A26</td>
<td>A26</td>
<td>A26</td>
</tr>
<tr>
<td>C27</td>
<td>C27</td>
<td>C27</td>
<td>C27</td>
<td>C27</td>
<td>C27</td>
<td>C27</td>
</tr>
<tr>
<td>U28</td>
<td>U28</td>
<td>U28</td>
<td>U28</td>
<td>U28</td>
<td>U28</td>
<td>U28</td>
</tr>
<tr>
<td>C29</td>
<td>C29</td>
<td>C29</td>
<td>C29</td>
<td>C29</td>
<td>C29</td>
<td>C29</td>
</tr>
<tr>
<td>G30</td>
<td>G30</td>
<td>G30</td>
<td>G30</td>
<td>G30</td>
<td>G30</td>
<td>G30</td>
</tr>
<tr>
<td>A31</td>
<td>A31</td>
<td>A31</td>
<td>A31</td>
<td>A31</td>
<td>A31</td>
<td>A31</td>
</tr>
<tr>
<td>G32</td>
<td>G32</td>
<td>G32</td>
<td>G32</td>
<td>G32</td>
<td>G32</td>
<td>G32</td>
</tr>
<tr>
<td>A33</td>
<td>A33</td>
<td>A33</td>
<td>A33</td>
<td>A33</td>
<td>A33</td>
<td>A33</td>
</tr>
<tr>
<td>C34</td>
<td>C34</td>
<td>C34</td>
<td>C34</td>
<td>C34</td>
<td>C34</td>
<td>C34</td>
</tr>
<tr>
<td>G35</td>
<td>G35</td>
<td>G35</td>
<td>G35</td>
<td>G35</td>
<td>G35</td>
<td>G35</td>
</tr>
<tr>
<td>C36</td>
<td>C36</td>
<td>C36</td>
<td>C36</td>
<td>C36</td>
<td>C36</td>
<td>C36</td>
</tr>
<tr>
<td>U37</td>
<td>U37</td>
<td>U37</td>
<td>U37</td>
<td>U37</td>
<td>U37</td>
<td>U37</td>
</tr>
<tr>
<td>G38</td>
<td>G38</td>
<td>G38</td>
<td>G38</td>
<td>G38</td>
<td>G38</td>
<td>G38</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
<td>G39 cWW C403</td>
</tr>
<tr>
<td>C40 cWW G402</td>
<td>G40 cWW G402</td>
<td>G40 cWW G402</td>
<td>G40 cWW G402</td>
<td>G40 cWW G402</td>
<td>G40 cWW G402</td>
<td>G40 cWW G402</td>
</tr>
<tr>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
<td>G41 cWW C401</td>
</tr>
<tr>
<td>C41 cWW G400</td>
<td>G41 cWW G400</td>
<td>G41 cWW G400</td>
<td>G41 cWW G400</td>
<td>G41 cWW G400</td>
<td>G41 cWW G400</td>
<td>G41 cWW G400</td>
</tr>
<tr>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
<td>C42 cWW A622</td>
</tr>
<tr>
<td>C43 cWW G399</td>
<td>C43 cWW G399</td>
<td>G399 cWW C43</td>
<td>G399 cWW C43</td>
<td>G399 cWW C43</td>
<td>G399 cWW C43</td>
<td>G399 cWW C43</td>
</tr>
<tr>
<td>G44 cWW C398</td>
<td>A44 cWW U398</td>
<td>A44 cWW U398</td>
<td>A44 cWW U398</td>
<td>A44 cWW U398</td>
<td>A44 cWW U398</td>
<td>A44 cWW U398</td>
</tr>
<tr>
<td>C45 cWW G396</td>
<td>G45 cWW C396</td>
<td>G45 cWW C396</td>
<td>G45 cWW C396</td>
<td>G45 cWW C396</td>
<td>G45 cWW C396</td>
<td>G45 cWW C396</td>
</tr>
<tr>
<td>G46 cWW C395</td>
<td>G46 cWW C395</td>
<td>C395 cWW G46</td>
<td>C395 cWW G46</td>
<td>C395 cWW G46</td>
<td>C395 cWW G46</td>
<td>C395 cWW G46</td>
</tr>
<tr>
<td>C47 cWW G361</td>
<td>G361 cWW C47</td>
<td>G361 cWW C47</td>
<td>G361 cWW C47</td>
<td>G361 cWW C47</td>
<td>G361 cWW C47</td>
<td>G361 cWW C47</td>
</tr>
<tr>
<td>C48 ---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U98 cWW G652</td>
<td>U98 cWW G652</td>
<td>G652 cWW U98</td>
<td>G652 cWW U98</td>
<td>G652 cWW U98</td>
<td>G652 cWW U98</td>
<td>G652 cWW U98</td>
</tr>
<tr>
<td>A55 cWW U165</td>
<td>A55 cWW U165</td>
<td>U165 cWW A55</td>
<td>U165 cWW A55</td>
<td>U165 cWW A55</td>
<td>U165 cWW A55</td>
<td>U165 cWW A55</td>
</tr>
<tr>
<td>U95 cWW G380</td>
<td>U95 cWW G380</td>
<td>G380 cWW U95</td>
<td>G380 cWW U95</td>
<td>G380 cWW U95</td>
<td>G380 cWW U95</td>
<td>G380 cWW U95</td>
</tr>
<tr>
<td>A60 cWW G107</td>
<td>A60 cWW G107</td>
<td>G107 cWW A60</td>
<td>G107 cWW A60</td>
<td>G107 cWW A60</td>
<td>G107 cWW A60</td>
<td>G107 cWW A60</td>
</tr>
<tr>
<td>U105 cWW G105</td>
<td>U105 cWW G105</td>
<td>G105 cWW U105</td>
<td>G105 cWW U105</td>
<td>G105 cWW U105</td>
<td>G105 cWW U105</td>
<td>G105 cWW U105</td>
</tr>
<tr>
<td>G63 cWW C104</td>
<td>G63 cWW C104</td>
<td>C104 cWW G63</td>
<td>C104 cWW G63</td>
<td>C104 cWW G63</td>
<td>C104 cWW G63</td>
<td>C104 cWW G63</td>
</tr>
<tr>
<td>U126 cWW G104</td>
<td>U126 cWW G104</td>
<td>G104 cWW U126</td>
<td>G104 cWW U126</td>
<td>G104 cWW U126</td>
<td>G104 cWW U126</td>
<td>G104 cWW U126</td>
</tr>
<tr>
<td>G65 cWW C103</td>
<td>G65 cWW C103</td>
<td>C103 cWW G65</td>
<td>C103 cWW G65</td>
<td>C103 cWW G65</td>
<td>C103 cWW G65</td>
<td>C103 cWW G65</td>
</tr>
<tr>
<td>U152 cWW G103</td>
<td>U152 cWW G103</td>
<td>G103 cWW U152</td>
<td>G103 cWW U152</td>
<td>G103 cWW U152</td>
<td>G103 cWW U152</td>
<td>G103 cWW U152</td>
</tr>
<tr>
<td>G66 cWW C102</td>
<td>G66 cWW C102</td>
<td>C102 cWW G66</td>
<td>C102 cWW G66</td>
<td>C102 cWW G66</td>
<td>C102 cWW G66</td>
<td>C102 cWW G66</td>
</tr>
<tr>
<td>U170 cWW G102</td>
<td>U170 cWW G102</td>
<td>G102 cWW U170</td>
<td>G102 cWW U170</td>
<td>G102 cWW U170</td>
<td>G102 cWW U170</td>
<td>G102 cWW U170</td>
</tr>
<tr>
<td>G68 cWW C100</td>
<td>G68 cWW C100</td>
<td>C100 cWW G68</td>
<td>C100 cWW G68</td>
<td>C100 cWW G68</td>
<td>C100 cWW G68</td>
<td>C100 cWW G68</td>
</tr>
<tr>
<td>U172 cWW G100</td>
<td>U172 cWW G100</td>
<td>G100 cWW U172</td>
<td>G100 cWW U172</td>
<td>G100 cWW U172</td>
<td>G100 cWW U172</td>
<td>G100 cWW U172</td>
</tr>
<tr>
<td>G69 cWW C99</td>
<td>G69 cWW C99</td>
<td>C99 cWW G69</td>
<td>C99 cWW G69</td>
<td>C99 cWW G69</td>
<td>C99 cWW G69</td>
<td>C99 cWW G69</td>
</tr>
<tr>
<td>T.th. (1Jse)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td>C103</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>C104</td>
<td>A147</td>
<td>U104</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>C105</td>
<td>A147</td>
<td>U104</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>C106</td>
<td>A147</td>
<td>U104</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>C107</td>
<td>A147</td>
<td>U104</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>C108</td>
<td>A147</td>
<td>U104</td>
<td>N185</td>
<td>A172</td>
<td>U103</td>
<td>N185</td>
</tr>
<tr>
<td>A109</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C110</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C111</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C112</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C113</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C114</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>A115</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C116</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C117</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C118</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C119</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C120</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C121</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C122</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C123</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C124</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C125</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C126</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C127</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C128</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C129</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C130</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C131</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C132</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C133</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C134</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C135</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C136</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C137</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C138</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C139</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C140</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C141</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C142</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C143</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C144</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C145</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C146</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C147</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C148</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C149</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C150</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C151</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C152</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C153</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C154</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C155</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C156</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C157</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C158</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C159</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>C160</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
