SPRING 2008 REPORT

BY MAURICIO ESGUERRA

A report submitted to
Dr. Wilma K. Olson
Rutgers, The State University of New Jersey

New Brunswick, New Jersey
May, 2008

ABSTRACT OF THE REPORT

Spring 2008 Report

by Mauricio Esguerra
Project Director: Wilma K. Olson

This report carries over the Fall of 2007 report in its first Chapter. It adds to it more data on clustering analysis of the space of torsion angles, and base-step parameters for RNA dinucleotide steps. It also contains a Chapter with the results presented at the IMA Workshop in Minnessota in 2007. The last Chapter has some initial tables and figures on the analysis of an RNA Dataset put together by Dr. Yurong Xin

Table of Contents

Chapter 1
Classification of RNA Conformations

The problem of classification of the space of conformations of RNA is not new, see for example, Olson 1972 [1], Saenger 1984 [2], and Gautheret 1993 [3]. This problem had only been addressed by a few researchers before the turn of the twenty first century, but starting in the year 2000 a vast amount of RNA structural information has become available with the elucidation of the structure of the 30S small ribosomal subunit of *Thermus thermophilus*, a bacterial ribosome [4, 5], and the 50S large ribosomal subunit of *Haloarcula marismortui*, an archaeal [i] ribosome [6].

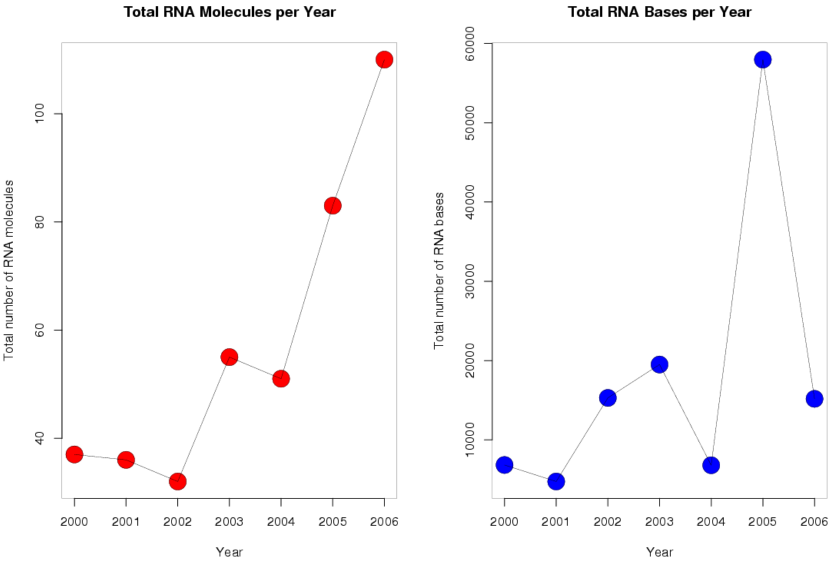

Figure 1.1: **Left:** Total number of RNA structures solved yearly by X-Ray crystallography between 2000 and 2006. **Right:** Total number of RNA bases added to the PDB database between 2000 and 2006.

Between 1972 and 2000 a total of 132 RNA structures with resolution greater than 3 Å, and comprising around 5500 nucleotide bases were found in the Protein Data Bank (PDB), and between 2000 and today a total of 460 RNA structures comprising around 140000 nucleotide bases have been found. That is, the increase in information due to the solution of large RNA structures is two orders of magnitude as pointed out by Noller [7] in 2005. Looking at the growth of RNA structural information from 2000 until today it's important to point out that, although the total number of RNA structures deposited in the PDB shows exponential growth (see left panel in Figure 1.1), the total number of RNA bases does not show a well defined trend (see right panel in Figure 1.1). This is due to the size preponderance of ribosomal structures. That is, in 2005 nineteen ribosomal structures were deposited in the PDB, whereas in 2006 only four were deposited. So, even though interest in RNA seems to be growing since ribosomal structures have become available in 2000, and two Nobel prizes were awarded for work

[i] I emphasize the phylogeny of rRNA's here since there is an ongoing discussion among biologists on whether archaea are closer to prokaryotes, or to eukaryotes.

in RNA in 2006, along with the exciting possibilities of deciphering even larger RNA virus structures, still the growth of the RNA structural field is far from that of proteins if weighed by the growth of RNA structural information in the past seven years. This fact might just be due to the smaller sizes and structural diversity of RNA molecules, which, as can be seen in Figure 1.2, is restricted to “compact” nucleotide ranges. A representative example of these characteristic ranges can be seen in Table 1.1 for structures larger than 300 bases.

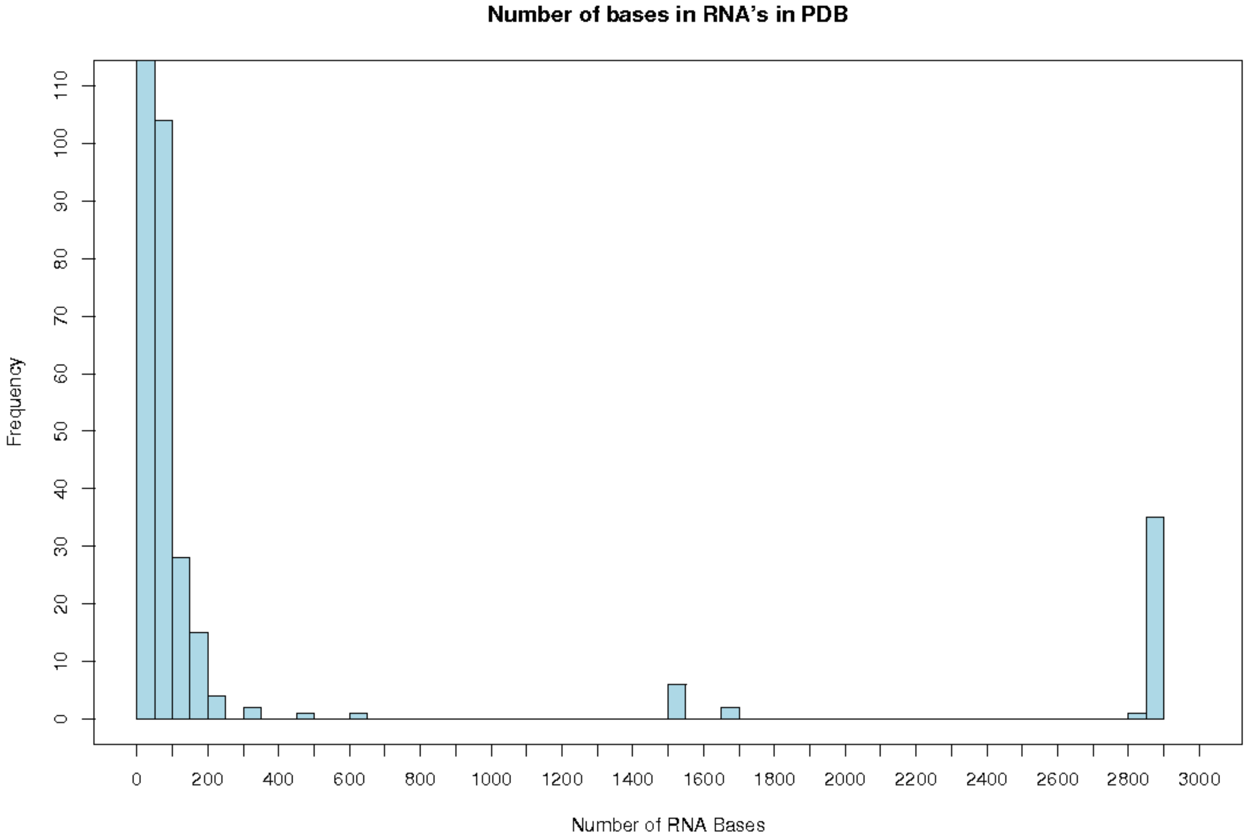

Figure 1.2: Frequency of nucleotide bases in RNA molecules found in the PDB classified by the size of RNA molecules. We define the size as the total number of nucleotide bases present per molecule.

Analysis of RNA conformational information contained in this structural data can be divided into three main perspectives: an atom based perspective; a bond based perspective; and a third, as yet unexplored to our knowledge, rigid-body based perspective. In the atom based perspective, either direct comparison of backbone atom positions is made [8], or a comparison of distances between a reduced set of atoms taken from the nucleotide backbone, sugar, and base [9]. The bond based perspective is divided into three main categories; the first considers the consecutive covalent bonds in the RNA backbone and the glycosidic bond between the sugar and base, that is, six backbone torsion angles and one glycosidic torsion angle [8, 10, 11, 12, 13]; or alternatively the pseudo-bonds between consecutive P and C4′ atoms and the resulting pseudo-torsion angles η and θ [1, 14, 15, 16]. The third category considers the networks of horizontal hydrogen bonding patterns coming from a definition of interacting edge boundaries in the nucleotide bases [17, 18, 19]. In this report we review one category of the bond based perspective. Namely we review the case where the covalent bonds between backbone atoms give rise to torsion angle space. We also make a first study of the rigid body based perspective using clustering analysis.

PDBID	Structure Name	Phylogenetic Group	Number of bases	Year
1l8v	Mutant of P4-P6 Domain of Group I Intron	Eukaryote	314	2002
1fg0	Central Loop in Domain V of 23S rRNA	Archaea	499	2000
2nz4	GlmS Ribozyme	Eukaryote	604	2006
1xmq	30S rRNA	Bacteria	1522	2004
1ffk	50S rRNA Subunit	Archaea	2828	2000

Table 1.1: Some large RNA structures ($>$300 bases) elucidated in the last 7 years.

1.1 Dinucleotide Torsion Angles

The covalent bond based perspective as mentioned in the previous section gives rise to six backbone torsion angles and one glycosidic torsion angle. This heptaparametric space has been the subject of several recent studies of RNA dinucleotide steps. Richardson and collaborators [10] have applied van der Waals radius filtering techniques on a database of 8636 nucleic acid residues from RNA X-Ray structures with resolution of 3.0 or better grouping all structures in 42 conformers which they refer to as rotamers. Berman et al. [12] reduced the data space of the large subunit of the *Haloarcula Marismortui* ribosome using Fourier transform filtering. Hershkovitz et al. [11] defined lower and upper bounds of torsion angle values by "binning" in one dimension. Pyle et al. [16] reduced the heptaparametric space to a biparametric one, defining a virtual bond between consecutive O and C4′ atoms in the RNA backbone.

Hershkovitz and collaborators [13] took a first step towards integrating clustering analysis formally in the study of RNA backbone torsion angles. In particular, they used the k-means partitional clustering algorithm. [ii] It's important to note that Hershkovitz et al. reduce the data set of all torsion angles in rRNA's large subunit, using their binning approach prior to k-means clustering.

Clustering analysis can be divided into two main methodologies, namely, hierarchical clustering and partitional clustering [20]. We have used particular cases of both methodologies to investigate thoroughly if "biased" [iii] data reduction is needed, as has been suggested by various authors [12, 13], or if the use of clustering analysis alone can be used to find in an efficient and clear manner subsets of RNA conformational space which possess a clear structural meaning.

1.1.1 Partitional Clustering for Torsion Angles

For partitional clustering the k-means algorithm as implemented in the software package **R** [21] was employed. [iv]

We consider the 2753 base-steps of the 23S subunit of the ribosome as vectors of seven dimensions composed of the previously mentioned backbone torsion angles α, β, γ, δ, ϵ, ζ, and the glycosidic

[ii] For a reason that still eludes the author, instead of using the more general and familiar terminology of clustering analysis, they refer to this method as if it was not a clustering analysis method. In one case they call their method scalar quantization, when one torsion angle at a time is clustered. They call it vector quantization when they want to cluster groups of more than one torsion angle at a time.

[iii] By biased data reduction we mean that a reduction of the whole data set is done by taking into account a particular bias imposed by us. This bias can be structural or sequence based.

[iv] Another partitional algorithm that could readily be used in the future is pam (partitioning around medoids). For now we've determined the average silhouette width, which is an analogous quantity to the average distortion **D**, (which will be defined later in the text) and plotted this value against the number of clusters as can be seen in Figure 1.4

torsion angle χ. The **R** software package has implemented four different k-means algorithms; *Hartigan-Wong*; *Lloyd*; *Forgy*; and *MacQueen*. For the data set used, that is, the large subunit of the ribosome (PDB code 1jj2), no noticeable differences were found with the four k-means algorithms when we group the data into two partitions, as can be seen in Figures 1.16 through 1.19.

One of the problems with k-means is that the number of clusters is not an emergent property of the data set but a parameter that has to be given to the algorithm. This problem has been given a good amount of attention in the statistical analysis community, in the area of clustering analysis. Hershkovitz et al. [13] use one method which is common in clustering analysis and find a so-called distortion measure **D**, also called the "within clusters sum of squares" in the more common clustering analysis area. That is, the squares of the elements of each cluster is found and then added. This quantity can be plotted against the number of clusters k that were selected as can be seen in Figure 1.3. The "optimal" number of clusters corresponds to the value of k where **D** becomes constant, which in Figure 1.3 is around 60. Where interestingly Figure1.3 is very similar to Figure 8 in the paper of Hershkovitz et al. [13], where the **D** value becomes constant also around 60. The main difference between our plot and theirs is that they exclude dinucleotide steps which are close to the A-type conformation. They further state that the A-type steps amount to over 60% of the data, meaning that the remaining 40% accounts for the majority of the conformational diversity of the space of torsion angles, since there is not a significant change in the value where **D** becomes constant.

There are more methods to determine the optimal amount of partitions that a data set can be split into. For example, one can also do another partitioning around medoids (PAM), and find an analogous quantity to the within clusters sum of squares, such quantity is called the average silhouette width, and the optimal number of clusters is that which maximizes this quantity. In Figure 1.4 we can see that the maximum corresponds to $k = 3$, nonetheless, we also see various local maxima, for example in $k = 16, 22, 28, 31, 36, 45, 48, 51, 57$, it's interesting to point out that in a very recent preprint by Berman et al. [22] they review the work on RNA backbone conformations and summarize that different research groups find 32, 37, and 42 discrete RNA conformations.

Another way of selecting k is just by visual inspection of the data. In our case if we take any of the scatterplots in Figures 1.16 thru 1.19, we can imagine that in some of them the data seems to be clustered around eight groups[v]. We choose eight groups also because this is the result that Duarte and Pyle obtain for their classification based on RNA pseudotorsions [14]. Following this argument we can color code our scatterplots for the cases where we select k to be eight and sixty as can be seen in Figures 1.5 and 1.6. In Figure 1.5 we see that for k=8, the k-means method clearly does not differentiate the clusters as one would expect, that is, one might expect to find eight clearly separate color regions for the ζ vs α plot, but we see that they overlap too much, this should not be surprising since we are just looking at the projections of a heptadimensional space in a bidimensional one. We have also plotted the cluster centers as black dots of greater diameter and we see that the cluster centers overlap in three separate regions. For Figure 1.6 the case is even more confusing since one would have to distinguish 60 different colors. This situation doesn't get better when instead of colors we use numbers from one thru sixty to show which points belong to which cluster. The previous results are a good argument to use data reduction approaches, we propose to introduce biased classifications on the data, whether the bias has its origin in taking sequence into account, or from chemical considerations like taking A-type conformations into account, for example, more interesting results can be obtained as has been shown by others [13].

[v] Checking the data in two dimensions more clearly it seem to me that at most one could say that there can be six groups

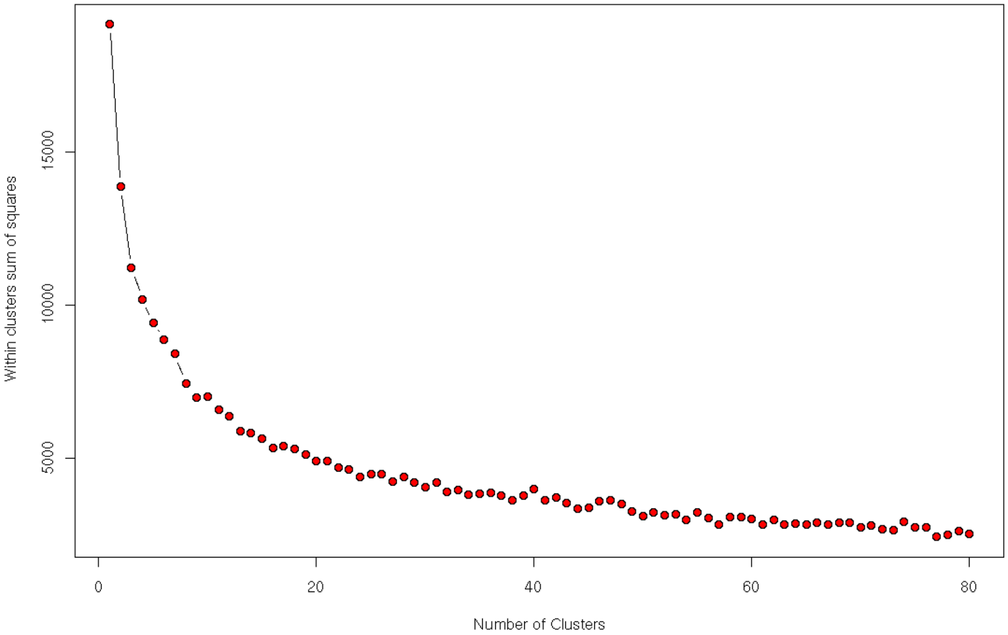

Figure 1.3: Sum of all within clusters sum of squares against number of clusters for data of all torsion angles in 23S rRNA.

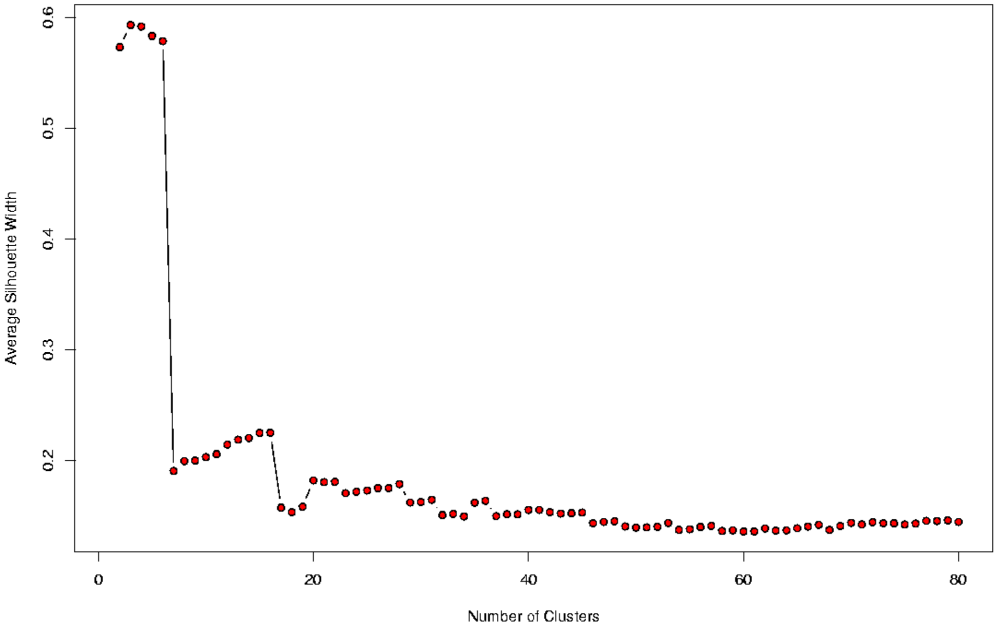

Figure 1.4: Average silhouette width against number of clusters for data of all torsion angles in 23S rRNA. The best clustering method and value of k is then defined as the model that maximizes a.s.w.