<td>C172</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>R3D Align E.c. (2avy)</td>
<td>Basepair SARA E.c. (Zavy)</td>
<td>ARTS DIAL E.c. (Zavy)</td>
<td>NW E.c. (Zavy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------------------</td>
<td>--------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C160</td>
<td>G196</td>
<td>G163</td>
<td>G164</td>
<td>G163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U164</td>
<td>G196</td>
<td>G157</td>
<td>G164</td>
<td>G164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U165</td>
<td>G156</td>
<td>G156</td>
<td>G156</td>
<td>G156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G168</td>
<td>G155</td>
<td>G166</td>
<td>G156</td>
<td>G156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G167</td>
<td>C154</td>
<td>U167</td>
<td>U154</td>
<td>U154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G158</td>
<td>C153</td>
<td>G148</td>
<td>G153</td>
<td>G153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C169</td>
<td>A152</td>
<td>C169</td>
<td>A152</td>
<td>A152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U170</td>
<td>A151</td>
<td>U170</td>
<td>A151</td>
<td>A151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A171</td>
<td>C67</td>
<td>A171</td>
<td>C67</td>
<td>C67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A172</td>
<td>A172</td>
<td>A172</td>
<td>A172</td>
<td>A172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A173</td>
<td>A171</td>
<td>U170</td>
<td>U170</td>
<td>U170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A174</td>
<td>A174</td>
<td>A174</td>
<td>A174</td>
<td>A174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C175</td>
<td>G148</td>
<td>G148</td>
<td>G148</td>
<td>G148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C176</td>
<td>G147</td>
<td>G147</td>
<td>G147</td>
<td>G147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C177</td>
<td>G146</td>
<td>G146</td>
<td>G146</td>
<td>G146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C178</td>
<td>C144</td>
<td>C144</td>
<td>C144</td>
<td>C144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A179</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U180</td>
<td>A195</td>
<td>U180</td>
<td>A195</td>
<td>A195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U182</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G183</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G184</td>
<td>C184</td>
<td>C184</td>
<td>C184</td>
<td>C184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C185</td>
<td>C185</td>
<td>C185</td>
<td>C185</td>
<td>C185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G186</td>
<td>C186</td>
<td>C186</td>
<td>C186</td>
<td>C186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U187</td>
<td>C187</td>
<td>C187</td>
<td>C187</td>
<td>C187</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C188</td>
<td>C188</td>
<td>C188</td>
<td>C188</td>
<td>C188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A189</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A191</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A192</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A193</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A194</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A195</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A196</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A197</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A198</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A199</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td>A190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G180</td>
<td>C180</td>
<td>C180</td>
<td>C180</td>
<td>C180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G180</td>
<td>C180</td>
<td>C180</td>
<td>C180</td>
<td>C180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td>C181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td>C182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td>A193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td>A194</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td>A195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td>A196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td>A197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td>A198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td>A199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>G458 U498 G474 U458 cWW G474</td>
<td>U458 cWW G474</td>
<td>—</td>
<td>—</td>
<td>U459 cWW U473</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G459 U498 A461 A459</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A460</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C461</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G462</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A487</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G474 U458 C458 G474 cWW U458</td>
<td>C474 cWW U458</td>
<td>—</td>
<td>—</td>
<td>C475 cWW U473</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G475 U457 C457 cWW G457</td>
<td>C475 cWW U457</td>
<td>—</td>
<td>—</td>
<td>C476 cWW U473</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G476 U456 C456 uSS U456</td>
<td>U476 UWW U456</td>
<td>—</td>
<td>—</td>
<td>U477 UWW U473</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A478</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G479 U453 U479</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U480 U452 U480 UWW U452</td>
<td>U480 UWW U452</td>
<td>—</td>
<td>—</td>
<td>U481 UWW U473</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G481</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A482</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G484</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U486 C485 U486</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G487 C487</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C488</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A489</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G490</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U491 U461 U491 cWW U461</td>
<td>U491 cWW U461</td>
<td>—</td>
<td>—</td>
<td>U492 cWW U492</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C492</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C494</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G495 U495 U495</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A496</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G497</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U498 U454 U498 U404</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G499</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C500</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C501</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C502</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C503</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C504</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C505</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C506</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C507</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C508</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C509</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C510</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C511</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C512</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C513</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C514</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C515</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C516</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C517</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C518</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C519</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C520</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C521</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C522</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A523</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C507</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A510</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G511</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U512</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C513</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C514</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C515</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C516</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C517</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C518</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C519</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C520</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C521</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C522</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A523</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C507</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A510</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G511</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U512</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C513</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C514</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C515</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C516</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C517</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C518</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C519</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C520</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C521</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C522</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A523</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C507</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A510</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G511</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U512</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C513</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C514</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C515</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C516</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C517</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C518</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C519</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C520</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C521</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C522</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A523</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C507</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A510</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G511</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>U512</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C513</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C514</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C515</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C516</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C517</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C518</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C519</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C520</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C521</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C522</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>T.th. (1</td>
<td>5se)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G710</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G731</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>G742</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>A732</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
<td>G165</td>
</tr>
<tr>
<td>T. th. (15)E</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>U835</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>G836</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>G837</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>G838</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U839</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U841</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U848</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>G849</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C850</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C851</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C852</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C853</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C854</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C855</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C856</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C857</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U858</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U859</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>G860</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C861</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C862</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C863</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C864</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C865</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C866</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C867</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>C868</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U871</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U872</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U873</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U874</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U875</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U876</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U877</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U878</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U879</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U880</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U881</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U882</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U883</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U884</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U885</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U886</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U887</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U888</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U889</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U890</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U891</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U892</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U893</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U894</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U895</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U896</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U897</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>U898</td>
<td>U804</td>
<td>GB51</td>
<td>U835</td>
<td>GB51</td>
<td>GB43</td>