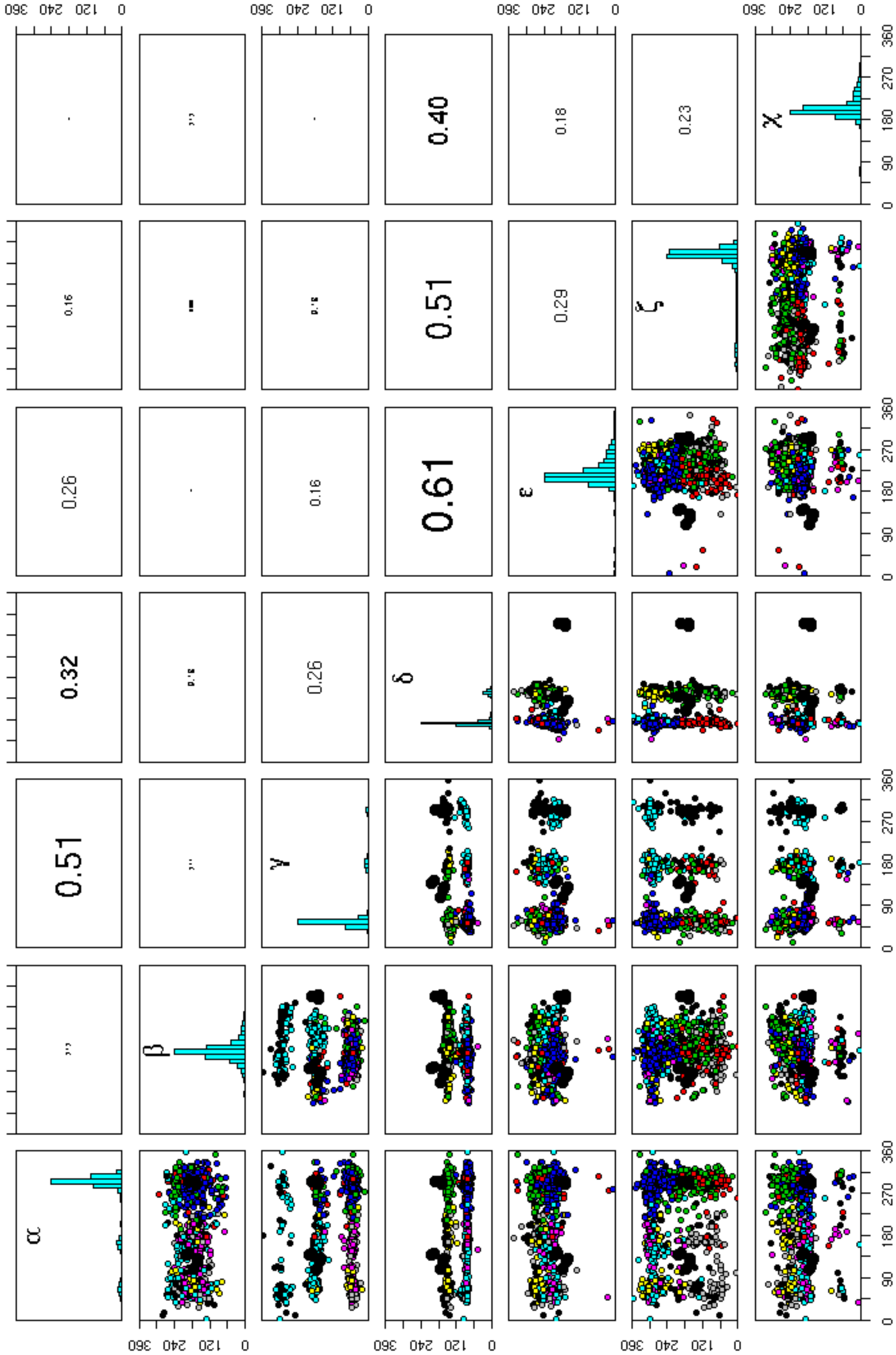

Figure 1.5: K-means clustering of heptadimensional torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA. The number of partitions is **8**. The large black dots represent cluster centers. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

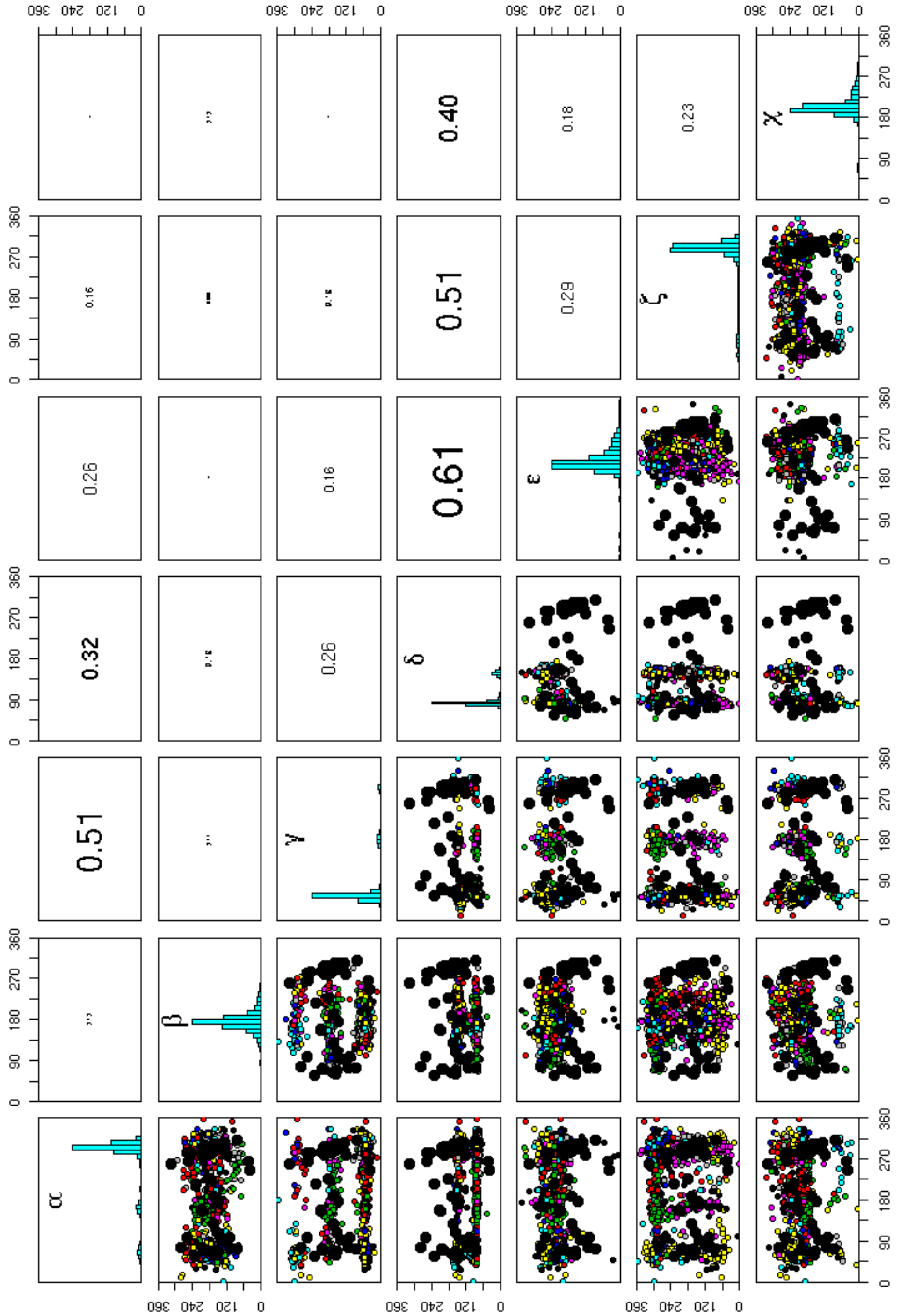

Figure 1.6: K-means clustering of heptadimensional torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA. The number of partitions is **60**. The large black dots represent cluster centers. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

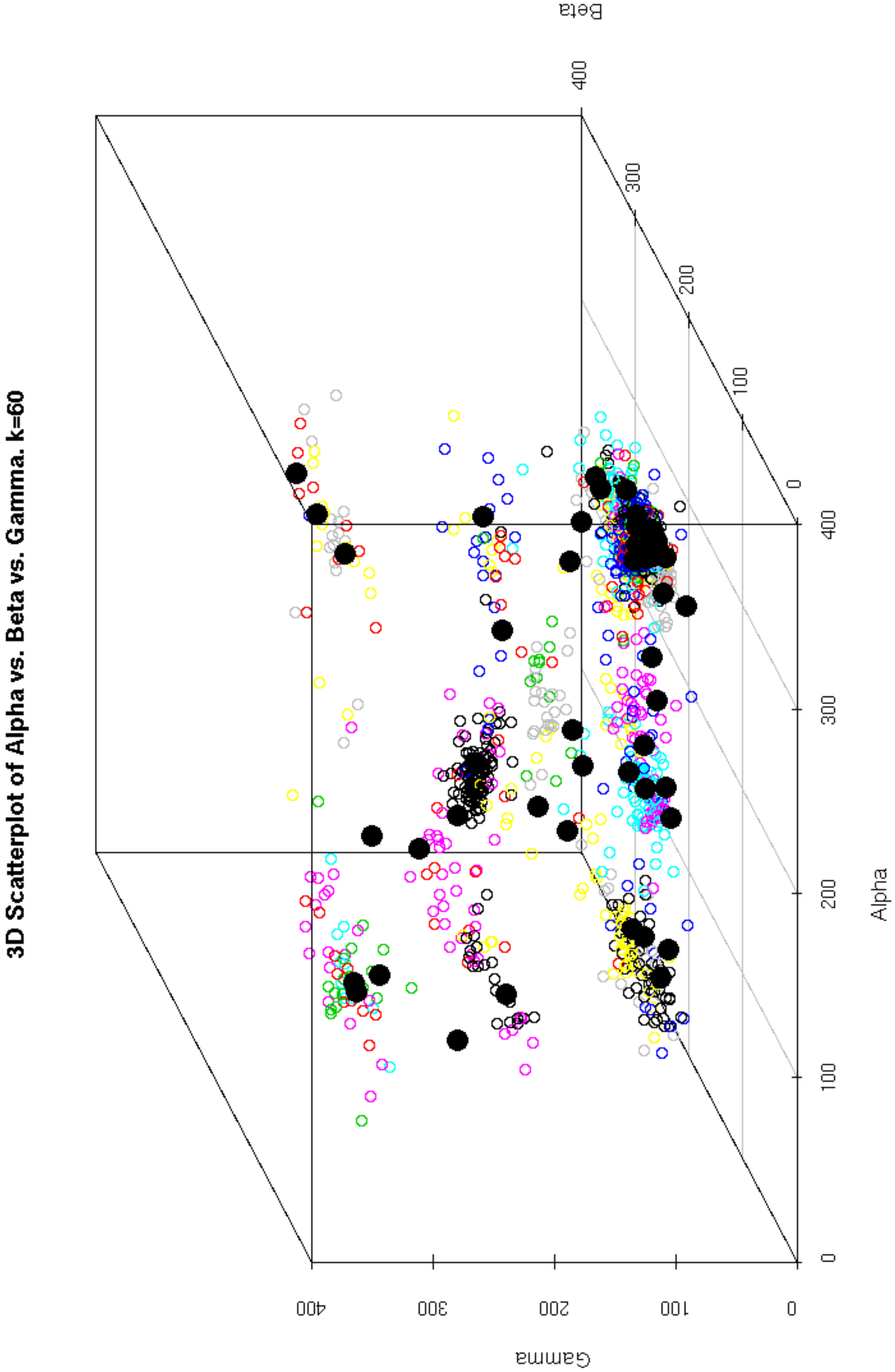

Figure 1.7: K-means clustering of the heptadimensional torsion angle vectors of 2753 dinucleotide steps of 23S rRNA. The axis of the three dimensional scatterplot corresponds to the torsion angles, α, β, and γ. The large black dots correspond with the cluster centers for clustering by using k-means with k=60.

1.1.2 Hierarchical Clustering for Torsion Angles

Other authors have used hierarchical clustering to analyze the torsion angles in nucleic acid structure, taking the Fröbenius norm and Ward's method as the distance definition for four different RNA representations (see Reijmers et al. [8]) on a "small" database similar to that of Duarte and Pyle [14]. This databases do not include ribosomal RNA's. The other case where hierarchical clustering has been used did not use torsion angles, but rather a set of 15 atoms belonging to the nucleotides sugar and backbone. The latter study used used the unweighted pair group method (UPGMA) to classify a database of RNA loop structures (see Huang et al. [23]).

In our case we used three distance definitions (Euclidean, Manhattan, and maximum), and four different clustering methodologies, that is, single, complete, average and centroid (see A. We tried to make a consensus analysis of the twelve trees obtained, but these trees are too large. The number of trees, that is, twelve, is also too small compared with the data vectors (2753 step vectors), for the algorithms to find a reasonable consensus. We were not even able to find consensus for two "near" clusters, where the "near" criteria was taken from a tree dissimilarity algorithm implemented in **R** cluster and whose result can be seen in Figure 1.9 where the previously mentioned "near" clusters refer to clusters five and nine. One typical suggestion in clustering analysis is to just use the single linkage method since this one uses as its grouping criteria the minimal distance between clusters, therefore giving a direct interpretation of the trees as just a direct minimal proximity relation depending on the metric being used (see Appendix A). [vi] In Figure 1.8 we see twelve clustering trees clustering all base-steps in the large subunit of the ribosome. It's noticeable that trees are more similar when the linkage method is the same than when the metric is the same. The main problem with the dendrograms in Figure 1.8 is very similar to that of determining the number of partitions in partitional clustering. In this case we have to determine which is the optimal tree height which would determine how many meaningful groups we have, for Figure 1.8 we have selected to draw boxes around a group of branches at the height where 36 branches are found. The reason for selecting 36 is because this was one of the maxima in Figure 1.4 and because itÂt's close to 37, the number of discrete nucleotide conformations suggested by Hershkovitz et al. [13]

1.2 Base-step Parameters

To our knowledge there has been no classification of rigid-body base-step parameters for RNA structures deposited at the PDB [vii]. It is important to note here that in crystal structures, RNA bases are determined more accurately than backbone torsion angles, as has been shown by Richardson and collaborators from analysis of van der Waals steric clashes. This can be seen more clearly in Figure 1.10, reproduced from Richardson's work [10], where the red and orange dots in the backbone atoms region denote steric clashes and the green and yellow dots in the base atoms region denote very good agreement with expected van der Waals distances.

1.2.1 Combining Fourier Averaging Results and Clustering Analysis

Using the coordinates files of 20 rRNA structures provided by Schneider at al.[12] we have used standard clustering analysis (CA) techniques to classify a set of non-ARNA base-steps using, rather than the torsion angles space, the base-step parameters space, that is, three translational parameters (ShiftD_x, SlideD_y, RiseD_z), and three rotational parameters (Tiltτ, Rollρ, Twistω), which are described by the hexaparametric vector ν:

[vi] The author still has to do consensus of single method trees.

[vii] The effort of putting together a database for such effort could be an interesting project to be considered.

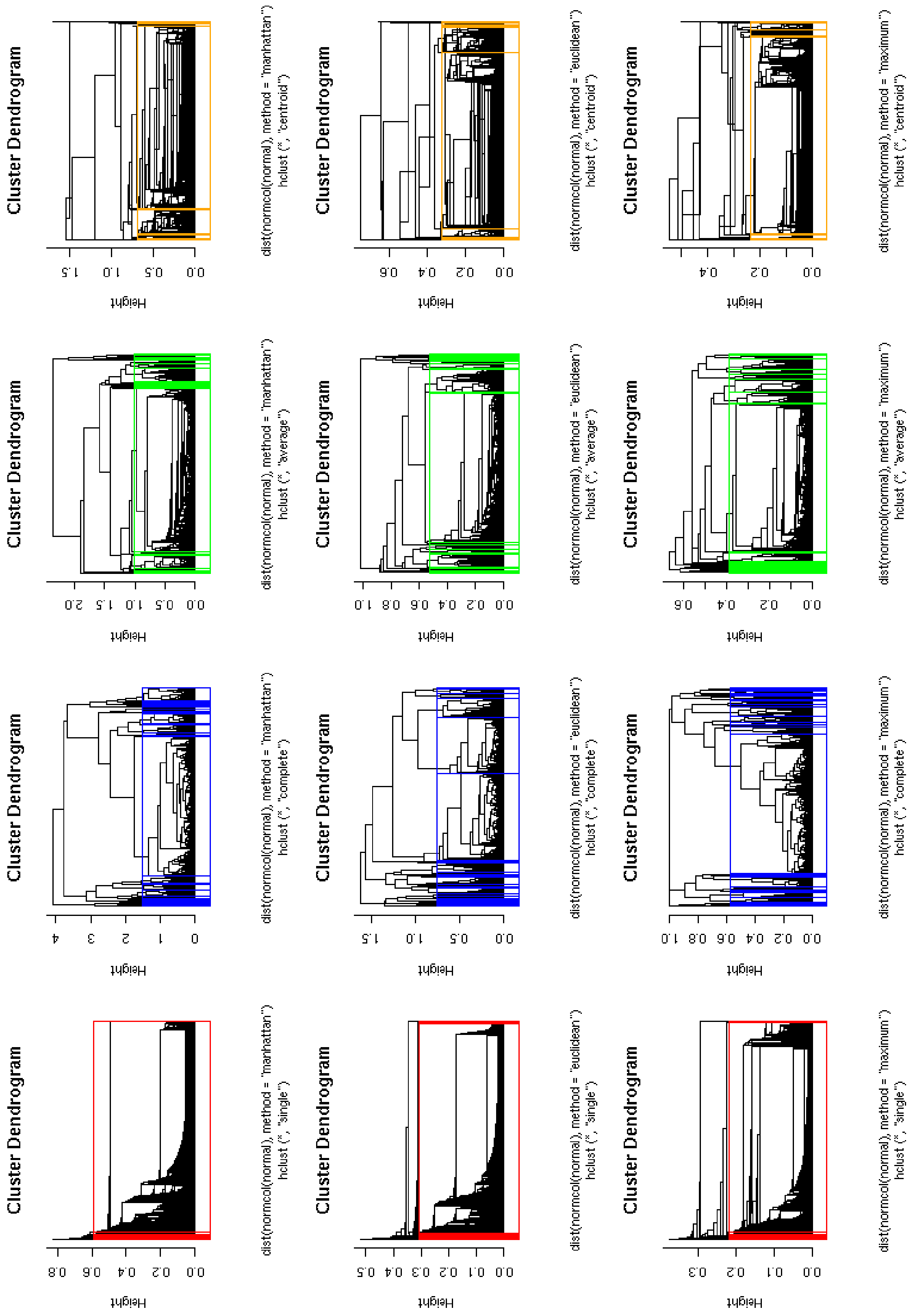

Figure 1.8: Hierarchical clustering for the twelve trees obtained from clustering of torsion angles of the large subunit of the ribosome (PDB-ID:1jj2). We have colored a box around branches for the case where the height of each tree has 36 branches.

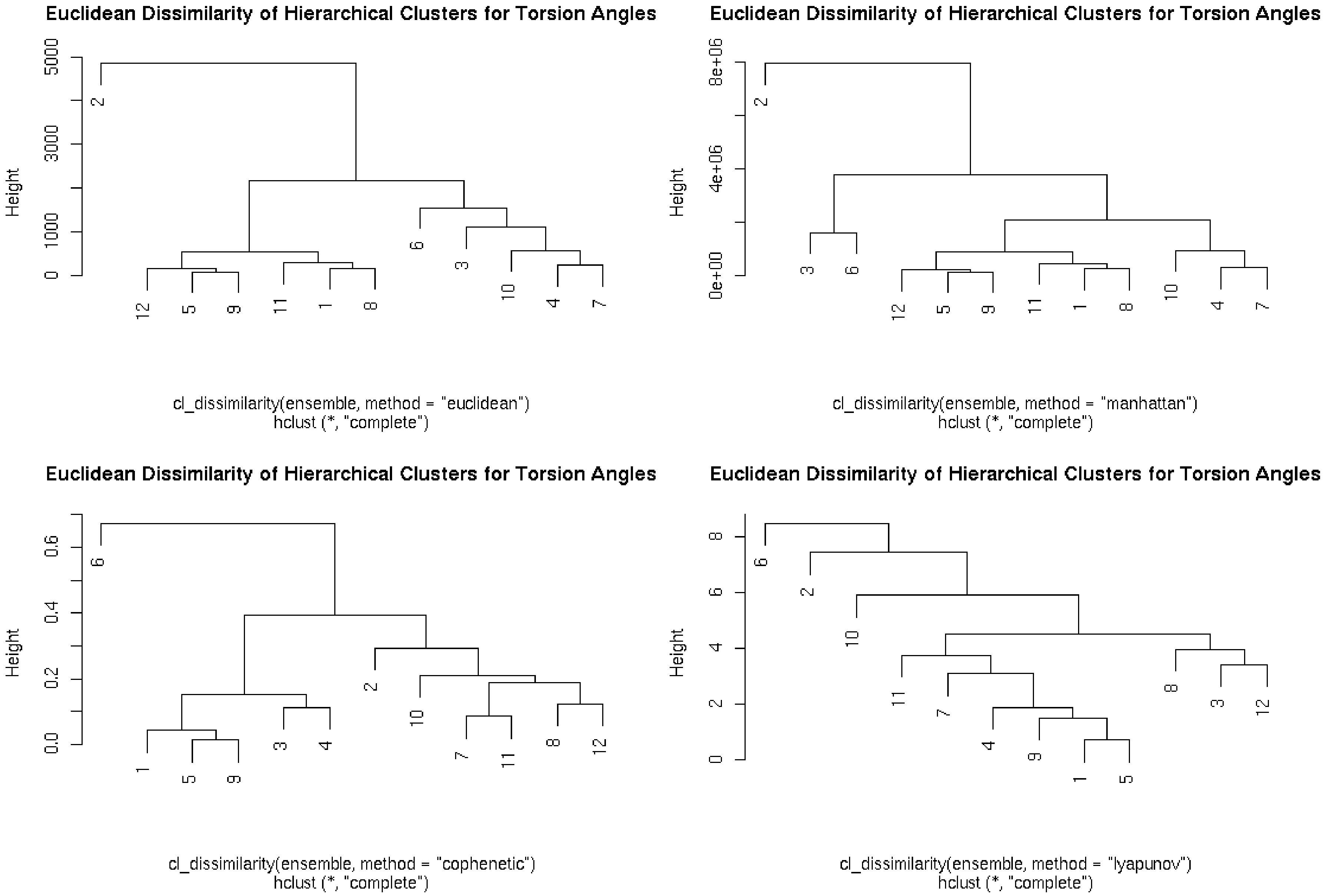

Figure 1.9: Cluster dissimilarities for the 12 combinations of metrics and methods used to obtain hierarchical clusterings of the 2753 heptadimensional torsion angle vectors of 23S rRNA.

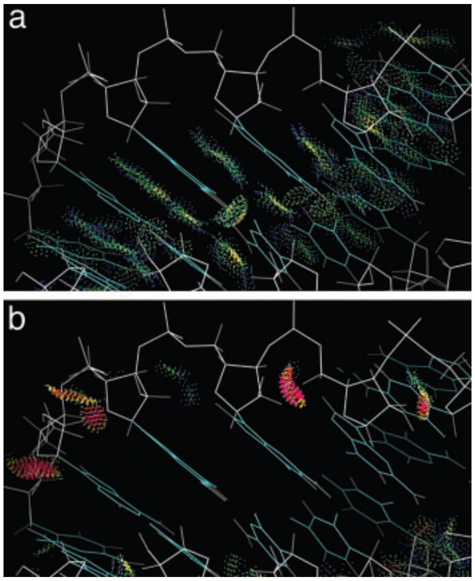

Figure 1.10: Figure taken from Richardson et al. [10] where the blue and green dots in a) mean very accurate van der Waals distances, and in b) the red and orange dots mean steric clashes, that is, distances outside the acceptable van der Waals range.

$$\nu = (D_x, D_y, D_z, \tau, \rho, \omega) \tag{1.1}$$

The results illustrated in Figures 1.11 and 1.12 were obtained by performing clustering analysis and consensus clustering on 20 structures provided by Schneider et al. [12]. These twenty structures were obtained by Schneider applying a Fourier averaging technique and lexicographical clustering to torsion angles of 23S rRNA. The methodology we used follows that used by others to recover the periodic table classification from multidimensional property vectors for elements [24, 25]. Table 1.2 shows the residue numbers of bases from 23S rRNA which belong to the main categories of Figure 1.12. To decide which residues of 23S rRNA belonged to the non-Atype clusters, a root mean squared deviation (RMSD) of 15 or less was required between step parameter vectors of 23S rRNA and the mean parameter vectors for the four non-Atype groups identified.

1.2.2 Partitional Clustering for Rigid Body Parameters

The same type of analysis that has been carried out for torsion angles can also be carried out for rigid body parameters, that is, partitional clustering, and hierarchical clustering, are used as standard statistical analysis methods to analyze our set of 2753 base-step parameter vectors. For the partitional clustering case, again, there is no known number of clusters in which the data must group, therefore we've calculated the within clusters sum of squares and also the average silhouette widths, for a particular selection of the number of partitions of the data for $k = [2-80]$. From figure 1.13 we can't conclude much. We see that the value of the within clusters sum of squares becomes constant around $k = 47$ and thereÂt's also a change of curvature around $k = 13$. For the case where the average silhoutte width has been computed, that is, figure 1.14, we see that the maximum is for $k = 2$, and there are some interesting maxima at $k = 9, 12$. Now that we have a clue as to which number of partitions the data optimally has we have plotted the k-means results for $k = 13$ and $k = 47$ in Figures bla and bla, and the PAM results for $k = 2, 9, 12$ in Figure bla.

We have also prefiltered the data according to the 16 possible RNA base steps, that is, AA, AG, GA,

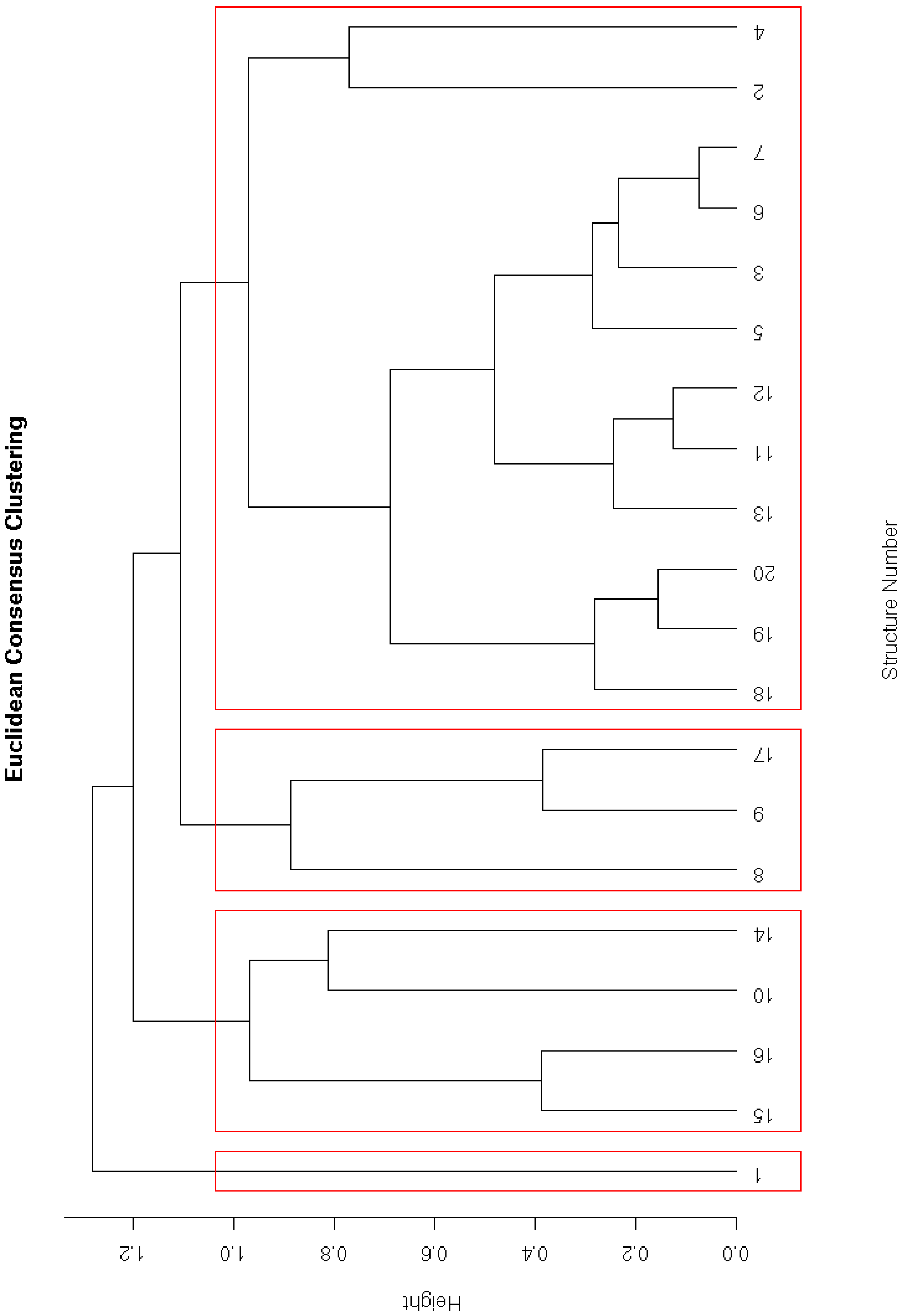

Figure 1.11: Dendrogram showing the results of consensus clustering of 20 non-Atype rRNA dinucleotides according to their hexadimensional base-step parameter vectors.

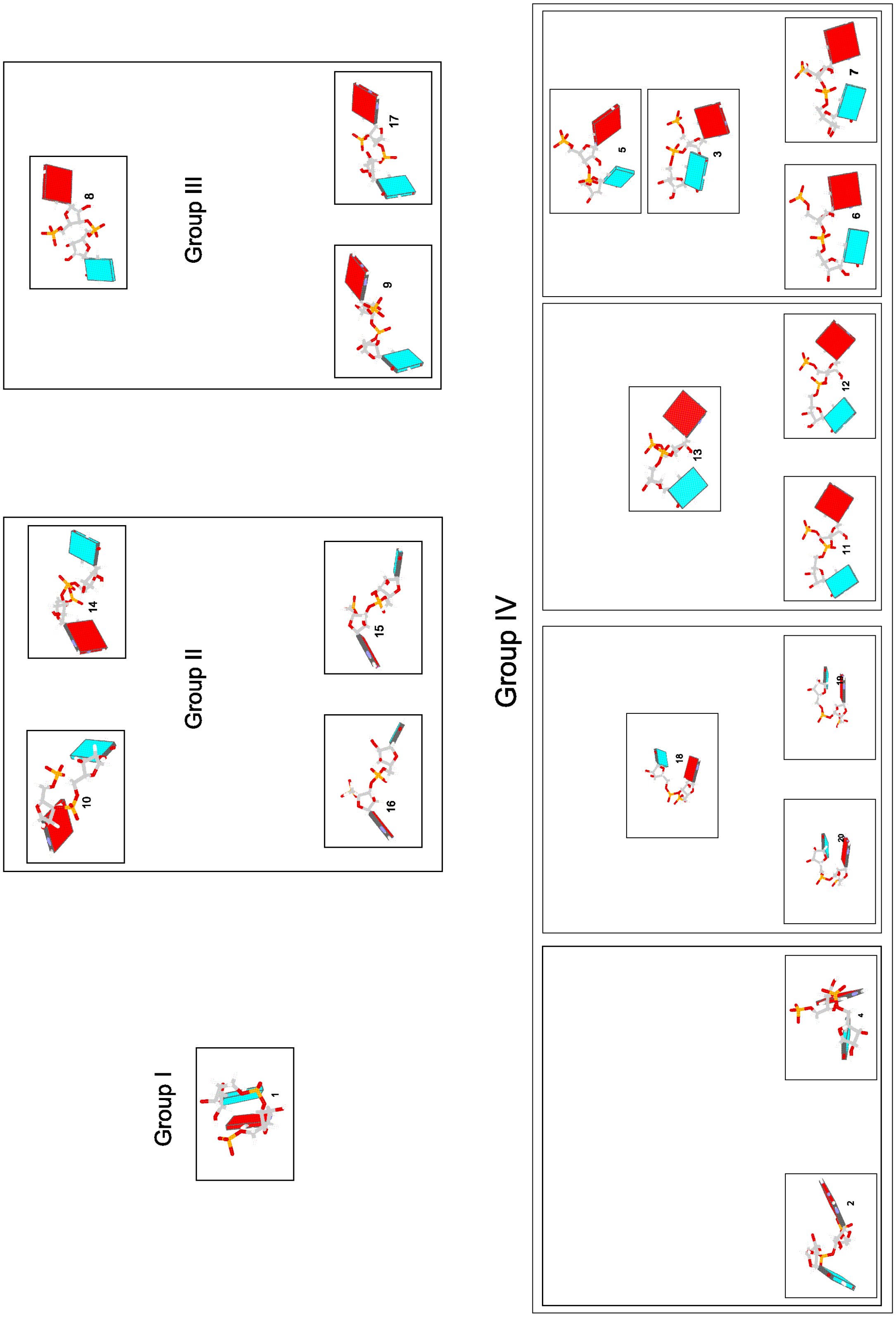

Figure 1.12: rRNA dinucleotide structures organized by clusters obtained from consensus clustering of their hexadimensional base-step parameter vectors.

Total Number of Nucleotides	**RMSD Limit**	**Group**	**Base-steps**	**Base-step Residue Number**	**Overlaps**
2754	< 15	I	3	892, 2006, 2390	
		II	5	459, 1279, 1653, 1919, 2302	
		III	1	2109	
		IV	35	79, 112, 128, 190, 213, 269, 358, 434, 488, 564, 706, 720, 775, 867, 966, 1292, 1503, 1543, 1614, 1766, 1874, 1908, 1971, 2017, 2257, 2427, 2516, 2540, 2755, 2782, 2810, 2826, 2874, 2882, 2913	
		IVa	1	882	
		IVb	807		
		IVc	9	306, 789, 854, 880, 1107, 1192, 1493, 1818, 2005	
		IVd	35	175, 213, 246, 264, 304, 358, 464, 518, 531, 534, 588, 795, 938, 1214, 1231, 1316, 1340, 1370, 1605, 1745, 1766, 1971, 1976, 2010, 2017, 2291, 2320, 2428, 2469, 2481, 2516, 2532, 2755, 2826, 2882	Only IVd with IV (213, 358, 1766, 1971, 2017, 2516, 2755, 2826, 2882)

Table 1.2: Residue numbers for base-steps with RMSD values less than 15 between the reference base-step vectors from the four groups of non-A-type RNA dinucleotide conformations and all base-step vectors found in the 23S strand of *Haloarcula marismortui* large ribosomal subunit.

GG, UU, UC, CU, CC, UA, UG, CA, CG, AU, AC, GU, and GC. Tables showing how many representatives steps there are belonging to non-helical, helical, and watson-crick sets, will be later included and discussed here.

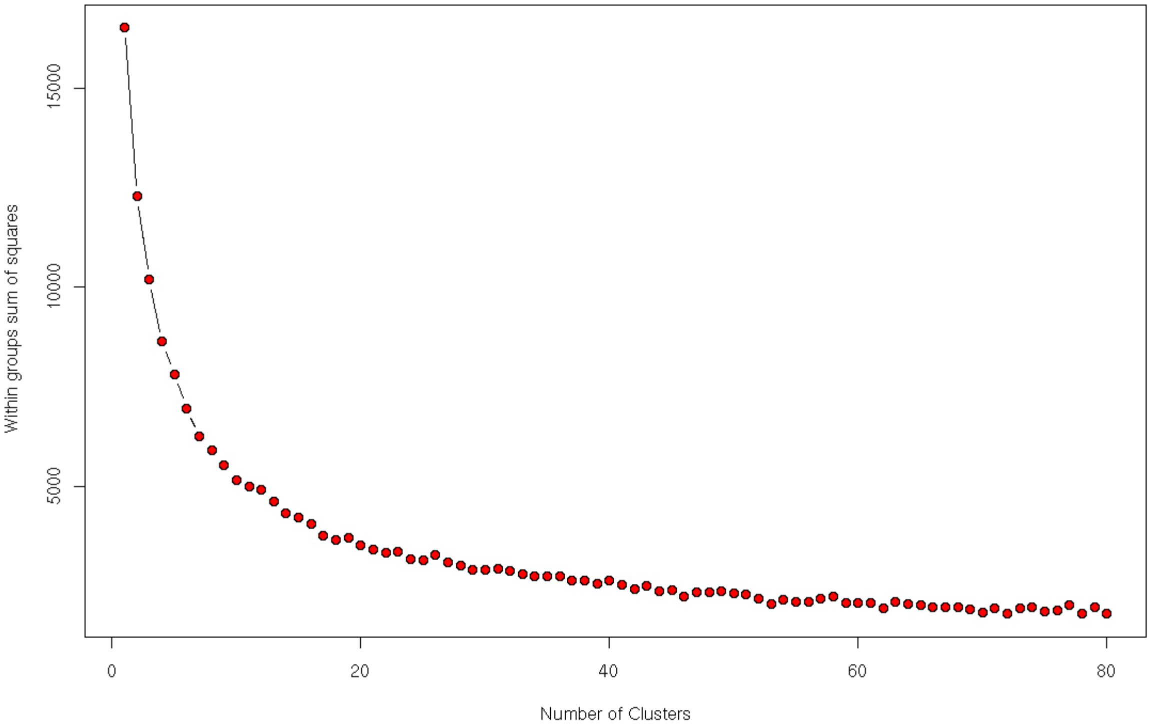

Figure 1.13: Sum of all within clusters sum of squares against number of clusters.

1.2.3 Hierarchical Clustering for Rigid Body Parameters

Also as has been carried out for torsion angles, hierarchical clustering has also been performed on rigid body parameters, the results are yet to be included here. A cluster dissimilarity tree can be seen in Figure 1.15 for the 12 trees resulting from the four clustering methods and three distance definitions used to cluster the base step data.

1.3 RNA Conformations

There are two main RNA conformations, A-RNA ,and A'RNA, and maybe even a third unconfirmed one A"RNA [2]. Their values for their standard torsion angles and step parameters can be seen in Tables 1.3 and 1.4

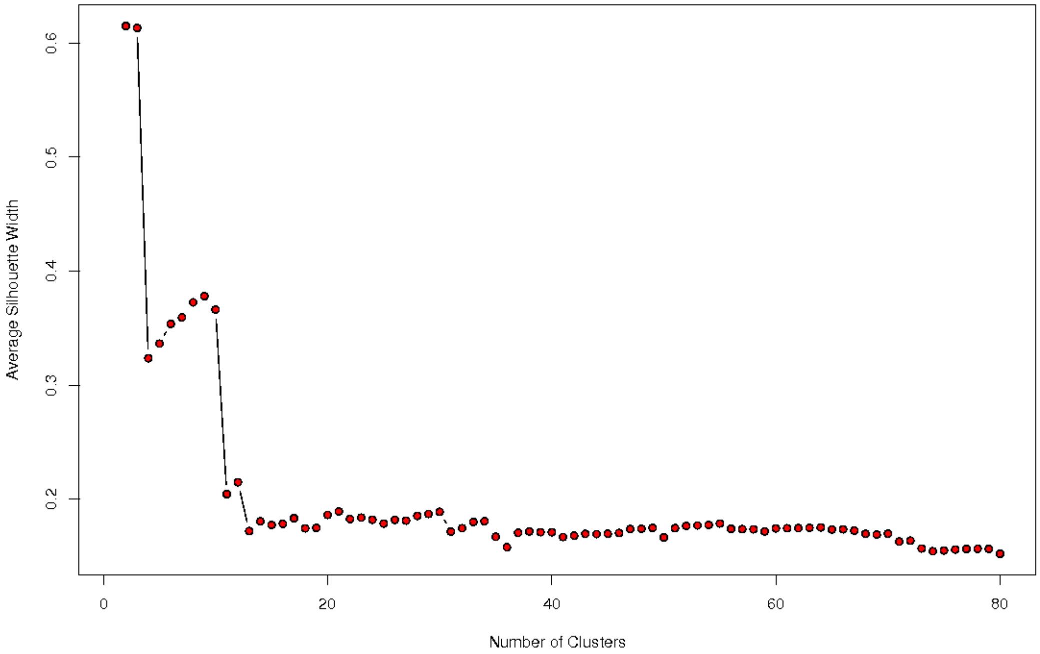

Figure 1.14: Average silhouette width against number of clusters.

Structure Name	α	β	γ	δ	ϵ	ζ	χ	**Reference**
A-RNA	-68.9	179.5	54.5	82.2	-153.9	-70.8	-161.1	Arnott
A'-RNA	-70.0	176.6	60.8	76.7	-153.4	-69.4	-163.4	Arnott
AII-RNA	-65.0	175.1	52.9	81.1	-166.0	-68.0	-157.0	Schneider

Table 1.3: Base step torsion angles for the different known RNA conformations.

Structure Name	Shift (D_x)	Slide (D_y)	Rise (D_z)	Tilt (τ)	Roll (ρ)	Twist (Ω)	**Reference**
A-DNA	0.36	-1.39	3.29	2.46	12.50	30.19	
B-DNA	0.44	0.47	3.33	4.63	1.77	35.67	
A-RNA	-0.08	-1.48	3.30	-0.43	8.64	31.57	Arnott
A'-RNA	0.05	-1.88	3.39	-0.12	5.43	29.52	Arnott
AII-RNA	1.01	-2.52	3.33	2.94	9.75	25.12	Schneider

Table 1.4: Base step parameters for the different known RNA conformations. Notice that the base step parameters are for single bases rather than base-pairs.