<td>GB51</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>G1017 cWW</td>
<td>U1017 cWW A1012</td>
<td>U1017 cWW A1012</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1018 cWW</td>
<td>U1018 cWW C1011</td>
<td>U1018 cWW C1011</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1019 cWW</td>
<td>U1019 cWW U1010</td>
<td>U1019 cWW U1010</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1020 cWW</td>
<td>U1020 cWW U1009</td>
<td>U1020 cWW U1009</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1021 cWW</td>
<td>U1021 cWW U1008</td>
<td>U1021 cWW U1008</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1022 cWW</td>
<td>U1022 cWW G1006</td>
<td>U1022 cWW G1006</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1023 cWW</td>
<td>——</td>
<td>U1023 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1024 cWW</td>
<td>——</td>
<td>U1024 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1025 cWW</td>
<td>——</td>
<td>U1025 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1026 cWW</td>
<td>——</td>
<td>U1026 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1027 cWW</td>
<td>——</td>
<td>U1027 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1028 cWW</td>
<td>——</td>
<td>U1028 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1029 cWW</td>
<td>——</td>
<td>U1029 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1030 cWW</td>
<td>——</td>
<td>U1030 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1031 cWW</td>
<td>——</td>
<td>U1031 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1032 cWW</td>
<td>——</td>
<td>U1032 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1033 cWW</td>
<td>——</td>
<td>U1033 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1034 cWW</td>
<td>——</td>
<td>U1034 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1035 cWW</td>
<td>——</td>
<td>U1035 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1036 cWW</td>
<td>——</td>
<td>U1036 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1037 cWW</td>
<td>——</td>
<td>U1037 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1038 cWW</td>
<td>——</td>
<td>U1038 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1039 cWW</td>
<td>——</td>
<td>U1039 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1040 cWW</td>
<td>——</td>
<td>U1040 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1041 cWW</td>
<td>——</td>
<td>U1041 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1042 cWW</td>
<td>——</td>
<td>U1042 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1043 cWW</td>
<td>——</td>
<td>U1043 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1044 cWW</td>
<td>——</td>
<td>U1044 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1045 cWW</td>
<td>——</td>
<td>U1045 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1046 cWW</td>
<td>——</td>
<td>U1046 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1047 cWW</td>
<td>——</td>
<td>U1047 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1048 cWW</td>
<td>——</td>
<td>U1048 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1049 cWW</td>
<td>——</td>
<td>U1049 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1050 cWW</td>
<td>——</td>
<td>U1050 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1051 cWW</td>
<td>——</td>
<td>U1051 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1052 cWW</td>
<td>——</td>
<td>U1052 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1053 cWW</td>
<td>——</td>
<td>U1053 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1054 cWW</td>
<td>——</td>
<td>U1054 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1055 cWW</td>
<td>——</td>
<td>U1055 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1056 cWW</td>
<td>——</td>
<td>U1056 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1057 cWW</td>
<td>——</td>
<td>U1057 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1058 cWW</td>
<td>——</td>
<td>U1058 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1059 cWW</td>
<td>——</td>
<td>U1059 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1060 cWW</td>
<td>——</td>
<td>U1060 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1061 cWW</td>
<td>——</td>
<td>U1061 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1062 cWW</td>
<td>——</td>
<td>U1062 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1063 cWW</td>
<td>——</td>
<td>U1063 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1064 cWW</td>
<td>——</td>
<td>U1064 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1065 cWW</td>
<td>——</td>
<td>U1065 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1066 cWW</td>
<td>——</td>
<td>U1066 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1067 cWW</td>
<td>——</td>
<td>U1067 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1068 cWW</td>
<td>——</td>
<td>U1068 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1069 cWW</td>
<td>——</td>
<td>U1069 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1070 cWW</td>
<td>——</td>
<td>U1070 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>G1071 cWW</td>
<td>——</td>
<td>U1071 cWW ——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>T.th. (1j5e)</td>
<td>R3D Align</td>
<td>Basepair</td>
<td>ARTS</td>
<td>SARA</td>
<td>DIAL</td>
<td>NW</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>E.c. (2avy)</td>
<td>E.c. (2avy)</td>
<td>E.c. (2avy)</td>
<td>E.c. (2avy)</td>
<td>E.c. (2avy)</td>
<td>E.c. (2avy)</td>
</tr>
<tr>
<td>C1129</td>
<td>G1134</td>
<td>---</td>
<td>G1134</td>
<td>A1130</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1130</td>
<td>ncWS</td>
<td>C1128</td>
<td>C1128</td>
<td>C1129</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1131</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1132</td>
<td>cWH</td>
<td>G1142</td>
<td>C1128</td>
<td>C1112</td>
<td>G1142</td>
<td>G1142</td>
</tr>
<tr>
<td>G1133</td>
<td>cWH</td>
<td>G1141</td>
<td>C1141</td>
<td>C1112</td>
<td>cWW</td>
<td>G1142</td>
</tr>
<tr>
<td>G1134</td>
<td>THW</td>
<td>C1129</td>
<td>C1144</td>
<td>---</td>
<td>C1129</td>
<td>---</td>
</tr>
<tr>
<td>G1135</td>
<td>cWH</td>
<td>C1140</td>
<td>C1140</td>
<td>C1112</td>
<td>G1142</td>
<td>C1128</td>
</tr>
<tr>
<td>U1135</td>
<td>cWH</td>
<td>G1138</td>
<td>G1138</td>
<td>G1134</td>
<td>C1144</td>
<td>C1128</td>
</tr>
<tr>
<td>U1136</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C1137</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1138</td>
<td>THW</td>
<td>U1135</td>
<td>U1135</td>
<td>C1138</td>
<td>G1134</td>
<td>C1138</td>
</tr>
<tr>
<td>G1139</td>
<td>cWH</td>
<td>G1142</td>
<td>G1142</td>
<td>C1139</td>
<td>cWW</td>
<td>G1142</td>
</tr>
<tr>
<td>G1140</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1141</td>
<td>cWH</td>
<td>G1133</td>
<td>G1133</td>
<td>C1134</td>
<td>cWH</td>
<td>G1134</td>
</tr>
<tr>
<td>G1142</td>
<td>cWH</td>
<td>G1132</td>
<td>G1132</td>
<td>C1134</td>
<td>cWH</td>
<td>G1134</td>
</tr>
<tr>
<td>G1143</td>
<td>cWH</td>
<td>G1139</td>
<td>G1139</td>
<td>C1134</td>
<td>cWH</td>
<td>G1134</td>
</tr>
<tr>
<td>U1143</td>
<td>sWH</td>
<td>C1127</td>
<td>C1127</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1144</td>
<td>cWH</td>
<td>C1128</td>
<td>C1128</td>
<td>---</td>
<td>C1132</td>
<td>C1132</td>
</tr>
<tr>
<td>U1145</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1146</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C1147</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1148</td>
<td>cWH</td>
<td>U1126</td>
<td>U1126</td>
<td>U1148</td>
<td>ncWS</td>
<td>U1135</td>
</tr>
<tr>
<td>U1149</td>
<td>cWH</td>
<td>C1124</td>
<td>C1124</td>
<td>C1149</td>
<td>cWH</td>
<td>C1124</td>
</tr>
<tr>
<td>G1149</td>
<td>cWH</td>
<td>A1280</td>
<td>A1280</td>
<td>C1149</td>
<td>cWH</td>
<td>A1280</td>
</tr>
<tr>
<td>U1150</td>
<td>cWH</td>
<td>U1123</td>
<td>U1123</td>
<td>C1150</td>
<td>cWH</td>
<td>U1123</td>
</tr>
<tr>
<td>A1151</td>
<td>cWH</td>
<td>A1122</td>
<td>A1122</td>
<td>A1151</td>
<td>cWH</td>
<td>A1122</td>
</tr>
<tr>
<td>A1152</td>
<td>cWH</td>