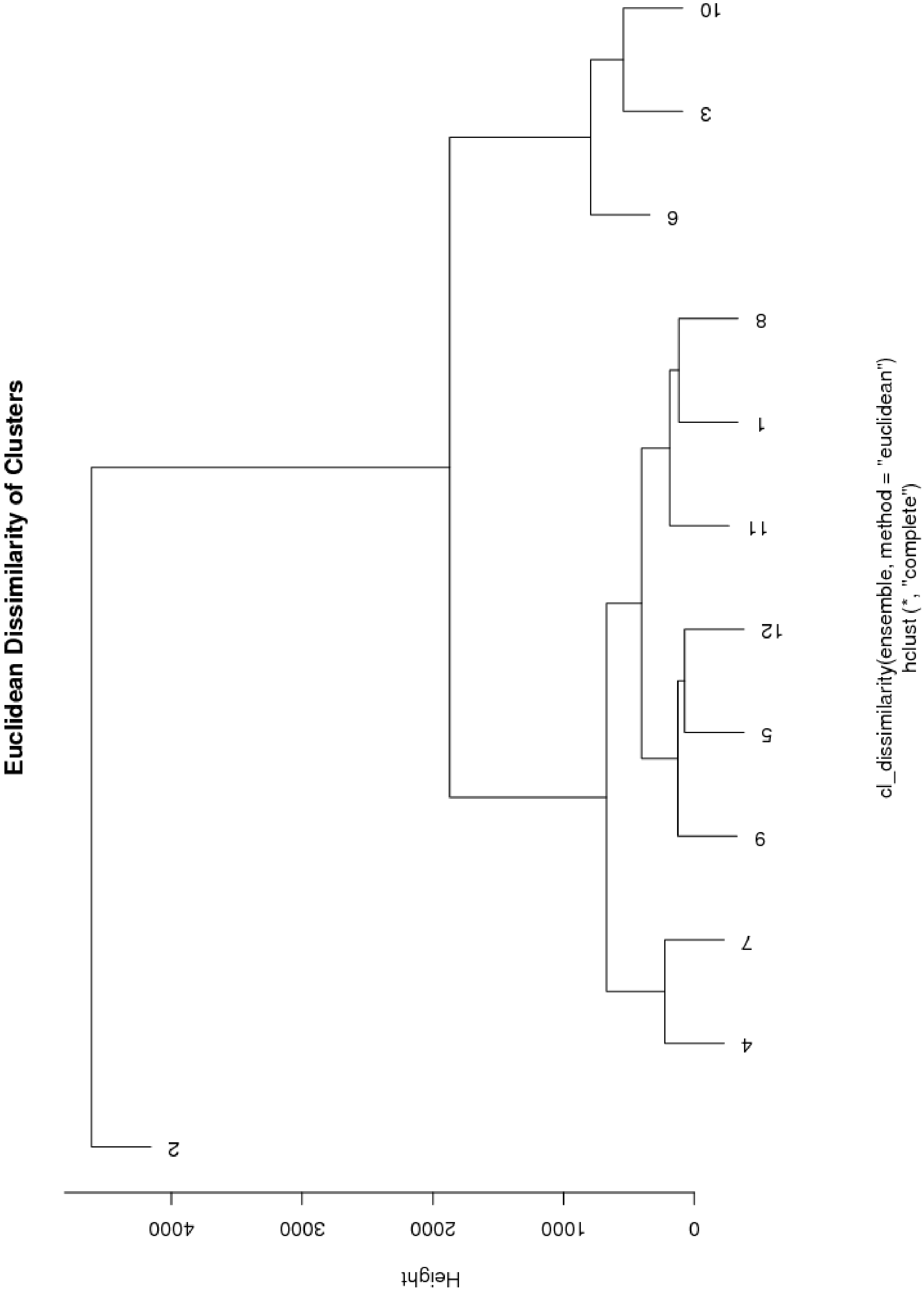

Figure 1.15: Cluster dissimilarities for the twelve hierarchical trees obtained from clustering of the six-dimensional base-step parameters obtained from the large subunit of the ribosome (PDB-ID:1jj2)

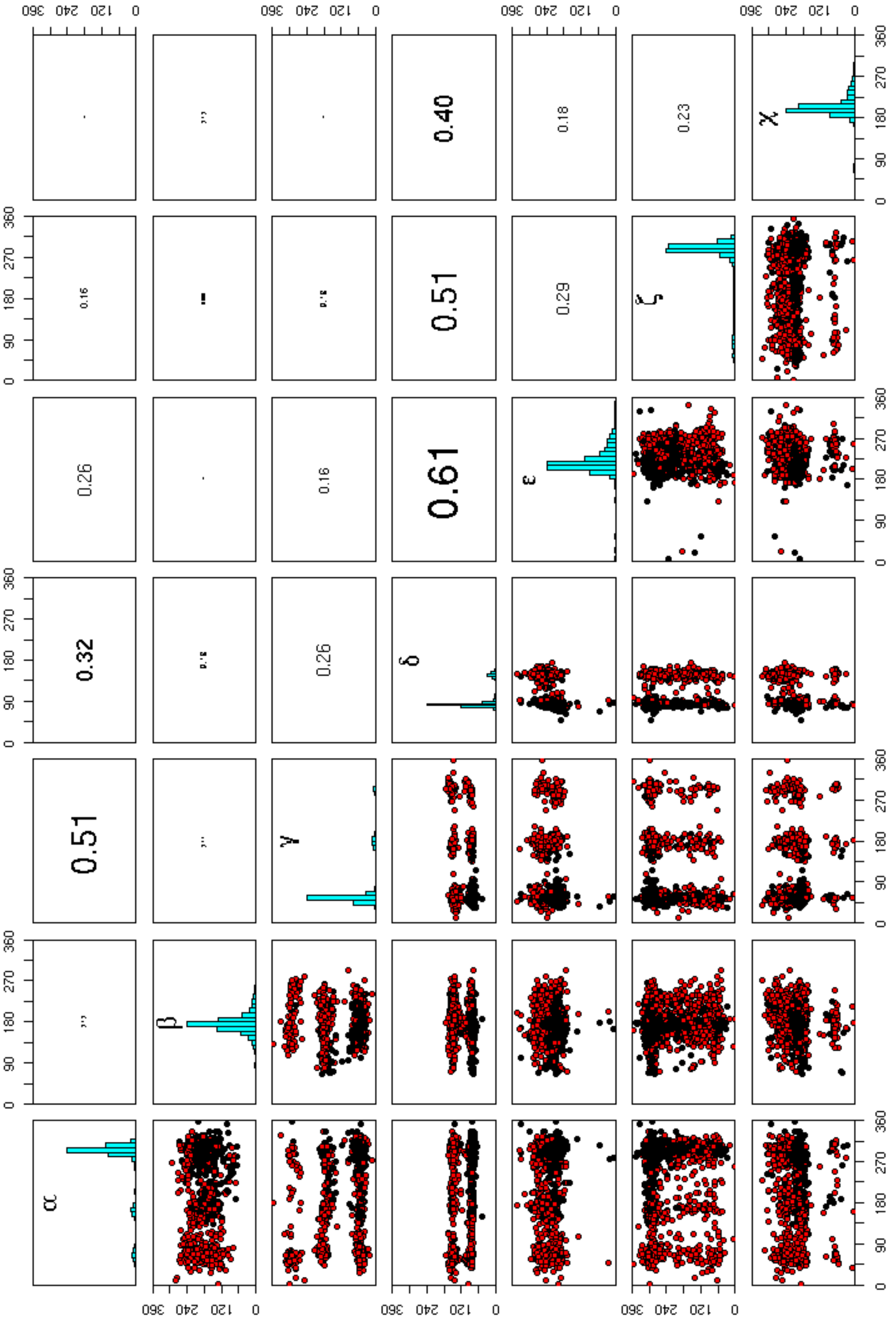

Figure 1.16: K-means of torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA using the *Hartigan-Wong* algorithm. The number of partitions is **2**. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

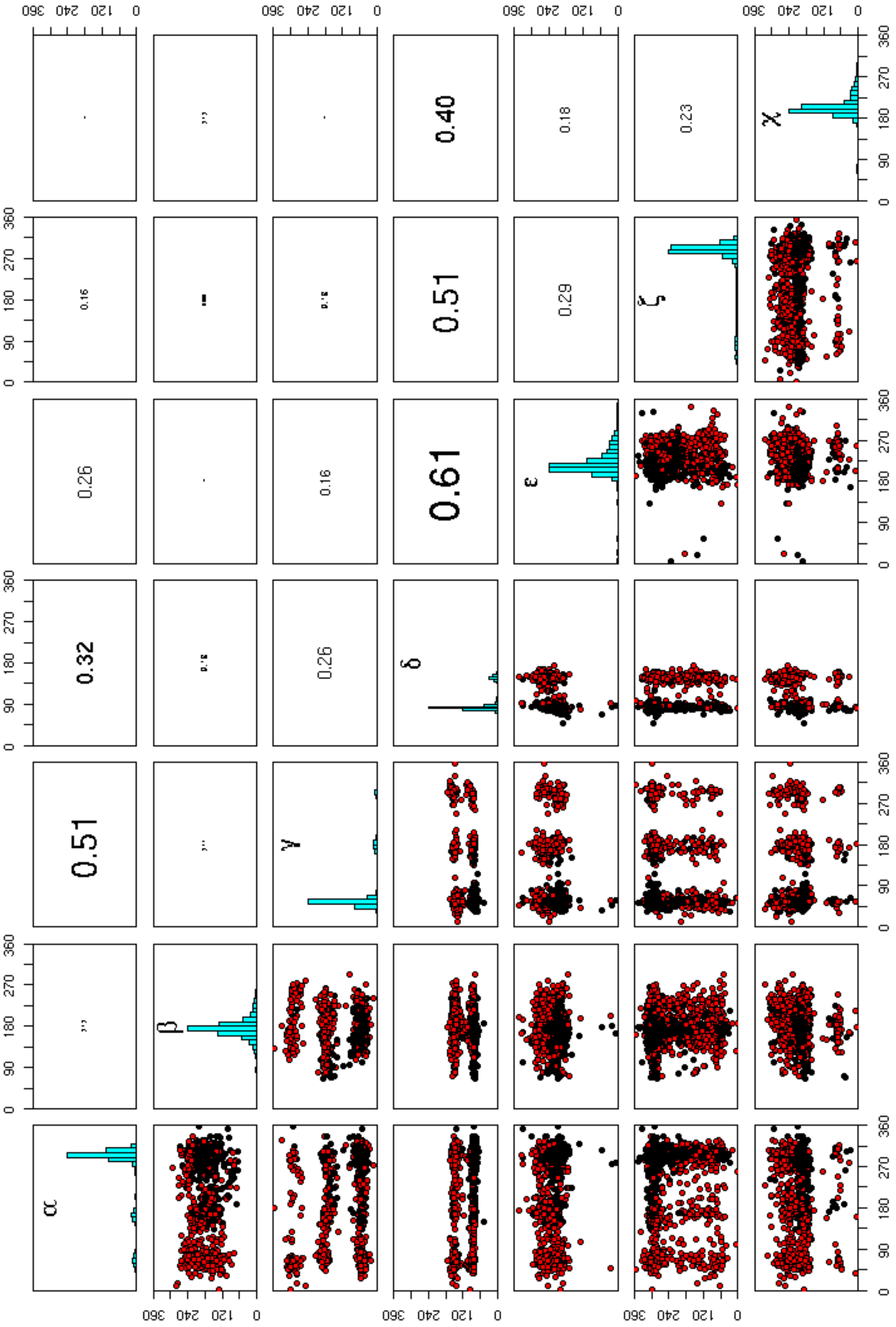

Figure 1.17: K-means of torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA using the *Lloyd* algorithm. The number of partitions is **2**. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

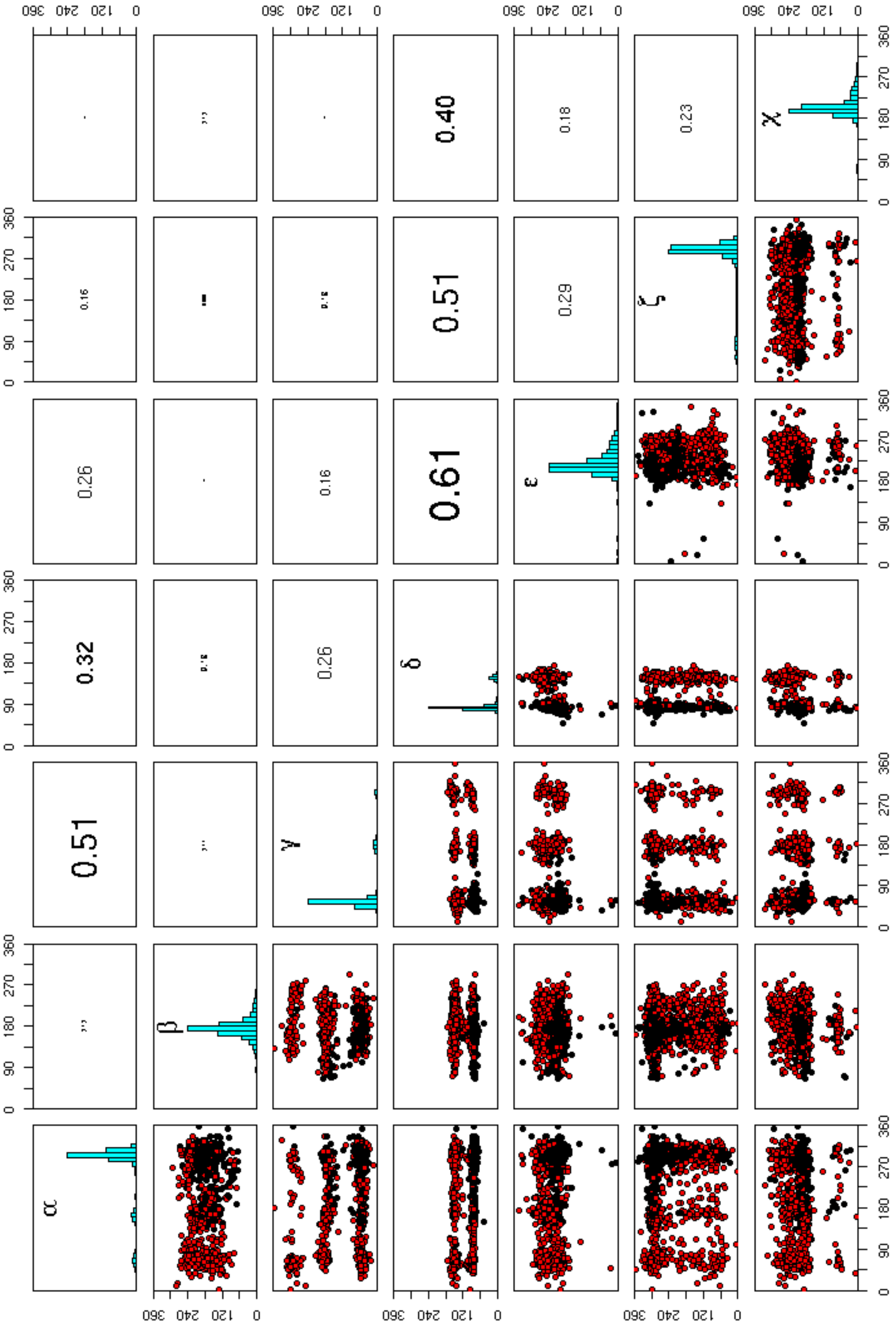

Figure 1.18: K-means of torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA using the *Forgy* algorithm. The number of partitions is **2**. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

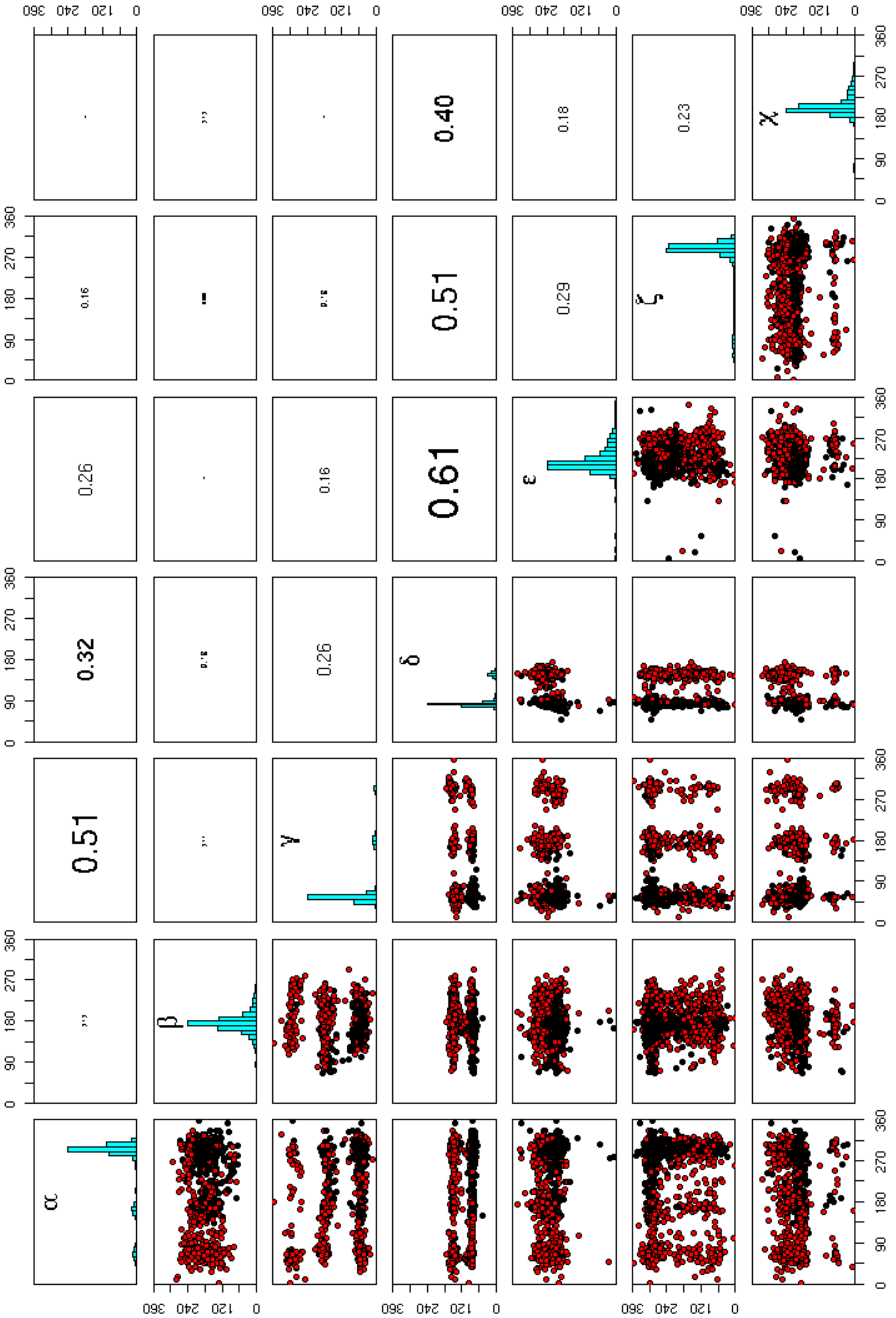

Figure 1.19: K-means of torsion angle vectors of 2753 dinucleotide steps present in 23S rRNA using the *McQueen* algorithm. The number of partitions is **2**. The upper diagonal matrix displays the values of the linear correlation coefficient r, and a histogram showing the torsion angle distribution is rendered in the diagonal.

References

[1] Olson, W. K.; Flory, P. J. *Biopolymers* **1972**, *11*, 1.

[2] Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag; London, 1984.

[3] Gautheret, D.; Major, F.; Cedergren, R. *Journal of Molecular Biology* **1993**, *229*, 1049.

[4] Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartschk, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327.

[5] Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; Yonath, A. *Cell* **2000**, *102*, 615.

[6] Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.

[7] Noller, H. F. *Science* **2005**, *309*, 1508.

[8] Reijmers, T. H.; Wehrens, R.; Buydens, L. M. C. *Journal of Chemical Information and Computer Science* **2001**, *41*, 1388.

[9] Sykes, M. T.; Levitt, M. *Journal of Molecular Biology* **2005**, *351*, 26.

[10] Murray, L. J. W.; III, W. B. A.; Richardson, D. C.; Richardson, J. S. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 13904.

[11] Hershkovitz, E.; Tannenbaum, E.; Howerton, S. B.; Sheth, A.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Research* **2003**, *31*, 6249.

[12] Schneider, B.; Moravek, Z.; Berman, H. *Nucleic Acids Research* **2004**, *32*, 1666.

[13] Hershkovitz, E.; Sapiro, G.; Tannenbaum, A.; Williams, L. D. *Transactions on Computational Biology and Bioinformatics* **2006**, *3*, 33.

[14] Duarte, C. M.; Pyle, A. M. *Journal of Molecular Biology* **1998**, *284*, 1465.

[15] Duarte, C. M.; Wadley, L. M.; Pyle, A. M. *Nucleic Acids Research* **2003**, *31*, 4755.

[16] Wadley, L. M.; Keating, K. S.; Duarte, C. M.; Pyle, A. M. *Journal of Molecular Biology* **2007**, *372*, 942.

[17] Westhof, E.; Fritsch, V. *Structure* **2000**, *8*, R55.

[18] Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Research* **2002**, *30*, 3497.

[19] Leontis, N. B.; Lescoute, A.; Westhof, E. *Current Opinion in Structural Biology* **2006**, *16*, 279.

[20] Jain, A. K.; Murthy, M. N.; Flynn, P. J. *ACM Computing Surveys* **1999**, *31*, 265.

[21] R Development Core Team, , R: A Language and Environment for Statistical Computing, Vienna, Austria (2007), ISBN 3-900051-07-0.

[22] Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; ; Berman, H. M. *RNA* **2008**,

[23] Huang, H.-C.; Nagaswamy, U.; Fox, G. E. *RNA* **2005**, *11*, 412.

[24] Restrepo, G.; Mesa, H.; Llanos, E. J.; Villaveces, J. L. *Journal of Chemical Information and Computer Science* **2004**, *44*, 68.

[25] Restrepo, G.; Llanos, E. J.; Meza, H. *Journal of Mathematical Chemistry* **2006**, *39*, 401.

Chapter 2
Sequence Analysis of Ribosomal Step Parameters

For the 2007 IMA Meeting, “RNA in Biology, Bioengineering and Nanotechnology”, that took place from October 29 to November 2, 2007 in Minneapolis, Minnesota, we sorted the base-steps of rRNA according to 16 possible sequence base-steps, that is, Purine-Purine (RR: AA, AG, GA, GG), Pyrimidine-Pyrimidine (YY: UU, UC, CU, CC), Pyrimidine-Purine (YR: UA, UG, CA, CG), and Purine-Pyrimidine (RY: AU, AC, GU, GC). In Table 2.1 we show how these 16 base-steps distribute in the complete rRNA structure (All), in the helical regions of rRNA found by the 3DNA software (Helical), and in base-steps whose first member is part of a Watson-Crick (WC) base-pair.

Ribosome 50S		All	Helical	Watson-Crick
RR	AA	202	102	12
	AG	232	164	60
	GA	250	177	64
	GG	249	238	160
YY	UU	75	54	20
	UC	147	107	66
	CU	142	123	78
	CC	200	186	137
YR	UA	124	90	24
	UG	176	143	62
	CA	163	96	60
	CG	226	192	126
RY	AU	116	79	37
	AC	189	127	60
	GU	188	141	90
	GC	196	181	94

Table 2.1: Distribution of base-step types in rRNA (PDB-ID:1jj2)

In Tables 2.2 and 2.3 we show the average values and the standard deviation for the base-step parameters for the 16 dinucleotide steps.

Once we have classified the base-step parameters for dinucleotide steps according to sequence we can also plot them automatically in a pairs plot. Such automatic plots, which have not been scaled according to a specific scale, are quite useful to identify uncommon base-steps. **Whether such uncommon base-step parameter values are due to bad cristallographic results, or to intrinsic structural peculiarities of such base-steps seems to be a question worthy of being addressed.** In Figures 2.33 to 2.48

	Shift			Slide			Rise		
	All	Helical	W-C	All	Helical	W-C	All	Helical	W-C
AA(mean)	-0.409	0.826	-0.133	-0.559	-0.418	-1.999	2.604	2.226	2.919
AA(sd)	6.780	3.535	0.611	2.708	2.389	0.980	5.101	2.487	0.934
AG(mean)	0.639	2.245	0.614	-1.157	-1.161	-1.719	2.982	3.110	3.151
AG(sd)	6.494	3.160	1.199	2.457	1.747	1.152	3.913	1.458	0.429
GA(mean)	-3.286	-2.142	0.578	-1.776	-1.549	-1.851	2.867	2.645	3.086
GA(sd)	6.193	4.100	0.855	2.413	2.078	0.768	3.575	2.784	0.672
GG(mean)	0.316	0.545	0.763	-1.705	-1.839	-2.064	2.936	3.072	3.196
GG(sd)	3.595	1.633	1.171	1.527	1.120	0.665	2.278	0.935	0.425
UU(mean)	-0.989	-0.095	0.761	-1.073	-1.087	-1.623	2.762	2.480	3.174
UU(sd)	6.087	2.306	1.022	1.914	1.921	0.677	4.192	2.314	0.268
UC(mean)	-0.010	0.421	0.138	-1.460	-1.298	-1.519	3.379	3.047	3.322
UC(sd)	6.121	2.071	1.124	1.693	0.944	0.867	3.678	1.332	0.345
CU(mean)	0.289	0.581	0.596	-1.435	-1.432	-1.539	3.191	3.278	3.385
CU(sd)	4.303	1.708	0.857	1.077	1.080	0.696	2.546	0.942	0.584
CC(mean)	0.472	0.244	0.165	-1.604	-1.546	-1.619	3.635	3.404	3.437
CC(sd)	3.352	1.414	0.938	1.066	0.781	0.500	1.857	1.139	0.491
UA(mean)	-2.232	-0.773	0.402	-1.282	-1.445	-1.466	3.048	2.330	2.938
UA(sd)	7.180	3.037	0.474	2.278	1.736	0.769	4.314	2.851	0.885
UG(mean)	-0.712	0.413	0.396	-1.278	-1.095	-1.660	2.545	2.800	3.220
UG(sd)	7.044	2.445	1.115	2.750	1.547	0.473	3.828	1.879	0.458
CA(mean)	-0.566	0.271	0.175	-1.159	-1.173	-1.254	3.109	2.877	3.119
CA(sd)	6.379	1.807	1.440	2.629	1.633	1.211	4.472	1.471	1.050
CG(mean)	0.932	0.873	0.807	-1.207	-1.333	-1.638	3.459	2.878	3.331
CG(sd)	5.131	1.910	1.589	1.910	1.035	0.638	3.181	1.398	0.419
AU(mean)	-1.586	1.181	0.848	-0.050	-0.805	-1.288	1.935	2.518	3.088
AU(sd)	10.773	2.062	1.267	11.914	1.628	0.399	5.399	1.934	0.238
AC(mean)	0.247	1.535	0.823	-1.200	-1.161	-1.406	2.856	2.950	3.191
AC(sd)	5.753	2.532	0.802	1.761	1.074	0.600	3.724	1.631	0.556
GU(mean)	-0.899	0.580	0.838	-1.630	-1.496	-1.473	2.887	2.937	3.004
GU(sd)	5.842	2.115	1.232	1.843	1.101	0.596	3.196	0.898	0.471
GC(mean)	-0.450	0.663	0.584	-1.520	-1.351	-1.539	2.633	2.730	3.047
GC(sd)	5.351	1.430	1.240	1.422	1.266	1.209	2.720	1.317	0.764

Table 2.2: Average base-step parameters Shift(D_x), Slide(D_y), and Rise(D_z), and their standard deviations for the 16 possible specific dinucleotide base-steps in rRNA (PDB-ID:1jj2)

	Tilt			Roll			Twist		
	All	Helical	W-C	All	Helical	W-C	All	Helical	W-C
AA(mean)	1.752	-3.965	-8.317	3.130	2.794	1.794	20.757	28.123	39.992
AA(sd)	52.388	46.911	43.597	43.872	48.766	12.441	61.833	69.499	44.488
AG(mean)	1.660	-0.290	4.970	-2.915	6.432	6.909	29.437	44.960	30.366
AG(sd)	40.762	33.407	12.249	54.999	44.517	17.874	60.077	49.794	14.675
GA(mean)	1.524	7.486	2.774	5.657	3.350	1.891	0.303	23.609	32.426
GA(sd)	40.488	48.392	8.416	47.487	38.171	18.947	62.473	52.173	10.706
GG(mean)	4.767	-0.406	2.018	2.980	0.242	4.702	27.322	35.409	34.969
GG(sd)	26.889	29.679	13.127	25.556	30.270	10.133	35.549	25.956	14.067
UU(mean)	4.590	7.830	4.119	14.585	2.299	7.251	17.974	26.520	32.371
UU(sd)	45.485	42.449	6.192	38.878	37.830	6.406	59.460	50.909	6.669
UC(mean)	-3.746	3.885	2.690	5.659	10.155	7.935	24.156	34.421	35.726
UC(sd)	45.258	19.279	6.744	38.300	27.183	6.568	61.100	33.273	17.096
CU(mean)	1.586	-0.455	0.619	15.989	11.728	7.085	31.730	35.311	31.520
CU(sd)	31.071	15.965	9.664	28.484	34.158	20.490	39.764	32.381	8.130
CC(mean)	-2.865	-0.014	-0.817	13.487	12.353	12.499	33.396	35.132	31.231
CC(sd)	20.623	19.864	9.442	21.393	21.634	23.199	29.108	22.634	15.752
UA(mean)	-5.082	17.366	5.997	15.921	13.854	5.854	14.453	23.904	31.072
UA(sd)	48.747	60.210	35.883	44.432	50.073	33.762	69.347	70.086	21.478
UG(mean)	2.816	11.150	2.876	11.702	11.719	11.405	16.030	35.716	31.355
UG(sd)	44.717	34.923	4.588	35.585	35.347	7.323	69.936	41.080	7.587
CA(mean)	-5.424	1.918	4.473	18.541	14.010	12.131	27.438	33.751	25.556
CA(sd)	46.535	32.314	33.935	49.656	51.725	36.917	64.201	44.243	37.512
CG(mean)	-3.526	6.437	-1.149	8.030	7.857	10.392	31.739	31.399	33.078
CG(sd)	35.351	54.584	16.207	29.225	15.785	6.968	48.690	43.902	18.347
AU(mean)	4.660	-3.373	6.439	12.288	11.167	7.512	18.943	22.843	34.141
AU(sd)	50.513	56.862	5.414	35.199	56.726	5.386	71.378	53.225	9.904
AC(mean)	4.070	5.696	6.374	5.919	4.142	1.707	27.191	42.259	33.983
AC(sd)	40.408	26.459	5.253	34.583	31.050	16.098	56.979	39.818	6.101
GU(mean)	5.035	3.387	7.095	7.111	6.286	7.078	21.209	40.686	38.729
GU(sd)	35.221	26.999	6.432	33.337	29.488	18.139	50.758	30.847	15.570
GC(mean)	5.477	6.392	7.383	9.002	-0.112	-0.117	22.118	33.511	31.661
GC(sd)	30.625	43.730	9.615	34.766	28.612	24.074	52.357	36.615	24.642

Table 2.3: Average base-step parameters Tilt(τ), Roll(ρ), and Twist(ω), and their standard deviations for the 16 possible specific dinucleotide base-steps in rRNA (PDB-ID:1jj2)

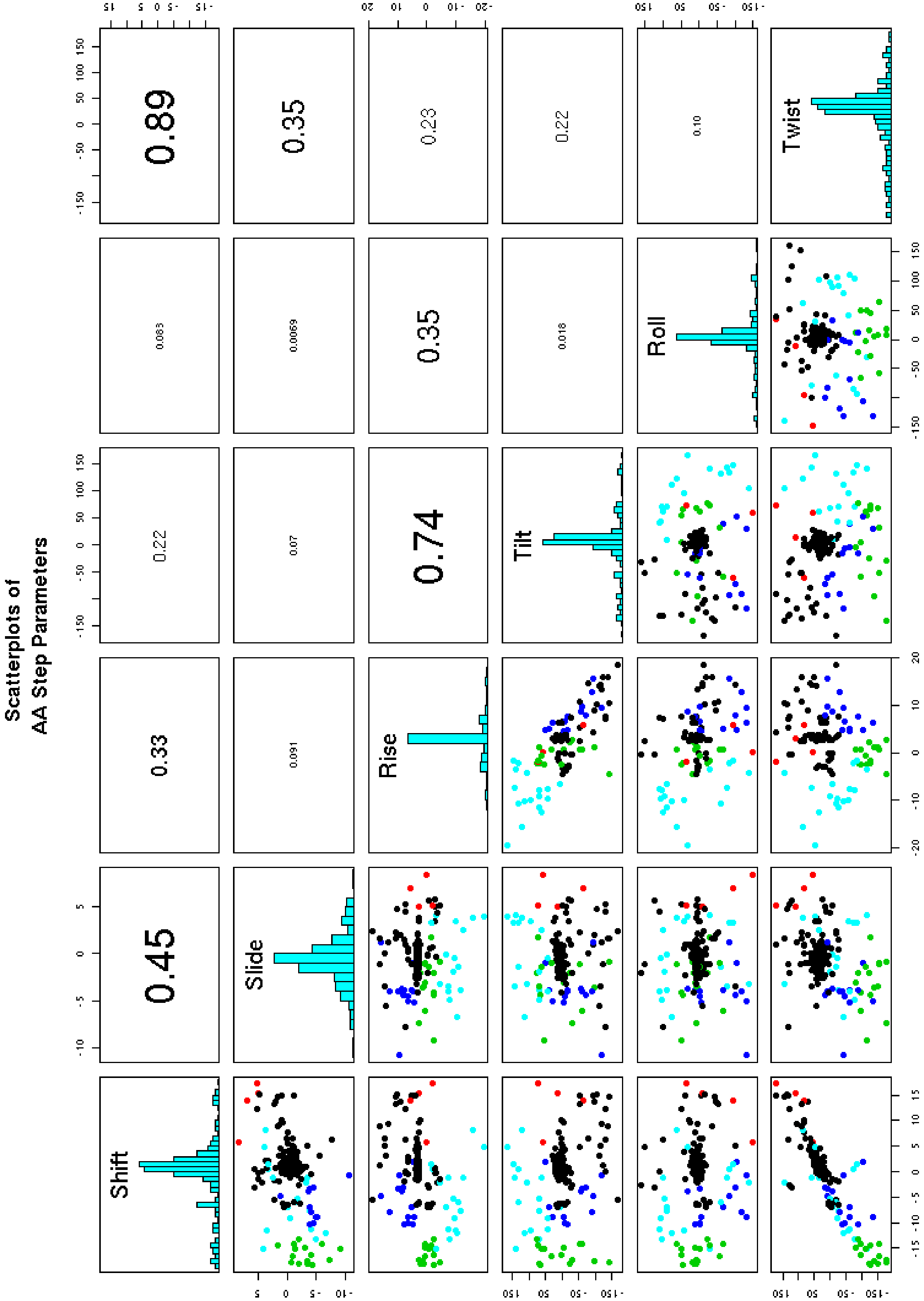

Figure 2.1: Scatterplots for step-parameters of **All** AA dinucleotide steps in 50S rRNA.

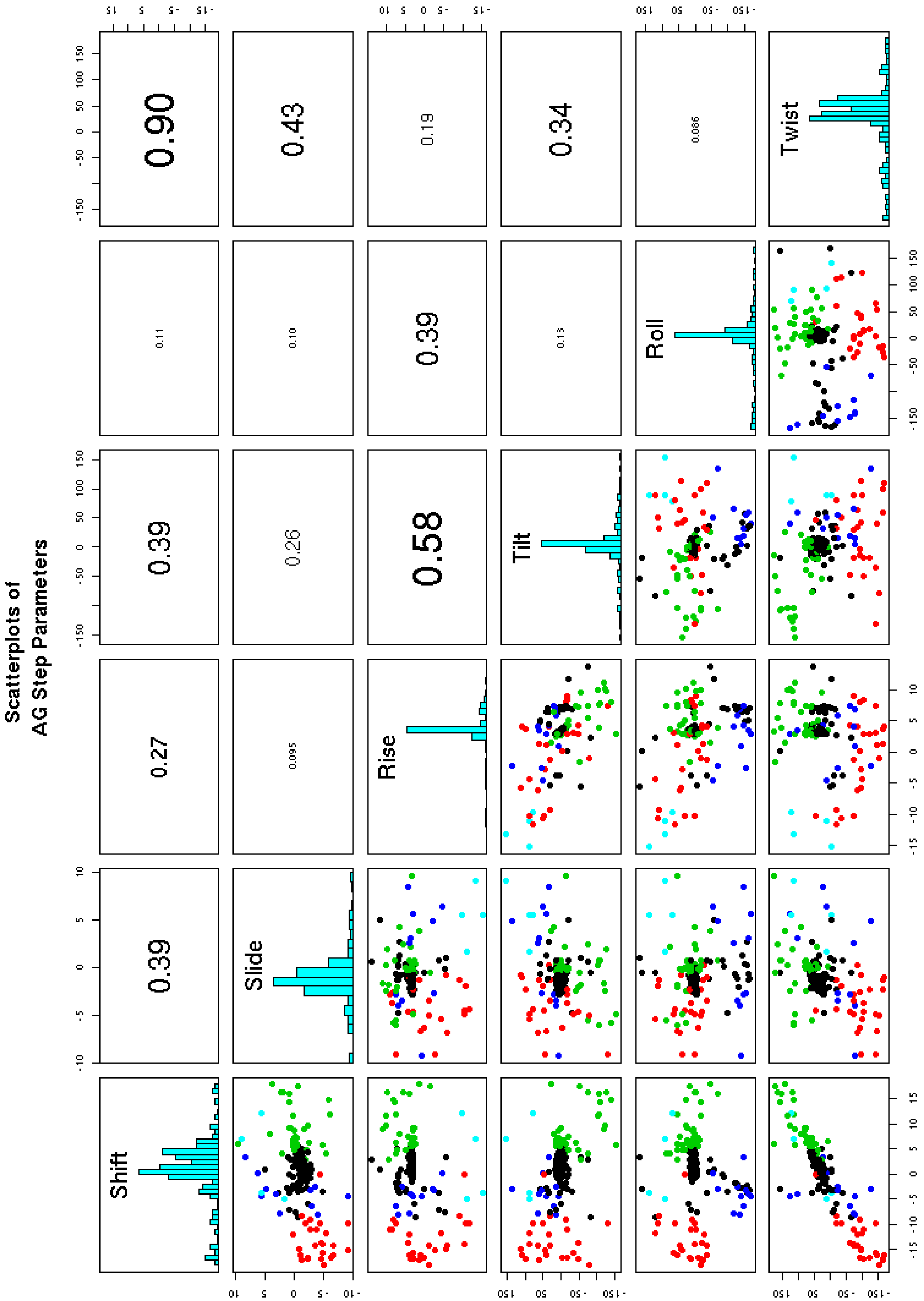

Figure 2.2: Scatterplots for step-parameters of **All** AG dinucleotide steps in 50S rRNA.

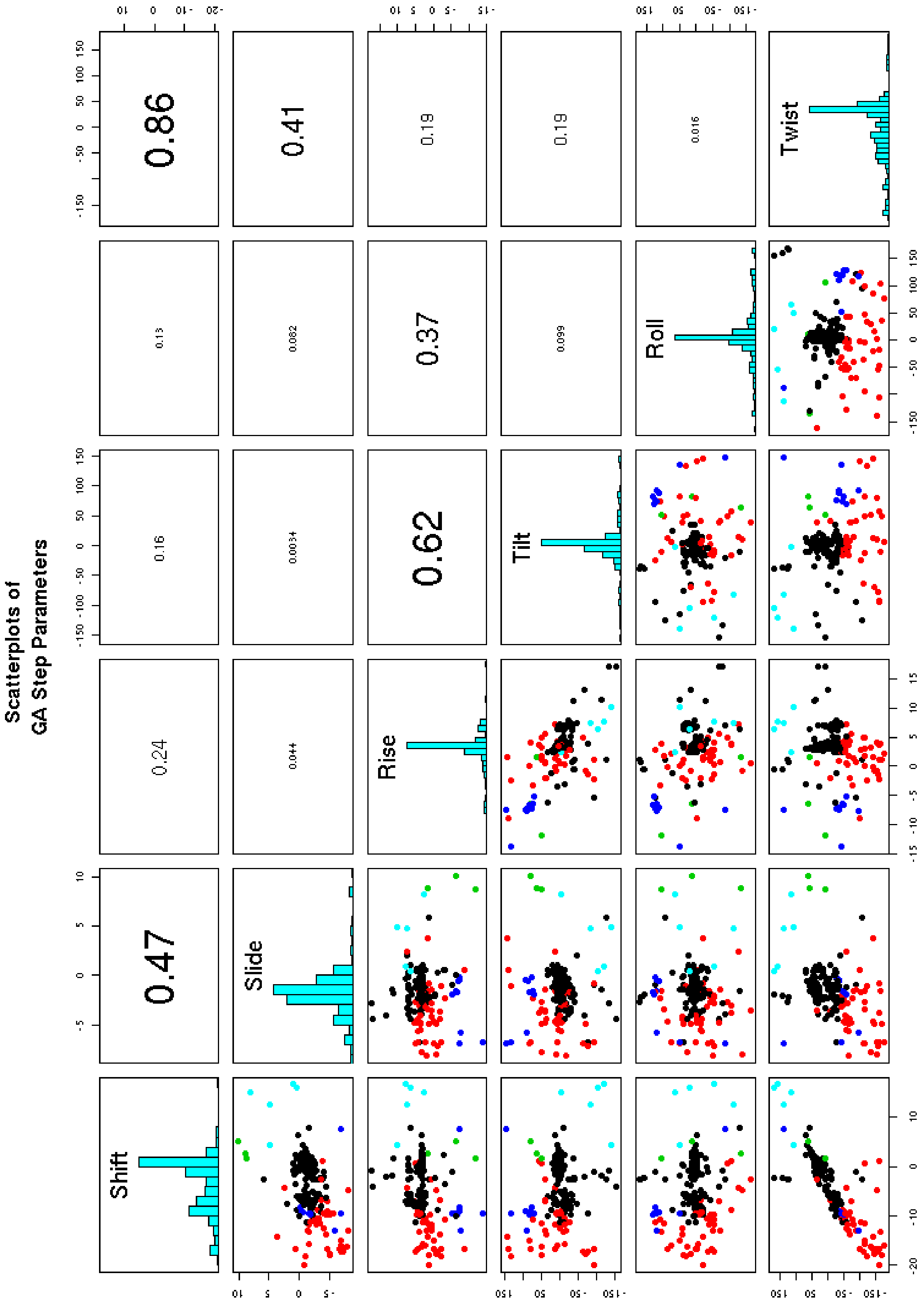

Figure 2.3: Scatterplots for step-parameters of **All** GA dinucleotide steps in 50S rRNA.