<td>U1121</td>
<td>U1121</td>
<td>A1152</td>
<td>cWH</td>
<td>U1121</td>
</tr>
<tr>
<td>C1153</td>
<td>cWH</td>
<td>G1120</td>
<td>G1120</td>
<td>G1153</td>
<td>cWH</td>
<td>G1120</td>
</tr>
<tr>
<td>C1154</td>
<td>cWH</td>
<td>G1139</td>
<td>G1139</td>
<td>G1154</td>
<td>cWH</td>
<td>G1139</td>
</tr>
<tr>
<td>C1155</td>
<td>cWH</td>
<td>U1118</td>
<td>U1118</td>
<td>G1155</td>
<td>cWH</td>
<td>U1118</td>
</tr>
<tr>
<td>G1156</td>
<td>sWH</td>
<td>A179</td>
<td>A179</td>
<td>G1156</td>
<td>cWH</td>
<td>A179</td>
</tr>
<tr>
<td>A1157</td>
<td>sWH</td>
<td>G1178</td>
<td>G1178</td>
<td>A1157</td>
<td>sWH</td>
<td>G1178</td>
</tr>
<tr>
<td>C1158</td>
<td>ncWS</td>
<td>G181</td>
<td>G181</td>
<td>C1158</td>
<td>ncWS</td>
<td>G181</td>
</tr>
<tr>
<td>C1159</td>
<td>cWH</td>
<td>G181</td>
<td>G181</td>
<td>A1159</td>
<td>cWH</td>
<td>G181</td>
</tr>
<tr>
<td>G1160</td>
<td>cWH</td>
<td>A176</td>
<td>A176</td>
<td>G1160</td>
<td>cWH</td>
<td>A176</td>
</tr>
<tr>
<td>G1161</td>
<td>cWH</td>
<td>G175</td>
<td>G175</td>
<td>G1161</td>
<td>cWH</td>
<td>G175</td>
</tr>
<tr>
<td>C1161</td>
<td>cWH</td>
<td>G117</td>
<td>G117</td>
<td>A1161</td>
<td>cWH</td>
<td>G117</td>
</tr>
<tr>
<td>C1162</td>
<td>sWH</td>
<td>G1175</td>
<td>G1175</td>
<td>A1162</td>
<td>sWH</td>
<td>G1175</td>
</tr>
<tr>
<td>C1163</td>
<td>cWH</td>
<td>U1173</td>
<td>U1173</td>
<td>G1163</td>
<td>cWH</td>
<td>U1173</td>
</tr>
<tr>
<td>C1164</td>
<td>cWH</td>
<td>C1172</td>
<td>C1172</td>
<td>C1164</td>
<td>cWH</td>
<td>C1172</td>
</tr>
<tr>
<td>A1165</td>
<td>cWH</td>
<td>U1171</td>
<td>U1171</td>
<td>A1165</td>
<td>cWH</td>
<td>U1171</td>
</tr>
<tr>
<td>C1166</td>
<td>sWH</td>
<td>A179</td>
<td>A179</td>
<td>A1166</td>
<td>sWH</td>
<td>A179</td>
</tr>
<tr>
<td>A1167</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1168</td>
<td>sWH</td>
<td>C1097</td>
<td>C1097</td>
<td>A1168</td>
<td>sWH</td>
<td>C1097</td>
</tr>
<tr>
<td>A1169</td>
<td>sWH</td>
<td>C1150</td>
<td>C1150</td>
<td>A1169</td>
<td>sWH</td>
<td>C1150</td>
</tr>
<tr>
<td>A1170</td>
<td>ncWS</td>
<td>G1089</td>
<td>G1089</td>
<td>A1170</td>
<td>ncWS</td>
<td>G1089</td>
</tr>
<tr>
<td>A1171</td>
<td>ncWS</td>
<td>G1089</td>
<td>G1089</td>
<td>A1171</td>
<td>ncWS</td>
<td>G1089</td>
</tr>
<tr>
<td>A1172</td>
<td>sWH</td>
<td>G1096</td>
<td>G1096</td>
<td>A1172</td>
<td>sWH</td>
<td>G1096</td>
</tr>
<tr>
<td>A1173</td>
<td>sWH</td>
<td>G1105</td>
<td>G1105</td>
<td>A1173</td>
<td>sWH</td>
<td>G1105</td>
</tr>
<tr>
<td>A1174</td>
<td>sWH</td>
<td>G1106</td>
<td>G1106</td>
<td>A1174</td>
<td>sWH</td>
<td>G1106</td>
</tr>
<tr>
<td>A1175</td>
<td>sWH</td>
<td>G1107</td>
<td>G1107</td>
<td>A1175</td>
<td>sWH</td>
<td>G1107</td>
</tr>
<tr>
<td>A1176</td>
<td>sWH</td>
<td>G1108</td>
<td>G1108</td>
<td>A1176</td>
<td>sWH</td>
<td>G1108</td>
</tr>
<tr>
<td>A1177</td>
<td>ncWS</td>
<td>G1146</td>
<td>G1146</td>
<td>A1177</td>
<td>ncWS</td>
<td>G1146</td>
</tr>
<tr>
<td>A1178</td>
<td>ncWS</td>
<td>C1158</td>
<td>C1158</td>
<td>A1178</td>
<td>ncWS</td>
<td>C1158</td>
</tr>
<tr>
<td>A1179</td>
<td>ncWS</td>
<td>G1156</td>
<td>G1156</td>
<td>A1179</td>
<td>ncWS</td>
<td>G1156</td>
</tr>
<tr>
<td>A1180</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1181</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1182</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1183</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1184</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1185</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1186</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1187</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1188</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1189</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

191
<table>
<thead>
<tr>
<th>T.th. (1j5e)</th>
<th>R3D Align</th>
<th>Basepair</th>
<th>ARTS</th>
<th>SARA</th>
<th>DIAL</th>
<th>NW</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1404</td>
<td>CWW</td>
<td>C1407</td>
<td>G1494</td>
<td>C1407</td>
<td>C1407</td>
<td>C1407</td>
</tr>
<tr>
<td>U1495</td>
<td>CWW</td>
<td>U1406</td>
<td>U1495</td>
<td>nWW</td>
<td>U1406</td>
<td>U1406</td>
</tr>
<tr>
<td>C1496</td>
<td>CWW</td>
<td>G1405</td>
<td>C1408</td>
<td>G1405</td>
<td>G1405</td>
<td>C1405</td>
</tr>
<tr>
<td>G1497</td>
<td>CWW</td>
<td>G1404</td>
<td>G1497</td>
<td>C1404</td>
<td>C1404</td>
<td>C1404</td>
</tr>
<tr>
<td>U1498</td>
<td>CWW</td>
<td>U1408</td>
<td>U1498</td>
<td>C1408</td>
<td>C1408</td>
<td>C1408</td>
</tr>
<tr>
<td>A1499</td>
<td>CWW</td>
<td>A1403</td>
<td>A1499</td>
<td>CWS</td>
<td>A1403</td>
<td>A1403</td>
</tr>
<tr>
<td>A1500</td>
<td>CWW</td>
<td>C1402</td>
<td>A1500</td>
<td>nCWW</td>
<td>C1402</td>
<td>C1402</td>
</tr>
<tr>
<td>C1501</td>
<td>CWW</td>
<td>G1401</td>
<td>C1501</td>
<td>G1401</td>
<td>G1401</td>
<td>G1401</td>
</tr>
<tr>
<td>A1502</td>
<td>---</td>
<td>---</td>
<td>A1502</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1503</td>
<td>---</td>
<td>---</td>
<td>A1503</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1504</td>
<td>CWW</td>
<td>C1399</td>
<td>G1504</td>
<td>CWW</td>
<td>C1399</td>
<td>C1399</td>
</tr>
<tr>
<td>G1505</td>
<td>---</td>
<td>---</td>
<td>G1505</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1506</td>
<td>---</td>
<td>---</td>
<td>U1506</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1507</td>
<td>CWW</td>
<td>U1528</td>
<td>A1507</td>
<td>CWW</td>
<td>U1528</td>
<td>U1528</td>
</tr>
<tr>
<td>G1508</td>
<td>CWW</td>
<td>C1527</td>
<td>A1508</td>
<td>CWW</td>
<td>C1527</td>
<td>C1527</td>
</tr>
<tr>
<td>C1509</td>
<td>CWW</td>
<td>G1526</td>
<td>C1509</td>
<td>G1526</td>
<td>G1526</td>
<td>G1526</td>
</tr>
<tr>
<td>G1510</td>
<td>CWW</td>
<td>G1525</td>
<td>G1510</td>
<td>G1525</td>
<td>G1525</td>
<td>G1525</td>
</tr>
<tr>
<td>G1511</td>
<td>CWW</td>
<td>A767</td>
<td>G1511</td>
<td>CWW</td>
<td>A767</td>
<td>A767</td>
</tr>
<tr>
<td>G1512</td>
<td>---</td>
<td>---</td>
<td>G1512</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1513</td>
<td>---</td>
<td>---</td>
<td>G1513</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1514</td>
<td>CWW</td>
<td>C1524</td>
<td>G1514</td>
<td>CWW</td>
<td>C1524</td>
<td>C1524</td>
</tr>
<tr>
<td>A1515</td>
<td>CWW</td>
<td>C1523</td>
<td>A1515</td>
<td>CWW</td>
<td>C1523</td>
<td>C1523</td>
</tr>
<tr>
<td>G1516</td>
<td>---</td>
<td>---</td>
<td>G1516</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1517</td>
<td>---</td>
<td>---</td>
<td>G1517</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1518</td>
<td>CWW</td>
<td>G1409</td>
<td>A1518</td>
<td>CWW</td>
<td>G1409</td>
<td>G1409</td>
</tr>
<tr>
<td>A1519</td>
<td>---</td>
<td>---</td>
<td>A1519</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1520</td>
<td>CWW</td>
<td>C1513</td>
<td>G1520</td>
<td>CWW</td>
<td>C1513</td>
<td>C1513</td>
</tr>
<tr>
<td>C1521</td>
<td>CWW</td>
<td>G1512</td>
<td>C1521</td>
<td>CWW</td>
<td>G1512</td>
<td>G1512</td>
</tr>
<tr>
<td>G1522</td>
<td>CWW</td>
<td>G1511</td>
<td>G1522</td>
<td>CWW</td>
<td>G1511</td>
<td>G1511</td>
</tr>
<tr>
<td>G1523</td>
<td>CWW</td>
<td>A767</td>
<td>G1523</td>
<td>CWW</td>
<td>A767</td>
<td>A767</td>
</tr>
<tr>
<td>G1524</td>
<td>---</td>
<td>---</td>
<td>G1524</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1525</td>
<td>---</td>
<td>---</td>
<td>G1525</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1526</td>
<td>CWW</td>
<td>C1509</td>
<td>G1526</td>
<td>CWW</td>
<td>C1509</td>
<td>C1509</td>
</tr>
<tr>
<td>C1527</td>
<td>CWW</td>
<td>A1508</td>
<td>C1527</td>
<td>CWW</td>
<td>A1508</td>
<td>A1508</td>
</tr>
<tr>
<td>G1528</td>
<td>CWW</td>
<td>A1507</td>
<td>G1528</td>
<td>CWW</td>
<td>A1507</td>
<td>A1507</td>
</tr>
<tr>
<td>G1529</td>
<td>CWW</td>
<td>C187</td>
<td>G1529</td>
<td>CWW</td>
<td>C187</td>
<td>C187</td>
</tr>
<tr>
<td>G1530</td>
<td>---</td>
<td>---</td>
<td>G1530</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1531</td>
<td>---</td>
<td>---</td>