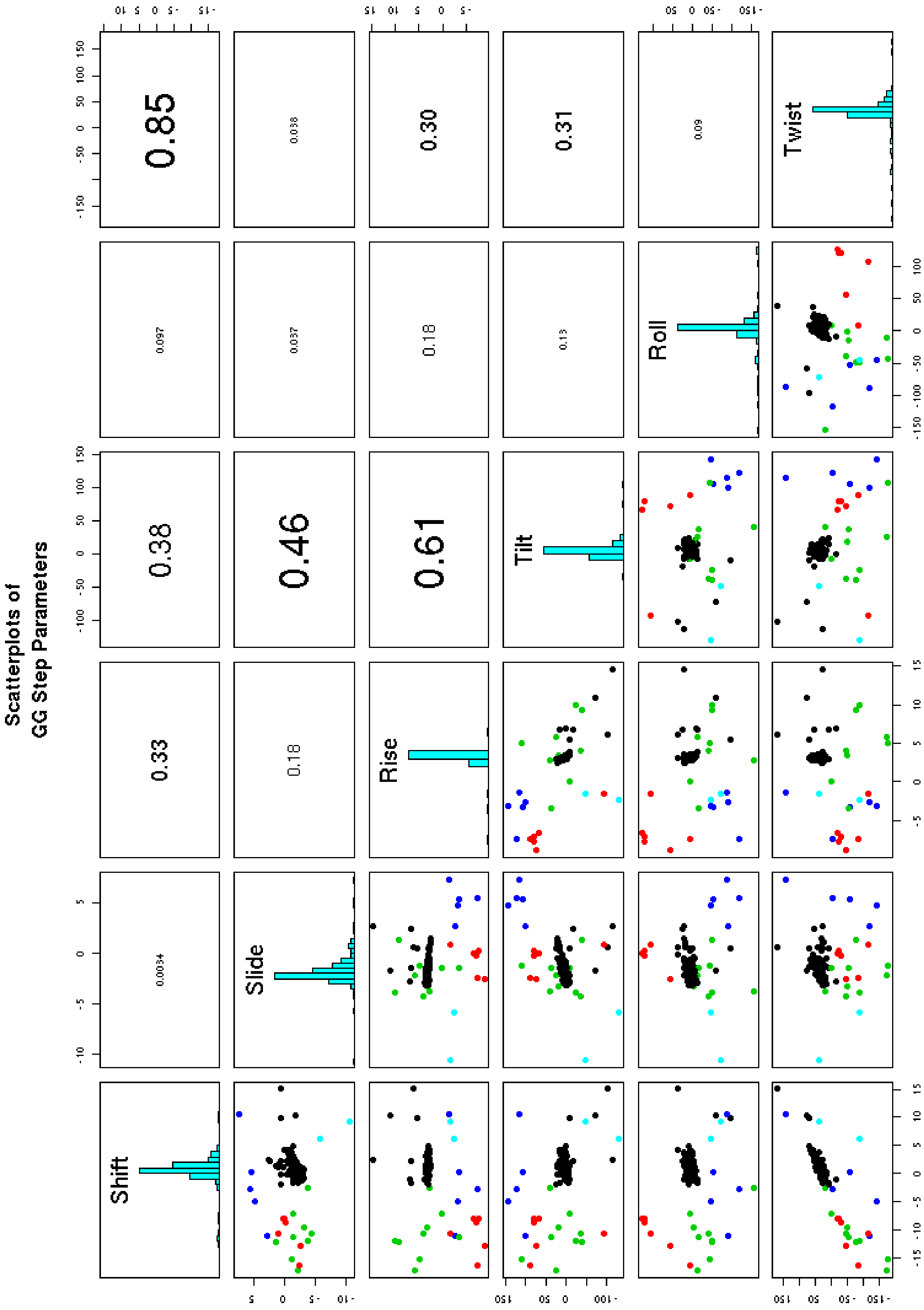

Figure 2.4: Scatterplots for step-parameters of **All** GG dinucleotide steps in 50S rRNA.

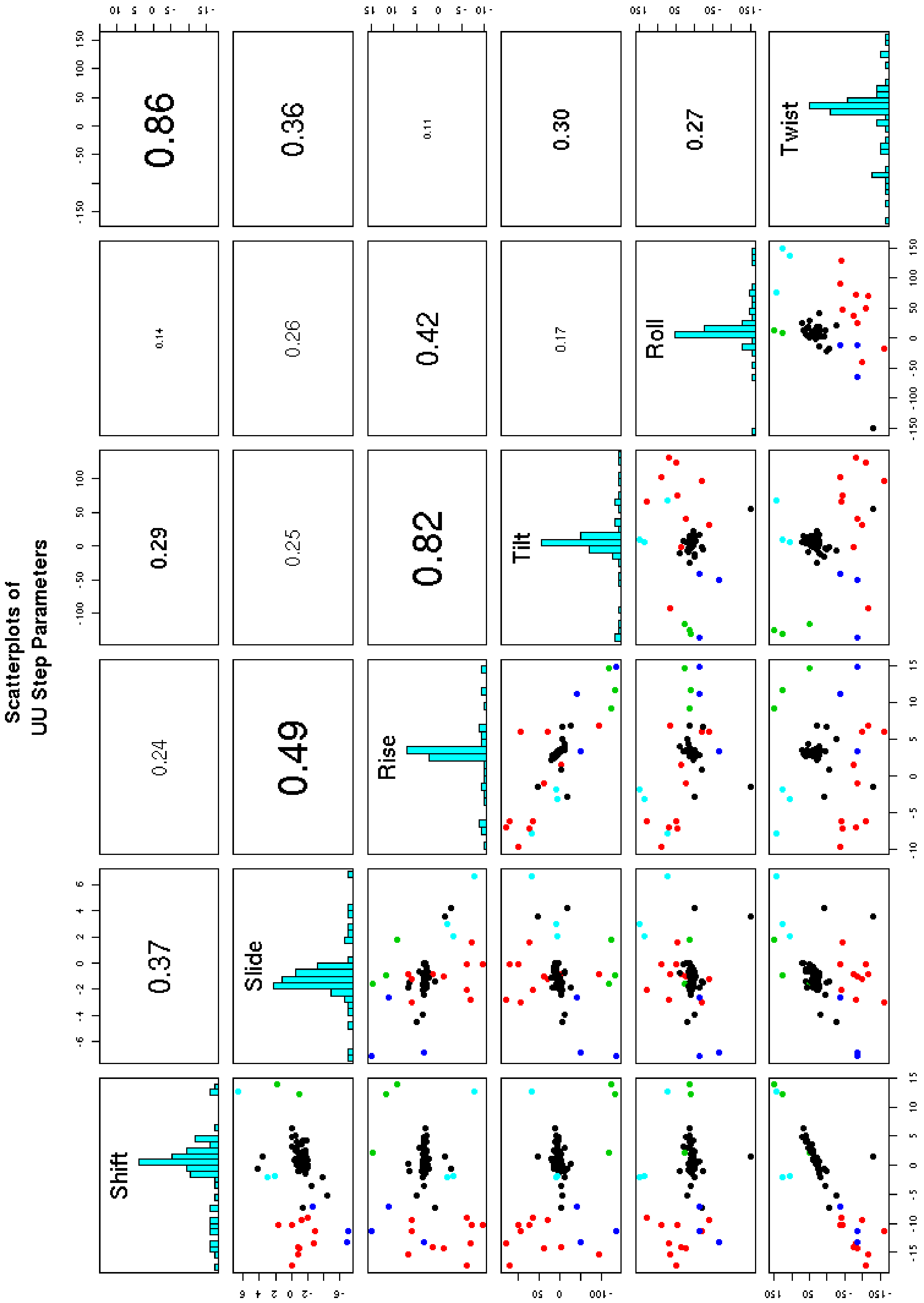

Figure 2.5: Scatterplots for step-parameters of **All** UU dinucleotide steps in 50S rRNA.

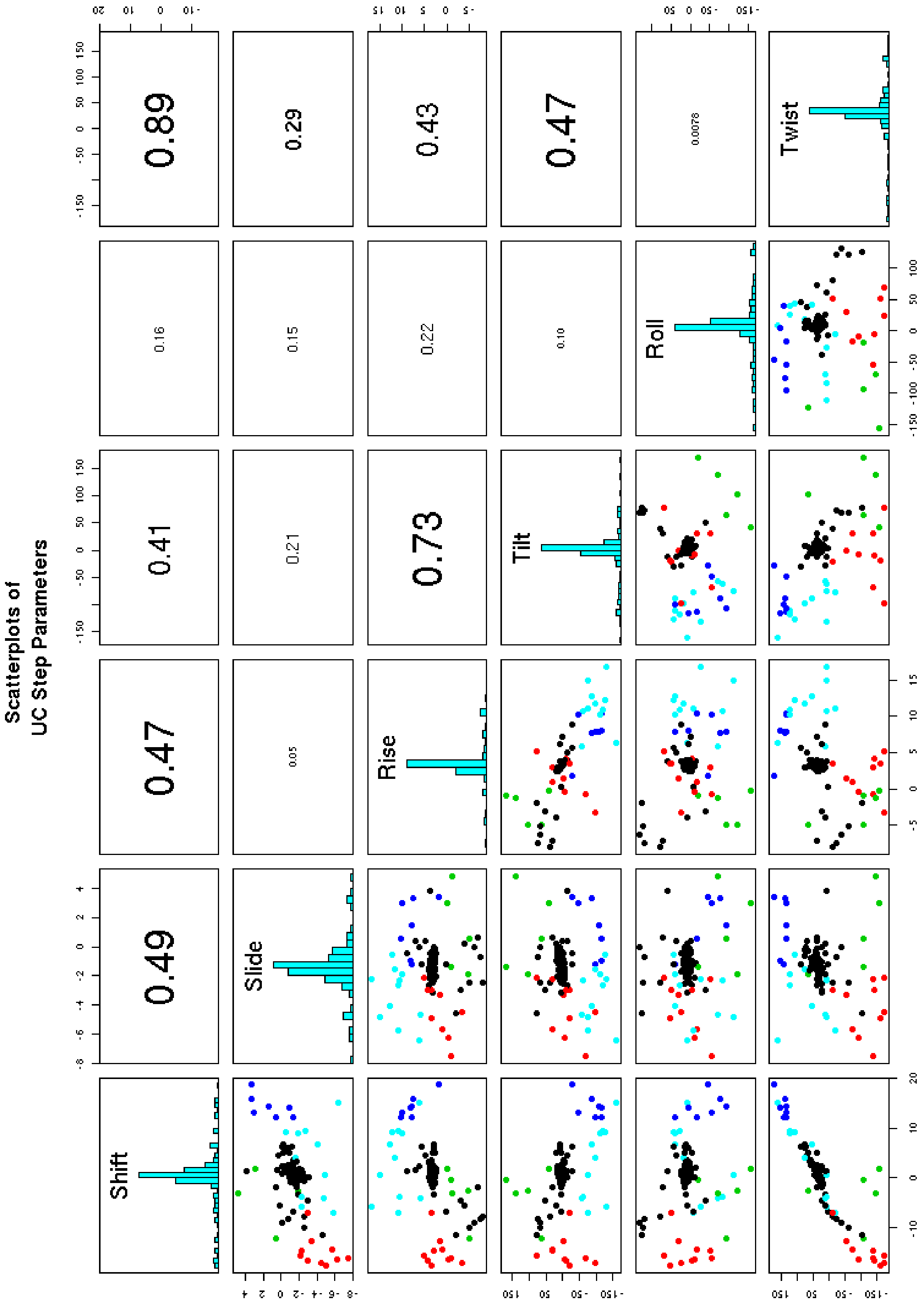

Figure 2.6: Scatterplots for step-parameters of **All** UC dinucleotide steps in 50S rRNA.

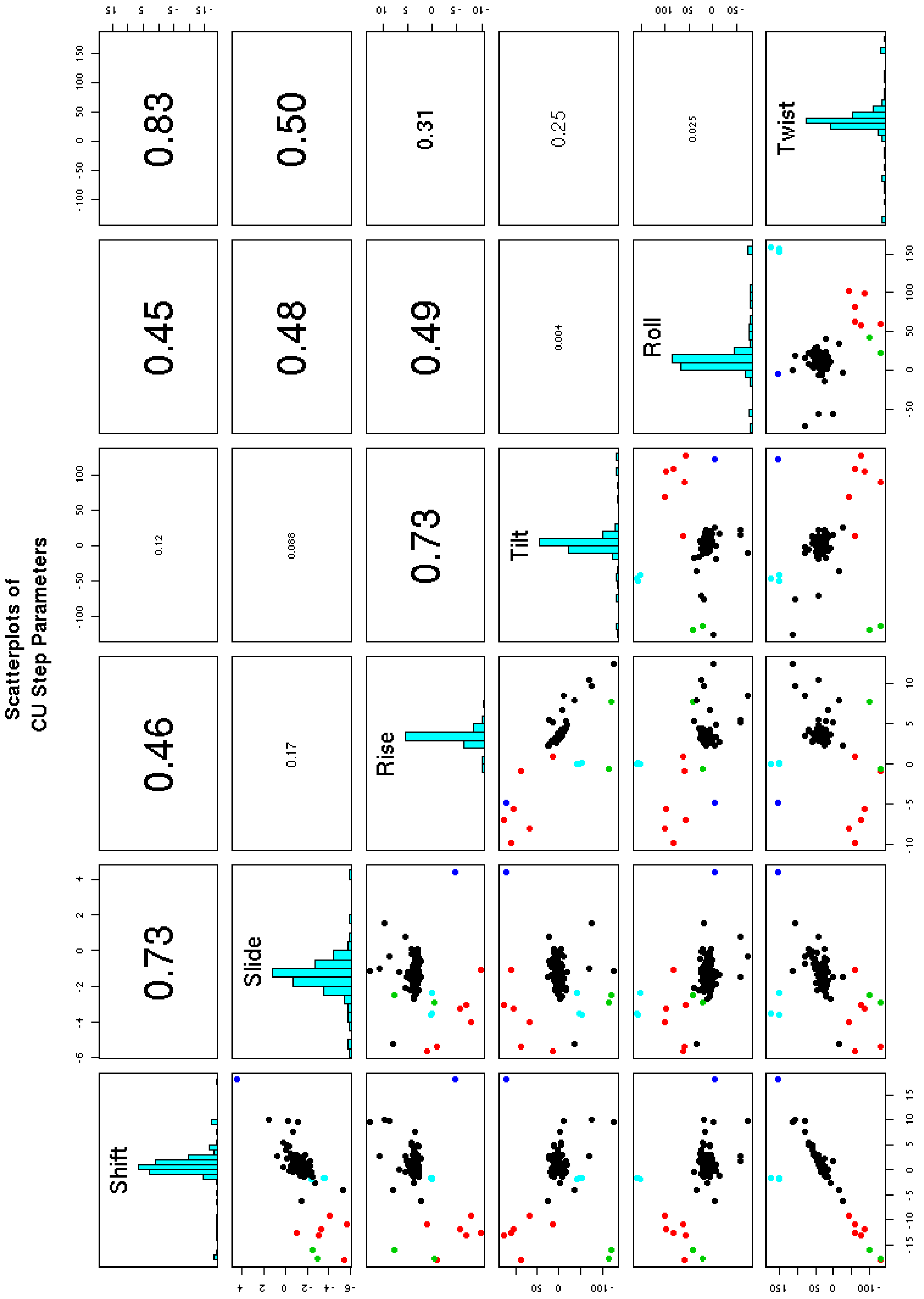

Figure 2.7: Scatterplots for step-parameters of **All** CU dinucleotide steps in 50S rRNA.

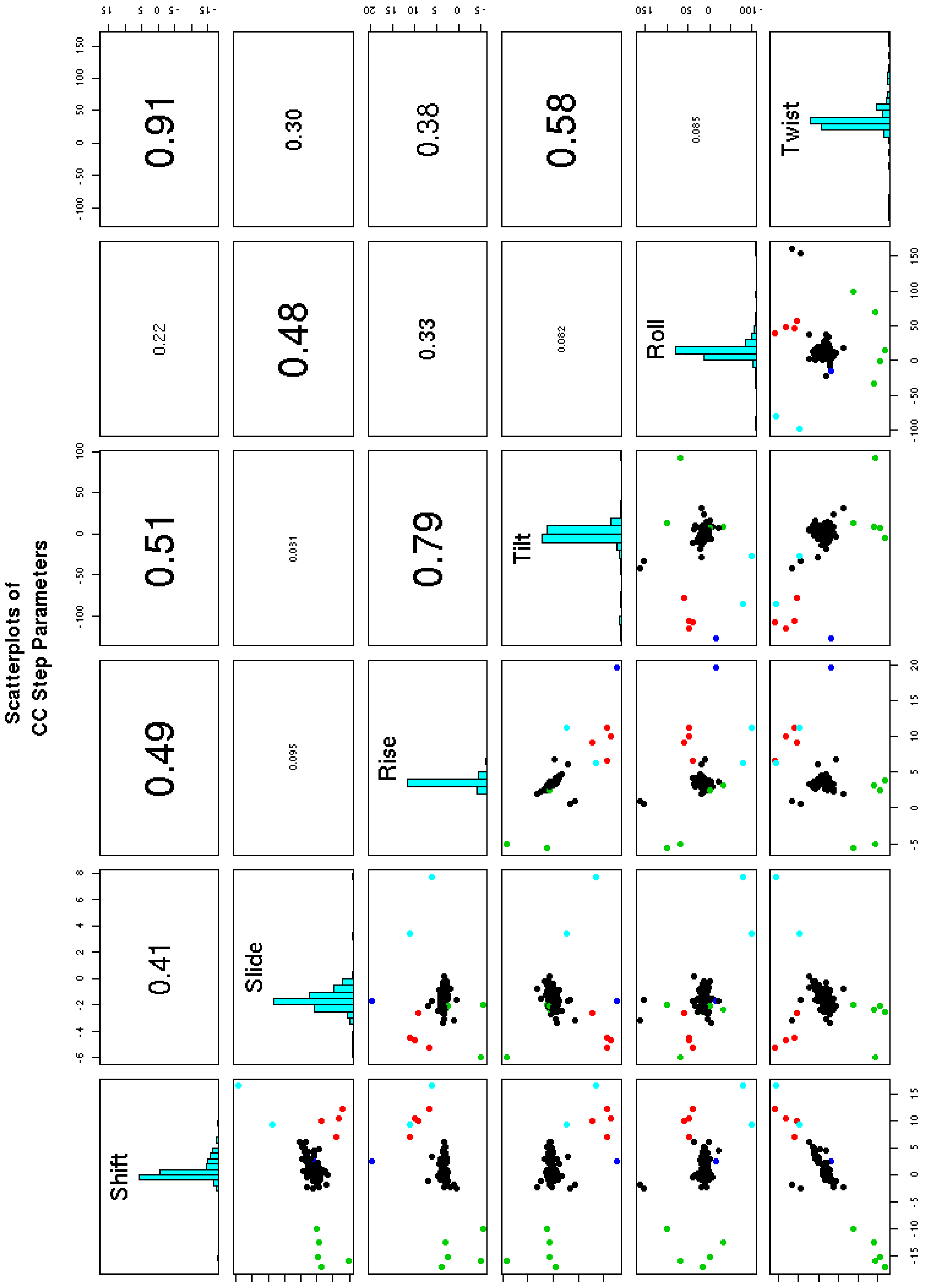

Figure 2.8: Scatterplots for step-parameters of **All** CC dinucleotide steps in 50S rRNA.

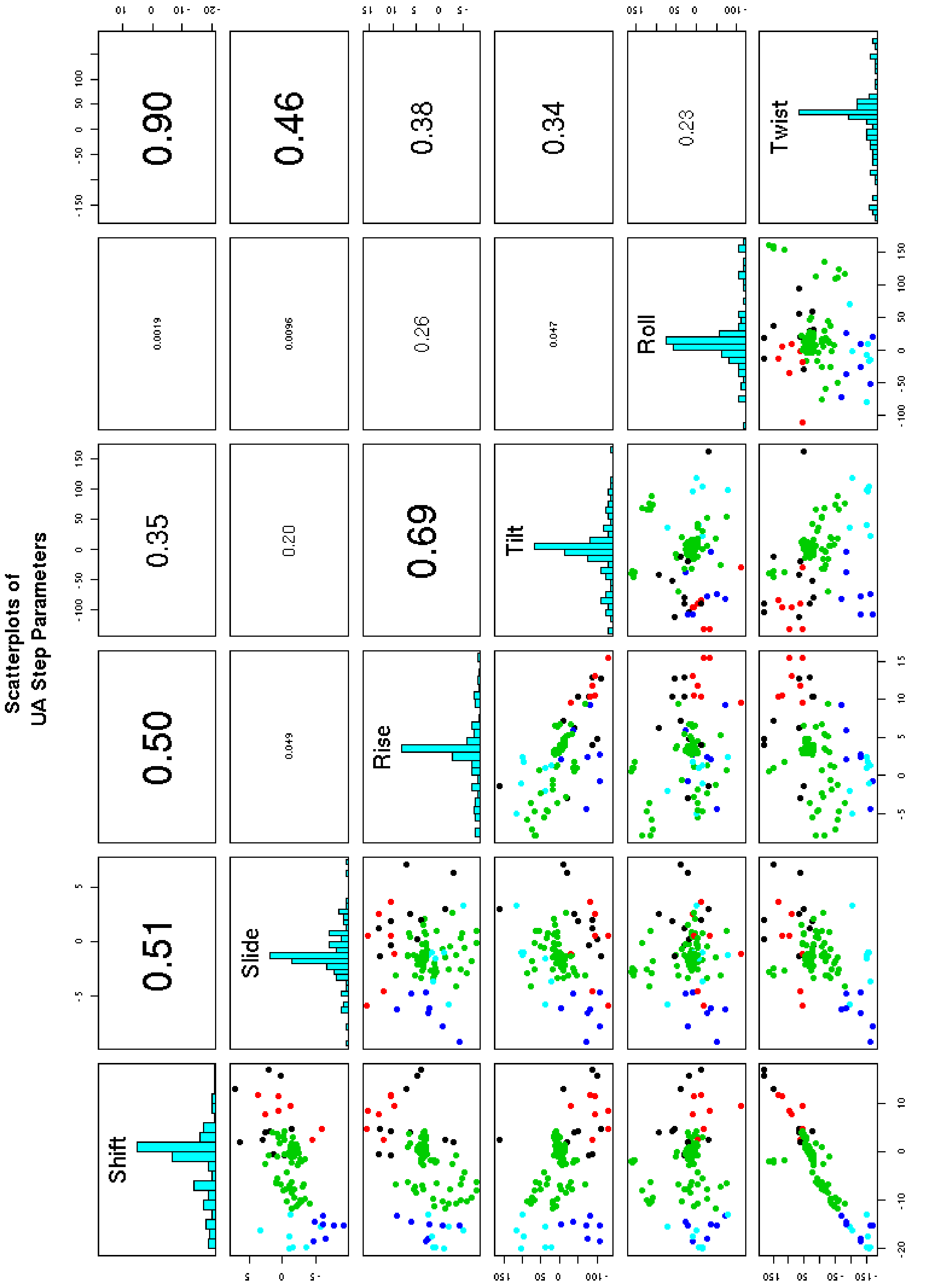

Figure 2.9: Scatterplots for step-parameters of **All** UA dinucleotide steps in 50S rRNA.