<td>G1531</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1532</td>
<td>CWW</td>
<td>U1532</td>
<td>U1532</td>
<td>CWW</td>
<td>U1532</td>
<td>U1532</td>
</tr>
<tr>
<td>C1533</td>
<td>CWW</td>
<td>C1533</td>
<td>C1533</td>
<td>CWW</td>
<td>C1533</td>
<td>C1533</td>
</tr>
<tr>
<td>A1534</td>
<td>---</td>
<td>---</td>
<td>A1534</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C1535</td>
<td>---</td>
<td>---</td>
<td>C1535</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1536</td>
<td>---</td>
<td>---</td>
<td>G1536</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1537</td>
<td>---</td>
<td>---</td>
<td>G1537</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A1538</td>
<td>---</td>
<td>---</td>
<td>A1538</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>G1539</td>
<td>---</td>
<td>---</td>
<td>G1539</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1540</td>
<td>---</td>
<td>---</td>
<td>U1540</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1541</td>
<td>---</td>
<td>---</td>
<td>U1541</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>U1542</td>
<td>---</td>
<td>---</td>
<td>U1542</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C1543</td>
<td>CWW</td>
<td>C1533</td>
<td>C1543</td>
<td>CWW</td>
<td>C1533</td>
<td>C1533</td>
</tr>
<tr>
<td>U1544</td>
<td>---</td>
<td>---</td>
<td>U1544</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
## Appendix C—R3D Align RNaseP Alignment

<table>
<thead>
<tr>
<th>1U9S</th>
<th>1NBS:B</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>P7</td>
<td>P7</td>
</tr>
<tr>
<td>G76 cWW</td>
<td>G86 cWW</td>
</tr>
<tr>
<td>C234</td>
<td>C239</td>
</tr>
<tr>
<td>P7</td>
<td>P7</td>
</tr>
<tr>
<td>G76 cWW</td>
<td>G87 cWW</td>
</tr>
<tr>
<td>C234</td>
<td>C239</td>
</tr>
<tr>
<td>P7</td>
<td>P7</td>
</tr>
<tr>
<td>G77 cWW</td>
<td>G88 cWW</td>
</tr>
<tr>
<td>C233</td>
<td>C238</td>
</tr>
<tr>
<td>P7</td>
<td>P7</td>
</tr>
<tr>
<td>U78 cWW</td>
<td>A89 cWW</td>
</tr>
<tr>
<td>A232</td>
<td>U236</td>
</tr>
<tr>
<td>P7</td>
<td>P7</td>
</tr>
<tr>
<td>G79 cWW</td>
<td>G90 cWW</td>
</tr>
<tr>
<td>C231</td>
<td>C235</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C80 cWW</td>
<td>C91 cWW</td>
</tr>
<tr>
<td>G94</td>
<td>G112</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C80 cSS</td>
<td>C91 cSS</td>
</tr>
<tr>
<td>A207</td>
<td>A208</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C81 cWW</td>
<td>C92 cWW</td>
</tr>
<tr>
<td>G93</td>
<td>G111</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C81 cSS</td>
<td>C92 cSS</td>
</tr>
<tr>
<td>A207</td>
<td>A208</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>A82 cWW</td>
<td>U93 cWW</td>
</tr>
<tr>
<td>A110</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G83 cWW</td>
<td>A94 cWW</td>
</tr>
<tr>
<td>U109</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G84 cWW</td>
<td>G95 cWW</td>
</tr>
<tr>
<td>C108</td>
<td></td>
</tr>
<tr>
<td>U85</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>A86</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>A87</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>C88</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>G89</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>P8</td>
</tr>
<tr>
<td>C96 cWW</td>
<td>G107</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>G97 tSH</td>
</tr>
<tr>
<td>...</td>
<td>A106</td>
</tr>
<tr>
<td>...</td>
<td>A98</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>A99</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>G100 cWS</td>
</tr>
<tr>
<td>...</td>
<td>A103</td>
</tr>
<tr>
<td>...</td>
<td>U101</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>C102</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>A103 cWS</td>
</tr>
<tr>
<td>...</td>
<td>G100</td>
</tr>
<tr>
<td>...</td>
<td>U104</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>A105</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>A106 tHS</td>
</tr>
<tr>
<td>...</td>
<td>G97</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C90 cWW</td>
<td>C108 cWW</td>
</tr>
<tr>
<td>G84</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>C91 cWW</td>
<td>U109 cWW</td>
</tr>
<tr>
<td>G83</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>U92 cWW</td>
<td>A110 cWW</td>
</tr>
<tr>
<td>A94</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G93 cWW</td>
<td>C81</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G111 cWW</td>
<td>C92</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G94 cWW</td>
<td>C80</td>
</tr>
<tr>
<td>P8</td>
<td>P8</td>
</tr>
<tr>
<td>G112 cWW</td>
<td>C91</td>
</tr>
<tr>
<td>Page</td>
<td>G94</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>P8</td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>G95</td>
</tr>
<tr>
<td>P9</td>
<td>C96</td>
</tr>
<tr>
<td>P9</td>
<td>G97</td>
</tr>
<tr>
<td>P9</td>
<td>G98</td>
</tr>
<tr>
<td>P9</td>
<td>C104</td>
</tr>
<tr>
<td>U101</td>
<td></td>
</tr>
<tr>
<td>A102</td>
<td></td>
</tr>
</tbody>
</table>

**Page 9**

<table>
<thead>
<tr>
<th>Page</th>
<th>G94</th>
<th>tSS</th>
<th>A208</th>
<th>Page</th>
<th>G94</th>
<th>tSS</th>
<th>A208</th>
</tr>
</thead>
<tbody>
<tr>
<td>P9</td>
<td>C105</td>
<td>cWW</td>
<td>G98</td>
<td>P9</td>
<td>C118</td>
<td>cWW</td>
<td>G125</td>
</tr>
<tr>
<td>P9</td>
<td>C106</td>
<td>cWW</td>
<td>G97</td>
<td>P9</td>
<td>U128</td>
<td>cWW</td>
<td>A115</td>
</tr>
<tr>
<td>P9</td>
<td>C107</td>
<td>cWW</td>
<td>C96</td>
<td>P9</td>
<td>G129</td>
<td>cWW</td>
<td>C114</td>
</tr>
<tr>
<td>A108</td>
<td></td>
<td></td>
<td></td>
<td>A130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>C109</td>
<td>cWW</td>
<td>G95</td>
<td>P9</td>
<td>C131</td>
<td>cWW</td>
<td>G113</td>
</tr>
<tr>
<td>P10</td>
<td>G110</td>
<td>cWW</td>
<td>C230</td>
<td>P10</td>
<td>G132</td>
<td>cWW</td>
<td>C234</td>
</tr>
<tr>
<td>P10</td>
<td>G111</td>
<td>cWW</td>
<td>C229</td>
<td>P10</td>
<td>G133</td>
<td>cWW</td>
<td>C233</td>
</tr>
<tr>
<td>A112</td>
<td>tSS</td>
<td>A227</td>
<td>C134</td>
<td>tWH</td>
<td>A231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A113</td>
<td></td>
<td></td>
<td></td>
<td>A114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A114</td>
<td></td>
<td></td>
<td></td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Page 10**

<table>
<thead>
<tr>
<th>Page</th>
<th>G108</th>
<th>tSS</th>
<th>A135</th>
<th>Page</th>
<th>G136</th>
<th>cWW</th>
<th>U174</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A140</td>
<td>cWW</td>
<td>A167</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A142</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A143</td>
<td>cWW</td>
<td>U165</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>C144</td>
<td>cWW</td>
<td>C164</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>C145</td>
<td>cWW</td>
<td>G163</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>C146</td>
<td>cWW</td>
<td>G162</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A139</td>
<td>tSS</td>
<td>A170</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A140</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>U147</td>
<td>tWH</td>
<td>A160</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>U147</td>
<td>cSS</td>
<td>A207</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>A148</td>
<td>tSH</td>
<td>C149</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>C149</td>
<td>cHS</td>
<td>A148</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>G150</td>
<td>cWW</td>
<td>U159</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td>P10.1</td>
<td>U151</td>
<td>cWW</td>
<td>A158</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>C152</td>
<td>cWW</td>
<td>G157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U153</td>
<td>tSW</td>
<td>G156</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U154</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>C155</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G156</td>
<td>tWS</td>
<td>U153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A158</td>
<td>cWW</td>
<td>U151</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U159</td>
<td>cWW</td>
<td>G150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A160</td>
<td>tHW</td>
<td>U147</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A160</td>
<td>tWW</td>
<td>A206</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U161</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G162</td>
<td>tSS</td>
<td>A208</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G163</td>
<td>cWW</td>
<td>C146</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>C164</td>
<td>cWW</td>
<td>G144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U165</td>
<td>cWW</td>
<td>A143</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G166</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A167</td>
<td>tHH</td>
<td>A141</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G168</td>
<td>cSH</td>
<td>U169</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U169</td>
<td>cHS</td>
<td>G168</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A170</td>
<td>tHS</td>
<td>A139</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U171</td>
<td>cWW</td>
<td>A138</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>C172</td>
<td>cWW</td>
<td>G137</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>C173</td>
<td>cWW</td>
<td>G136</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U174</td>
<td>cWW</td>
<td>A135</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>U175</td>
<td>tWW</td>
<td>A179</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>G176</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A177</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G110</td>
<td>P10.1</td>
<td>A178</td>
<td>tSS</td>
<td>G132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10.1</td>
<td>A179</td>
<td>tWW</td>
<td>U175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>G115</td>
<td>cWW</td>
<td>C228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>G180</td>
<td>cWW</td>
<td>C232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>U116</td>
<td>cWW</td>
<td>A227</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>U181</td>
<td>cWW</td>
<td>A231</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>G117</td>
<td>cWW</td>
<td>C224</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>G182</td>
<td>cWW</td>
<td>U228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C118</td>
<td>cWW</td>
<td>G223</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C183</td>
<td>cWW</td>
<td>G227</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C119</td>
<td>tSS</td>
<td>A173</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C184</td>
<td>tSS</td>
<td>A221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C119</td>
<td>cWW</td>
<td>G222</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>C184</td>
<td>cWW</td>
<td>G226</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A120</td>
<td>tHH</td>
<td>A172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A185</td>
<td>nWH</td>
<td>G220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>C121</td>
<td>---</td>
<td>J11/12</td>
<td>C186</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A122</td>
<td>tWH</td>
<td>A126</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A187</td>
<td>ntWH</td>
<td>A191</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A122</td>
<td>cSS</td>
<td>U170</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A187</td>
<td>nSH</td>
<td>U218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>G123</td>
<td>tSW</td>
<td>A174</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>G188</td>
<td>tSW</td>
<td>A222</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A124</td>
<td>---</td>
<td>J11/12</td>
<td>U189</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>G125</td>
<td>cSS</td>
<td>A129</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>J11/12</td>
<td>G190</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A126</td>
<td>tHW</td>
<td>A122</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A191</td>
<td>nWH</td>
<td>A187</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A126</td>
<td>tWS</td>
<td>G171</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A191</td>
<td>ntWS</td>
<td>G219</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A126</td>
<td>ncSH</td>
<td>A127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A191</td>
<td>cSH</td>
<td>C192</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A127</td>
<td>cHS</td>
<td>A126</td>
<td>J11/12</td>
<td>C192</td>
<td>cWH</td>
<td>G224</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>J11/12</td>
<td>A127</td>
<td></td>
<td></td>
<td>J11/12</td>
<td>C192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>G128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A129</td>
<td>cSS</td>
<td>G125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>G130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J11/12</td>
<td>A131</td>
<td>cSH</td>
<td>C132</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C132</td>
<td>cHS</td>
<td>A131</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C132</td>
<td>cWW</td>
<td>G168</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C133</td>
<td>cWW</td>
<td>G167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G134</td>
<td>cWW</td>
<td>C164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C135</td>
<td>cWW</td>
<td>G163</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C135</td>
<td>cSS</td>
<td>A185</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C136</td>
<td>cWW</td>
<td>G162</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C136</td>
<td>cSS</td>
<td>A184</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C139</td>
<td>cWW</td>
<td>G161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G140</td>
<td>cWW</td>
<td>C160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G141</td>
<td>cWW</td>
<td>U158</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C142</td>
<td>cWW</td>
<td>G157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J11/12</td>
<td>G193</td>
<td>cWW</td>
<td>C223</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J11/12</td>
<td>A194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J11/12</td>
<td>A195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J11/12</td>
<td>G196</td>
<td>cSW</td>
<td>U189</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>U197</td>
<td>cWW</td>
<td>A216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C143</td>
<td>cWW</td>
<td>G156</td>
<td>P12</td>
<td>C198</td>
<td>cWW</td>
<td>G215</td>
</tr>
<tr>
<td>P12</td>
<td>G144</td>
<td>cWW</td>
<td>C155</td>
<td>P12</td>
<td>U199</td>
<td>cWW</td>
<td>A214</td>
</tr>
<tr>
<td>P12</td>
<td>G145</td>
<td>cWW</td>
<td>C154</td>
<td>P12</td>
<td>C200</td>
<td>cWW</td>
<td>G213</td>
</tr>
<tr>
<td>P12</td>
<td>G146</td>
<td>cWW</td>
<td>C153</td>
<td>P12</td>
<td>A201</td>
<td>cWW</td>
<td>U212</td>
</tr>
<tr>
<td>P12</td>
<td>G147</td>
<td>cWW</td>
<td>C152</td>
<td>P12</td>
<td>C202</td>
<td>cWW</td>
<td>G211</td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>U203</td>
<td>cWW</td>
<td>G210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>A204</td>
<td>cWW</td>
<td>U209</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G205</td>
<td>cSH</td>
<td>A208</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A206</td>
<td>cWW</td>
<td>A160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A207</td>
<td>cSS</td>
<td>U147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A208</td>
<td>cSS</td>
<td>C146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A208</td>
<td>cSS</td>
<td>G162</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A208</td>
<td>cHS</td>
<td>G205</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>U209</td>
<td>cWW</td>
<td>A204</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>G210</td>
<td>cWW</td>