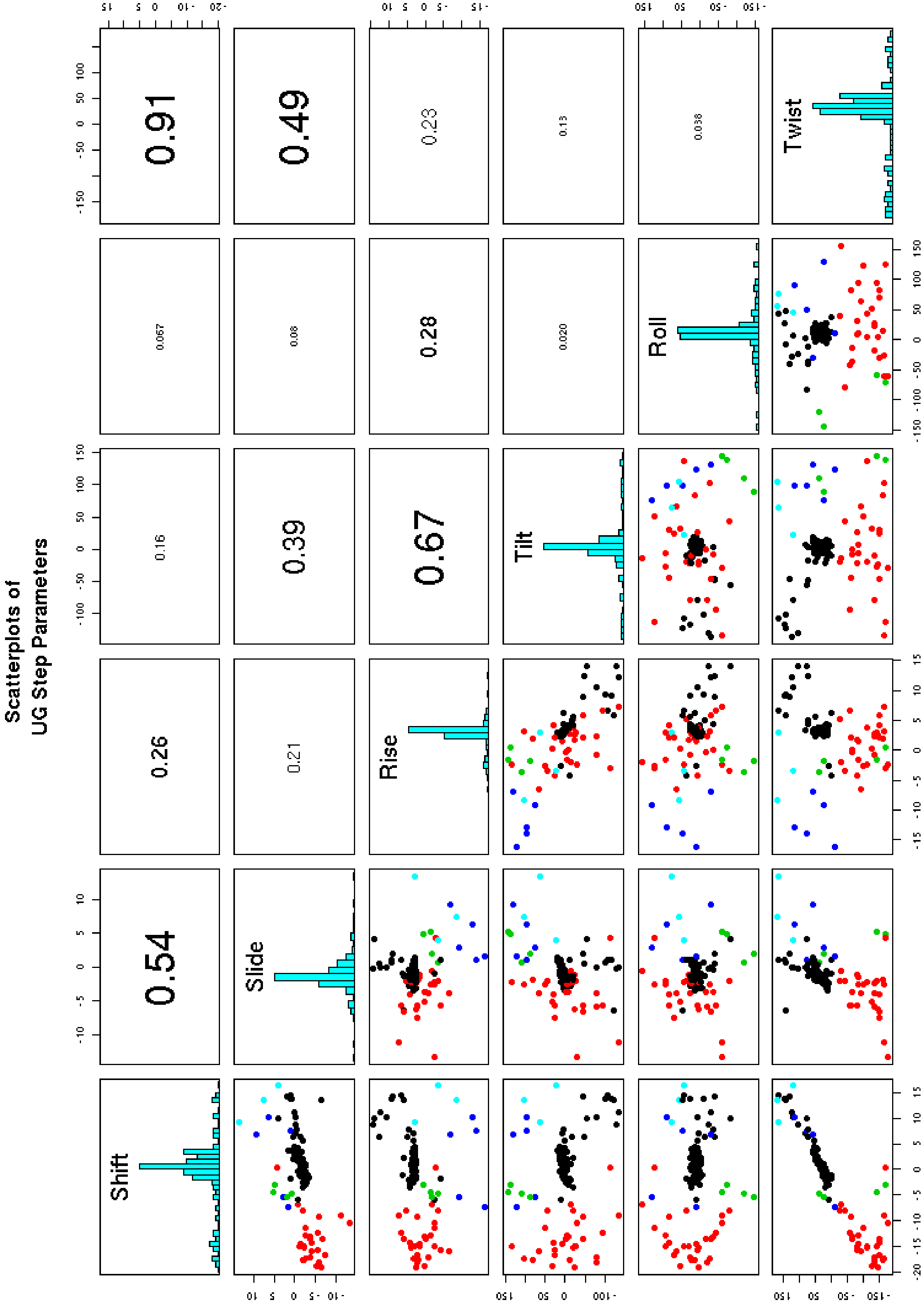

Figure 2.10: Scatterplots for step-parameters of **All** UG dinucleotide steps in 50S rRNA.

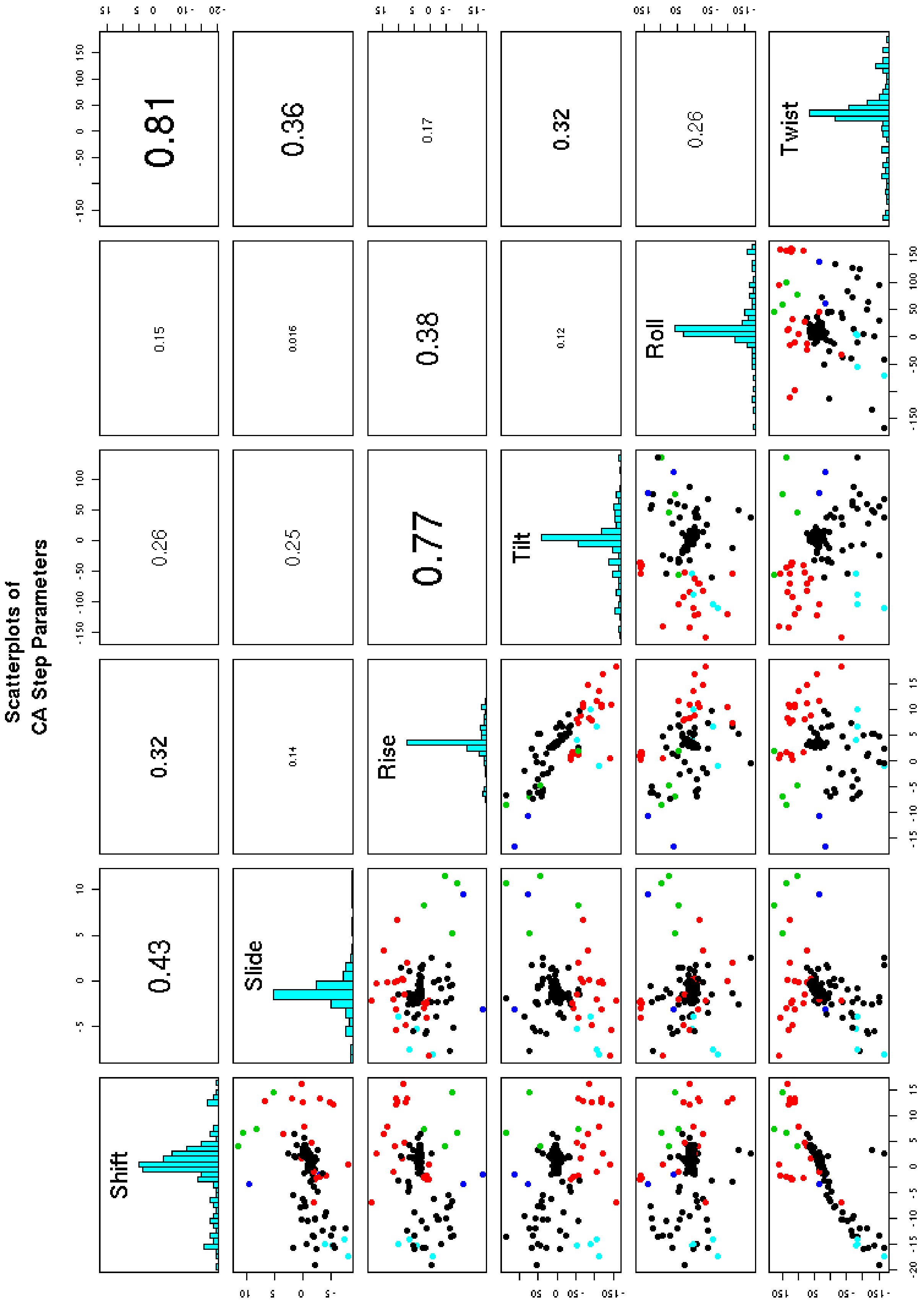

Figure 2.11: Scatterplots for step-parameters of **All** CA dinucleotide steps in 50S rRNA.

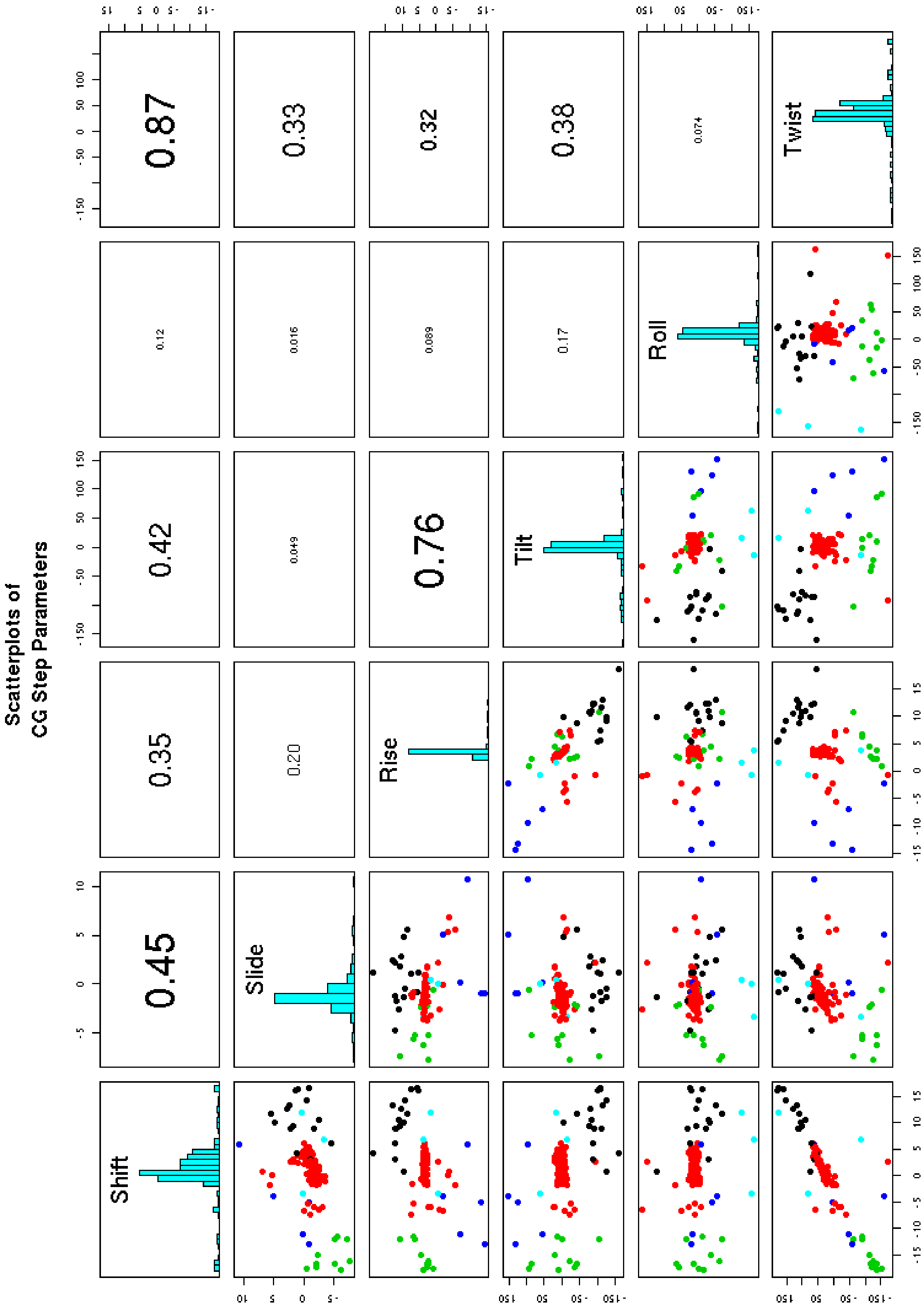

Figure 2.12: Scatterplots for step-parameters of **All** CG dinucleotide steps in 50S rRNA.

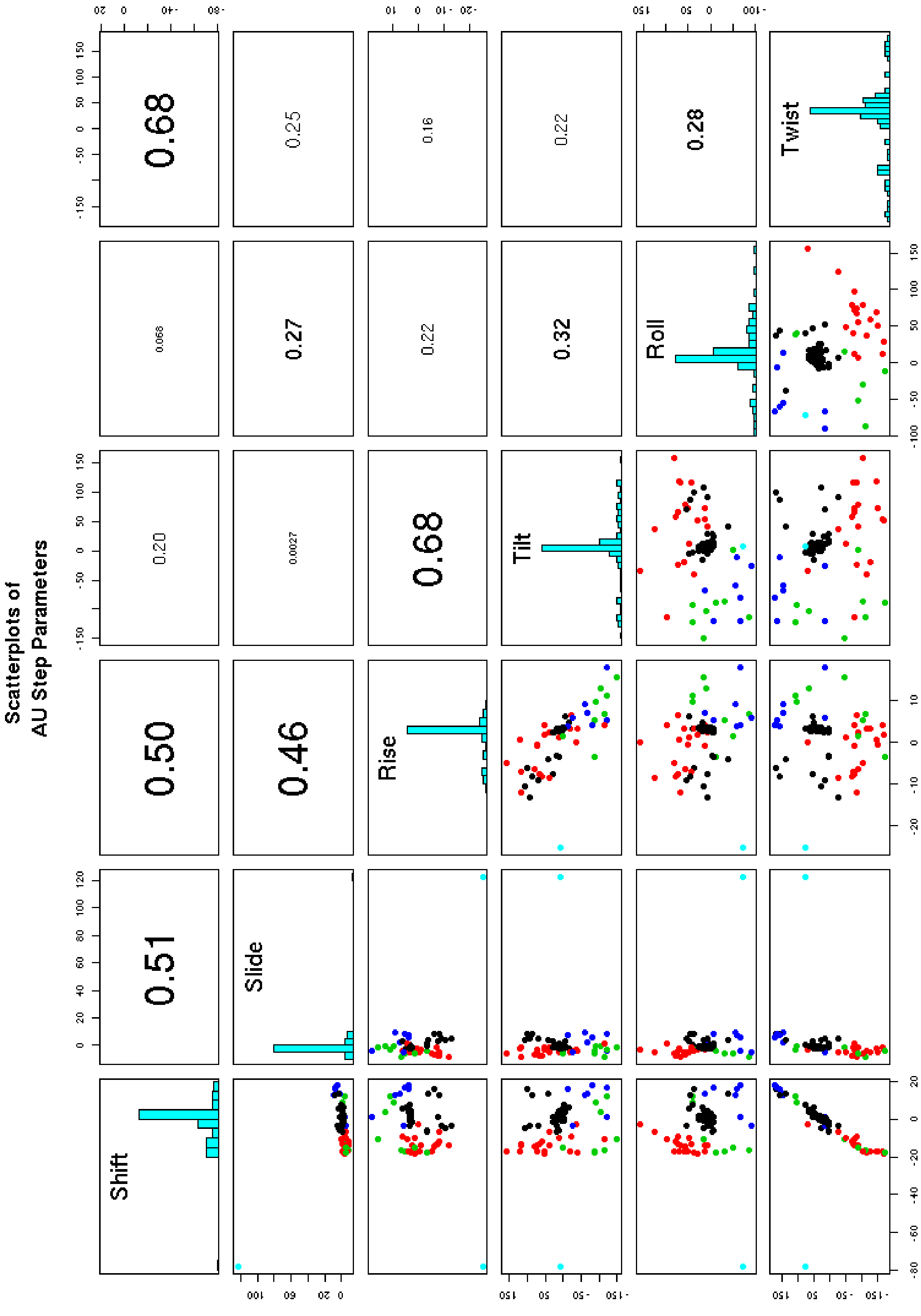

Figure 2.13: Scatterplots for step-parameters of **All** AU dinucleotide steps in 50S rRNA.

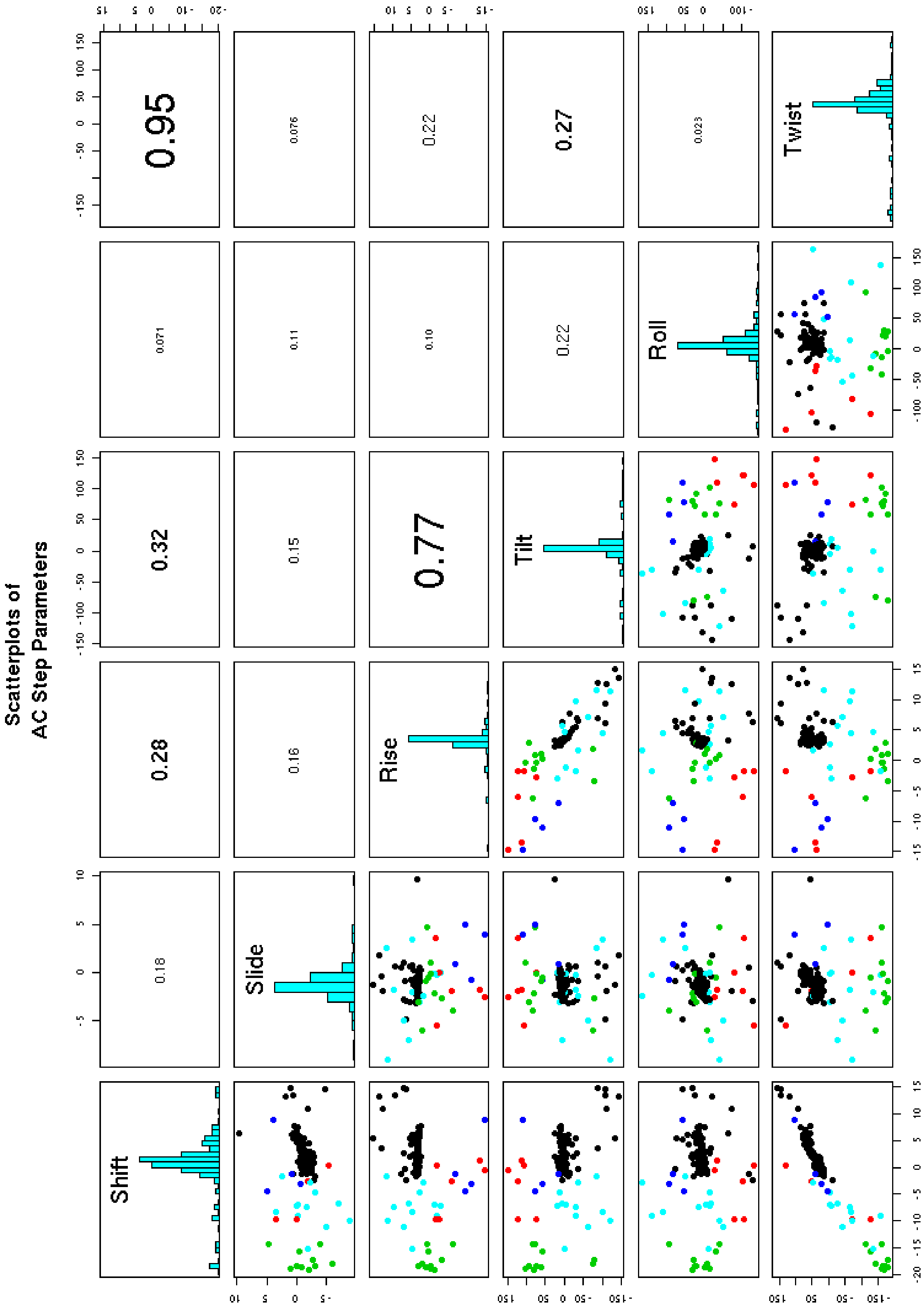

Figure 2.14: Scatterplots for step-parameters of **All** AC dinucleotide steps in 50S rRNA.