<td>U203</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C152</td>
<td>cWW</td>
<td>G147</td>
<td>P12</td>
<td>G211</td>
<td>cWW</td>
<td>C202</td>
</tr>
<tr>
<td>P12</td>
<td>C153</td>
<td>cWW</td>
<td>G146</td>
<td>P12</td>
<td>U212</td>
<td>cWW</td>
<td>A201</td>
</tr>
<tr>
<td>P12</td>
<td>C154</td>
<td>cWW</td>
<td>G145</td>
<td>P12</td>
<td>G213</td>
<td>cWW</td>
<td>C200</td>
</tr>
<tr>
<td>P12</td>
<td>C155</td>
<td>cWW</td>
<td>G144</td>
<td>P12</td>
<td>A214</td>
<td>cWW</td>
<td>U199</td>
</tr>
<tr>
<td>P12</td>
<td>G156</td>
<td>cWW</td>
<td>C143</td>
<td>P12</td>
<td>G215</td>
<td>cWW</td>
<td>C198</td>
</tr>
<tr>
<td>P12</td>
<td>G157</td>
<td>cWW</td>
<td>C142</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>U158</td>
<td>cWW</td>
<td>G141</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G159</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>C160</td>
<td>cWW</td>
<td>G140</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G161</td>
<td>cWW</td>
<td>C139</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G162</td>
<td>cWW</td>
<td>C136</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G163</td>
<td>cWW</td>
<td>C135</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G163</td>
<td>tSS</td>
<td>A185</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A165</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G167</td>
<td>cWW</td>
<td>C133</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>G168</td>
<td>cWW</td>
<td>C132</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>---</td>
<td>---</td>
<td>P12</td>
<td>A216</td>
<td>cWW</td>
<td>U197</td>
<td></td>
</tr>
<tr>
<td>J12/11</td>
<td>G169</td>
<td>---</td>
<td>J12/11</td>
<td>G217</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>J12/11</td>
<td>U170</td>
<td>cSS</td>
<td>A122</td>
<td>J12/11</td>
<td>U218</td>
<td>cSS</td>
<td>A187</td>
</tr>
<tr>
<td>J12/11</td>
<td>U170</td>
<td>tWH</td>
<td>A174</td>
<td>J12/11</td>
<td>U218</td>
<td>tWH</td>
<td>A222</td>
</tr>
<tr>
<td>J12/11</td>
<td>G171</td>
<td>tSW</td>
<td>A126</td>
<td>J12/11</td>
<td>G219</td>
<td>ntSW</td>
<td>A191</td>
</tr>
<tr>
<td>J12/11</td>
<td>A172</td>
<td>tHH</td>
<td>A120</td>
<td>J12/11</td>
<td>G220</td>
<td>ntHH</td>
<td>A185</td>
</tr>
<tr>
<td>J12/11</td>
<td>A173</td>
<td>tSS</td>
<td>C119</td>
<td>J12/11</td>
<td>A221</td>
<td>tSS</td>
<td>C184</td>
</tr>
<tr>
<td>J12/11</td>
<td>A174</td>
<td>tWS</td>
<td>G123</td>
<td>J12/11</td>
<td>A222</td>
<td>tWS</td>
<td>G188</td>
</tr>
<tr>
<td>J12/11</td>
<td>A174</td>
<td>tSS</td>
<td>A129</td>
<td>J12/11</td>
<td>A222</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>J12/11</td>
<td>A174</td>
<td>tHW</td>
<td>U170</td>
<td>J12/11</td>
<td>A222</td>
<td>tHW</td>
<td>U218</td>
</tr>
<tr>
<td>J12/11</td>
<td>A174</td>
<td>---</td>
<td>J12/11</td>
<td>A222</td>
<td>cSH</td>
<td>C223</td>
<td></td>
</tr>
<tr>
<td>J12/11</td>
<td>C175</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G176</td>
<td>cWW</td>
<td>C194</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G177</td>
<td>cHW</td>
<td>G188</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G177</td>
<td>cWW</td>
<td>C193</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C178</td>
<td>cWW</td>
<td>A192</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G179</td>
<td>cWW</td>
<td>C191</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G180</td>
<td>cWW</td>
<td>C190</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>G181</td>
<td>cWW</td>
<td>C189</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G182</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U183</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A184</td>
<td>cSS</td>
<td>C136</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A185</td>
<td>cSS</td>
<td>C135</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A185</td>
<td>tSS</td>
<td>G163</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G186</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A187</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G188</td>
<td>cWH</td>
<td>G177</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C189</td>
<td>cWW</td>
<td>G181</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C190</td>
<td>cWW</td>
<td>G180</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C191</td>
<td>cWW</td>
<td>G179</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>A192</td>
<td>cWW</td>
<td>C178</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C193</td>
<td>cWW</td>
<td>G177</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>C194</td>
<td>cWW</td>
<td>G176</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G195</td>
<td>cWW</td>
<td>C219</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C196</td>
<td>cWW</td>
<td>G217</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C197</td>
<td>cWW</td>
<td>G216</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>U198</td>
<td>cWW</td>
<td>G215</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G199</td>
<td>cWW</td>
<td>C214</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G200</td>
<td>cWW</td>
<td>C213</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C201</td>
<td>cWW</td>
<td>G212</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C202</td>
<td>cWW</td>
<td>G211</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>U203</td>
<td>cWW</td>
<td>A210</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G204</td>
<td>cWW</td>
<td>C209</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G205</td>
<td>tSH</td>
<td>A208</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C206</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A207</td>
<td>cSS</td>
<td>C81</td>
<td>---</td>
<td>C92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A208</td>
<td>cSS</td>
<td>C80</td>
<td>---</td>
<td>C91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A208</td>
<td>tSS</td>
<td>G94</td>
<td>---</td>
<td>G112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A208</td>
<td>tHS</td>
<td>G205</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C209</td>
<td>cWW</td>
<td>G204</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A210</td>
<td>cWW</td>
<td>U203</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G211</td>
<td>cWW</td>
<td>C202</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G212</td>
<td>cWW</td>
<td>C201</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C213</td>
<td>cWW</td>
<td>G200</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C214</td>
<td>cWW</td>
<td>G199</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G215</td>
<td>cWW</td>
<td>U198</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G216</td>
<td>cWW</td>
<td>C197</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G217</td>
<td>cWW</td>
<td>C196</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>G217</td>
<td>cSW</td>
<td>A220</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C219</td>
<td>cWW</td>
<td>G195</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>A220</td>
<td>cWS</td>
<td>G217</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>C221</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>G222</td>
<td>cWW</td>
<td>C119</td>
<td>P11</td>
<td>G226</td>
<td>cWW</td>
<td>C184</td>
</tr>
<tr>
<td>P11</td>
<td>G223</td>
<td>cWW</td>
<td>C118</td>
<td>P11</td>
<td>G227</td>
<td>cWW</td>
<td>C183</td>
</tr>
<tr>
<td>P11</td>
<td>C224</td>
<td>cWW</td>
<td>G117</td>
<td>P11</td>
<td>U228</td>
<td>cWW</td>
<td>G182</td>
</tr>
<tr>
<td>P10</td>
<td>A225</td>
<td>---</td>
<td>A229</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>A232</td>
<td>cWW</td>
<td>U78</td>
<td>P7</td>
<td>U236</td>
<td>cWW</td>
<td>A89</td>
</tr>
<tr>
<td>P7</td>
<td>C233</td>
<td>cWW</td>
<td>G77</td>
<td>P7</td>
<td>C237</td>
<td>cWW</td>
<td>G88</td>
</tr>
<tr>
<td>P7</td>
<td>C234</td>
<td>cWW</td>
<td>G76</td>
<td>---</td>
<td>C87</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>