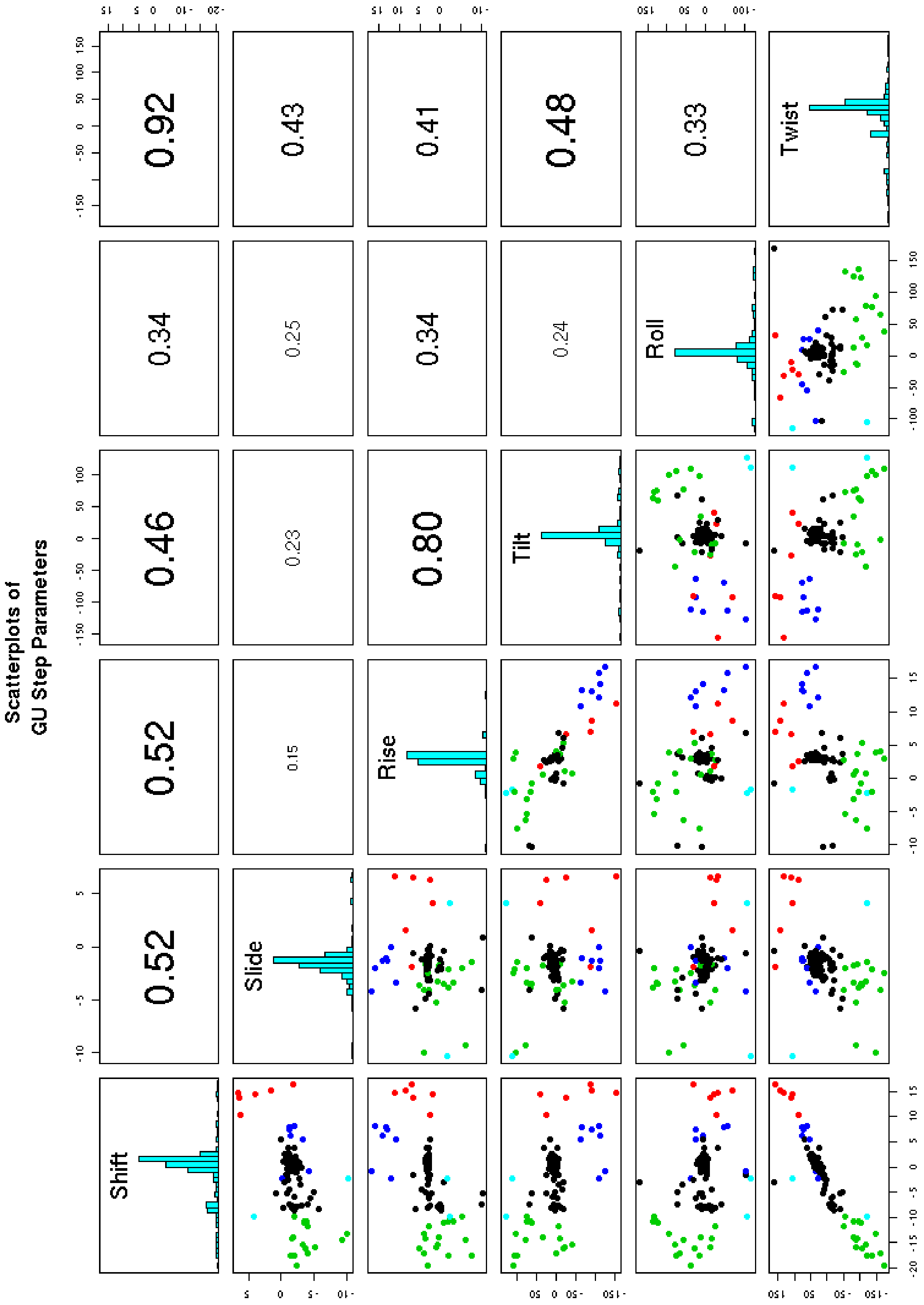

Figure 2.15: Scatterplots for step-parameters of **All** GU dinucleotide steps in 50S rRNA.

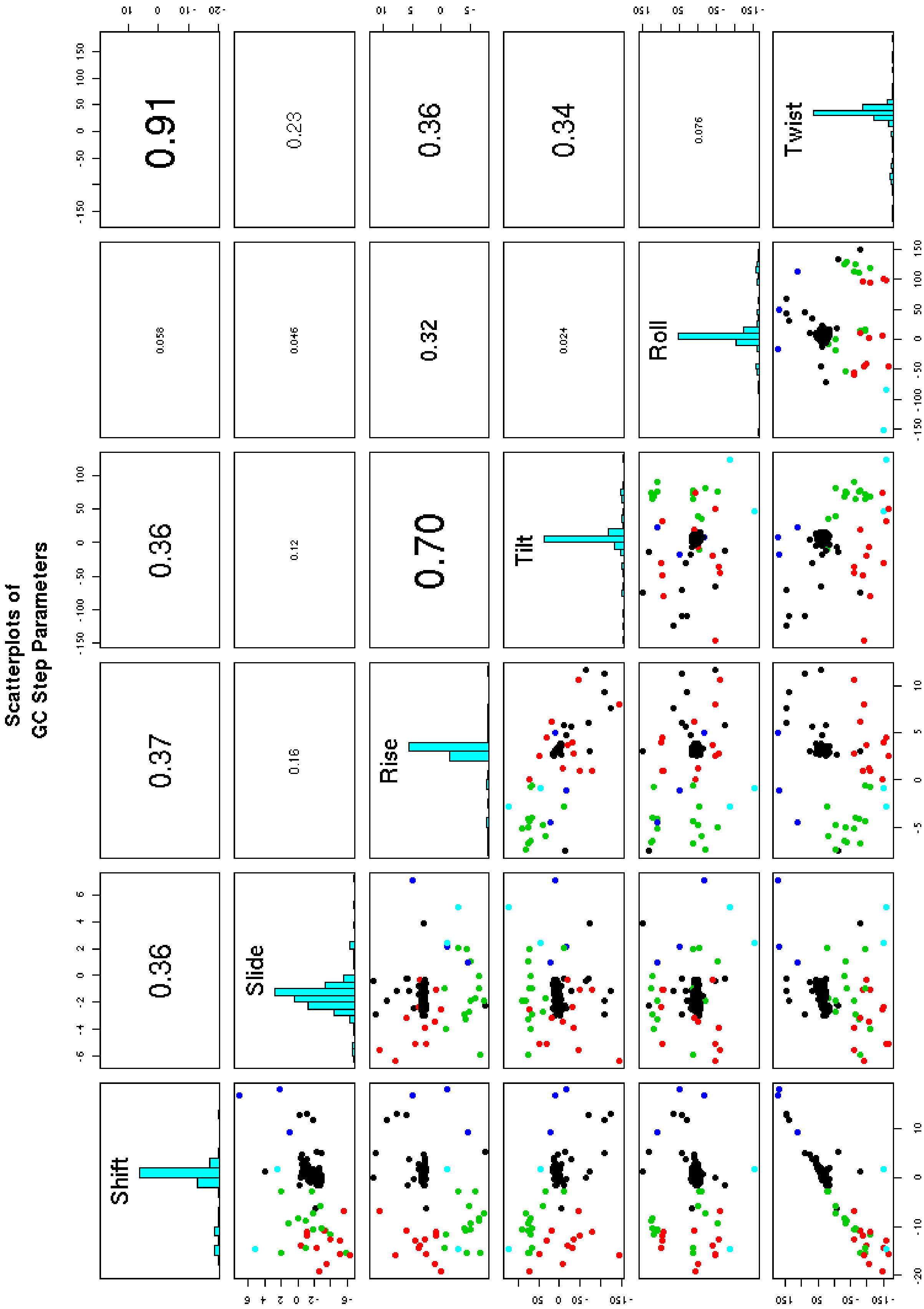

Figure 2.16: Scatterplots for step-parameters of **All** GC dinucleotide steps in 50S rRNA.

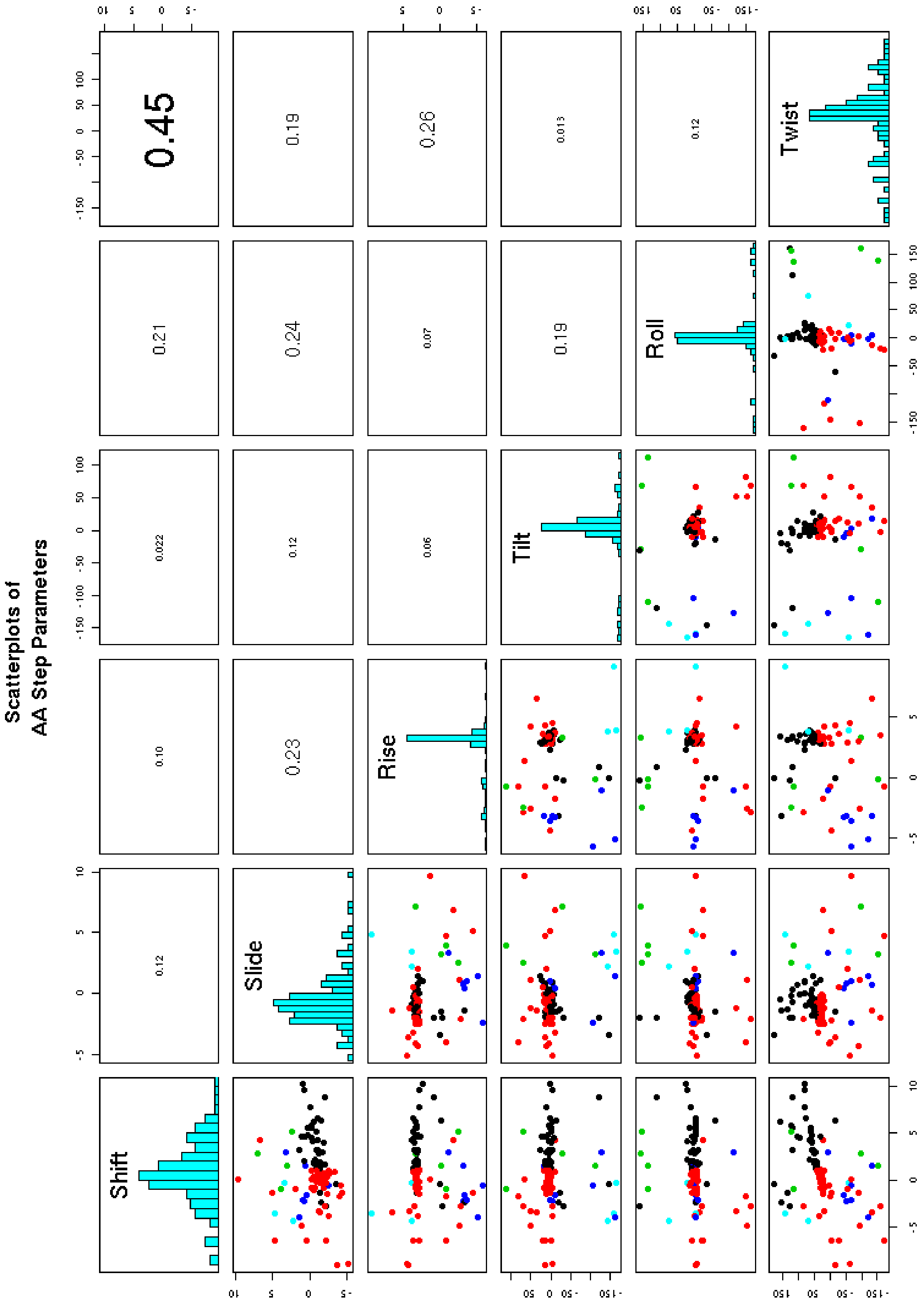

Figure 2.17: Scatterplots for step-parameters of **Helical** AA dinucleotide steps in 50S rRNA.

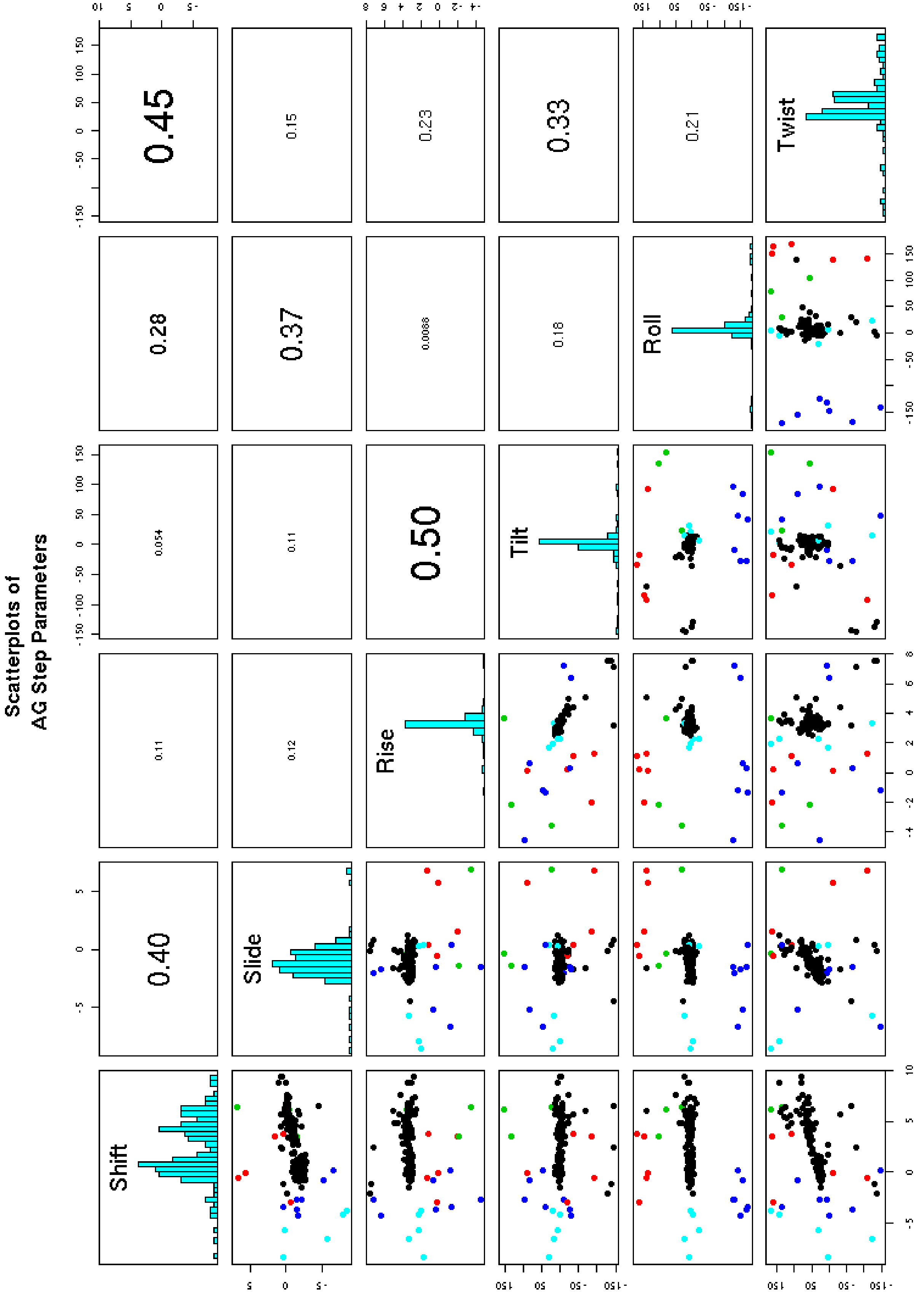

Figure 2.18: Scatterplots for step-parameters of **Helical** AG dinucleotide steps in 50S rRNA.

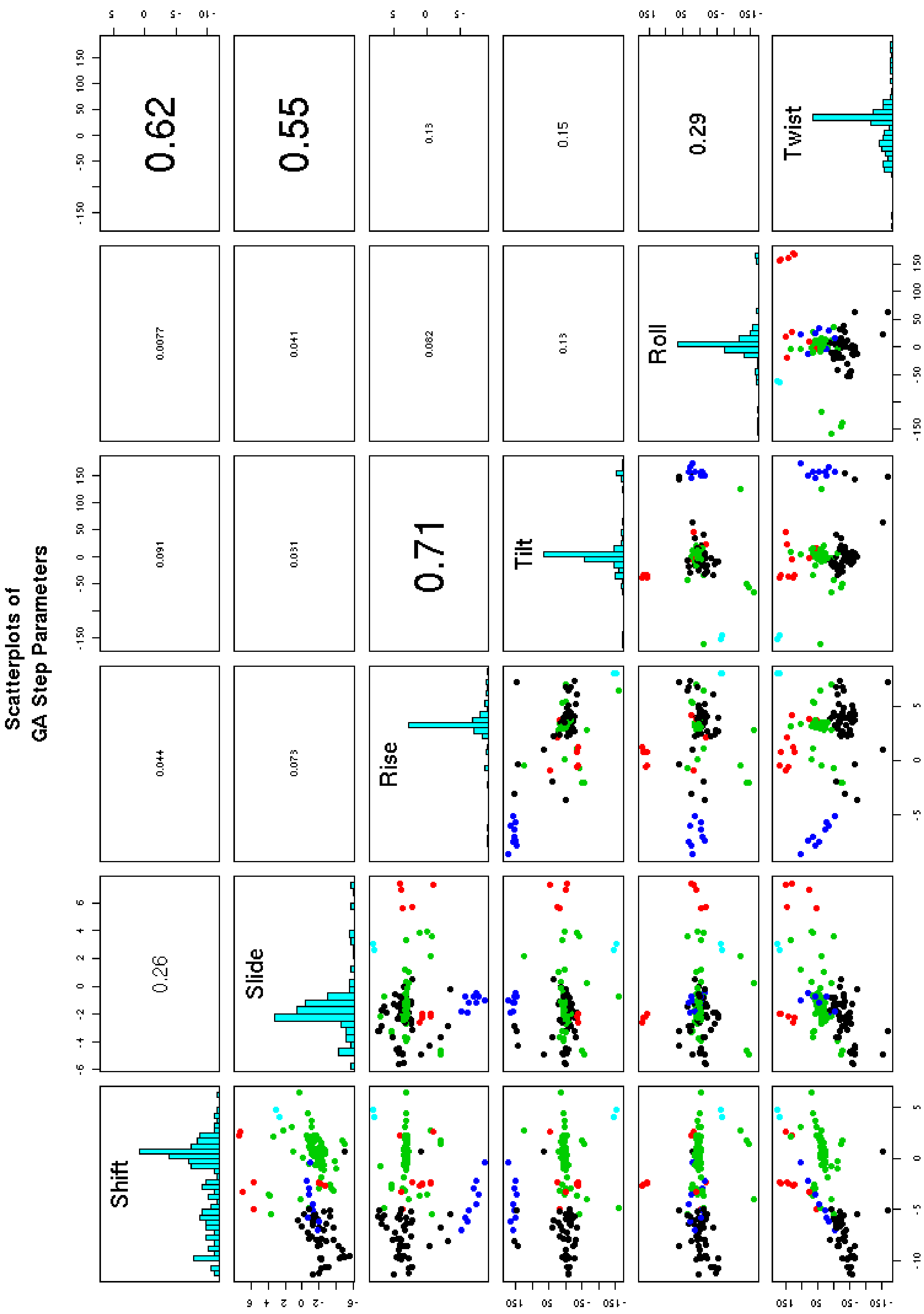

Figure 2.19: Scatterplots for step-parameters of **Helical** GA dinucleotide steps in 50S rRNA.

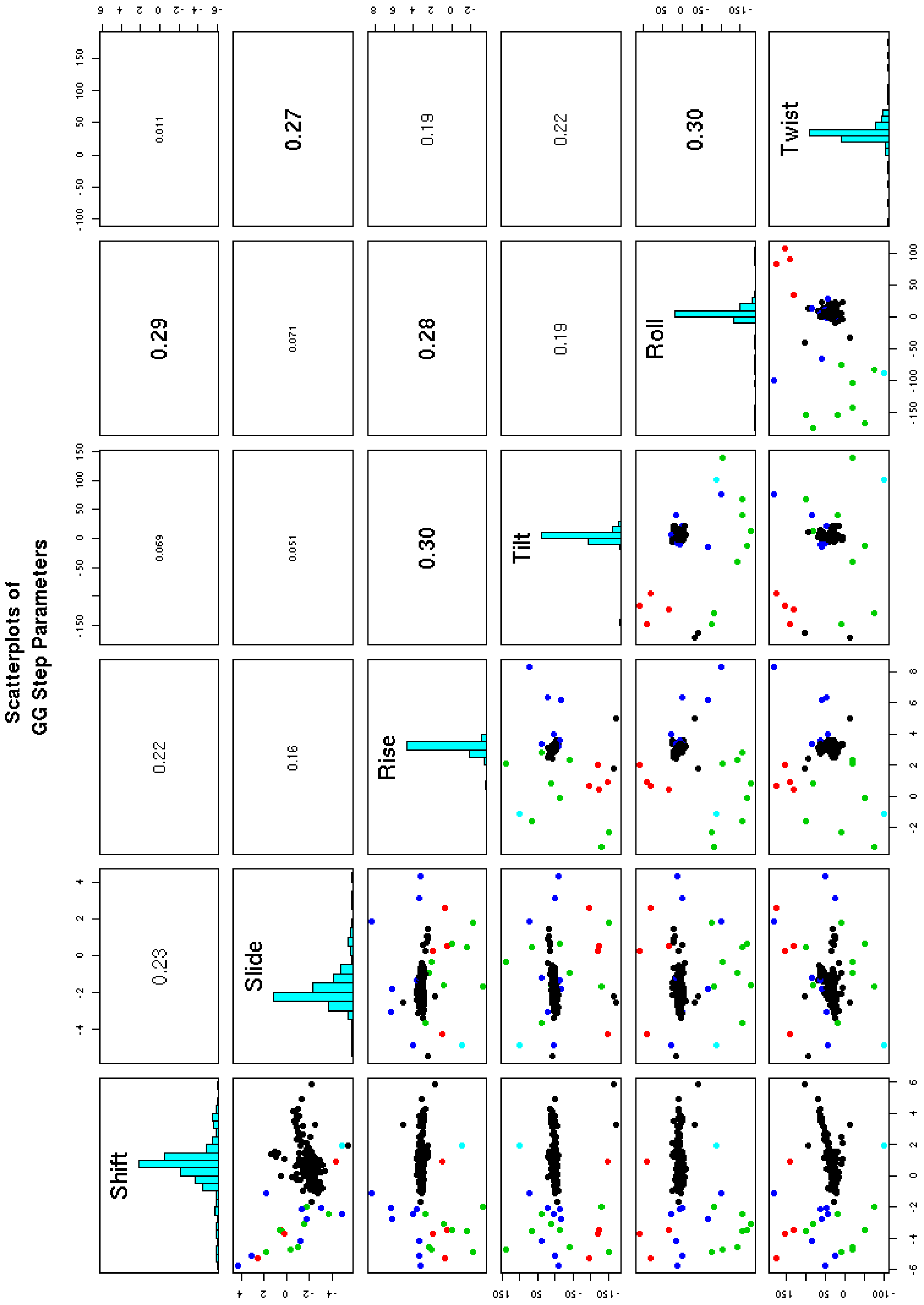

Figure 2.20: Scatterplots for step-parameters of **Helical** GG dinucleotide steps in 50S rRNA.

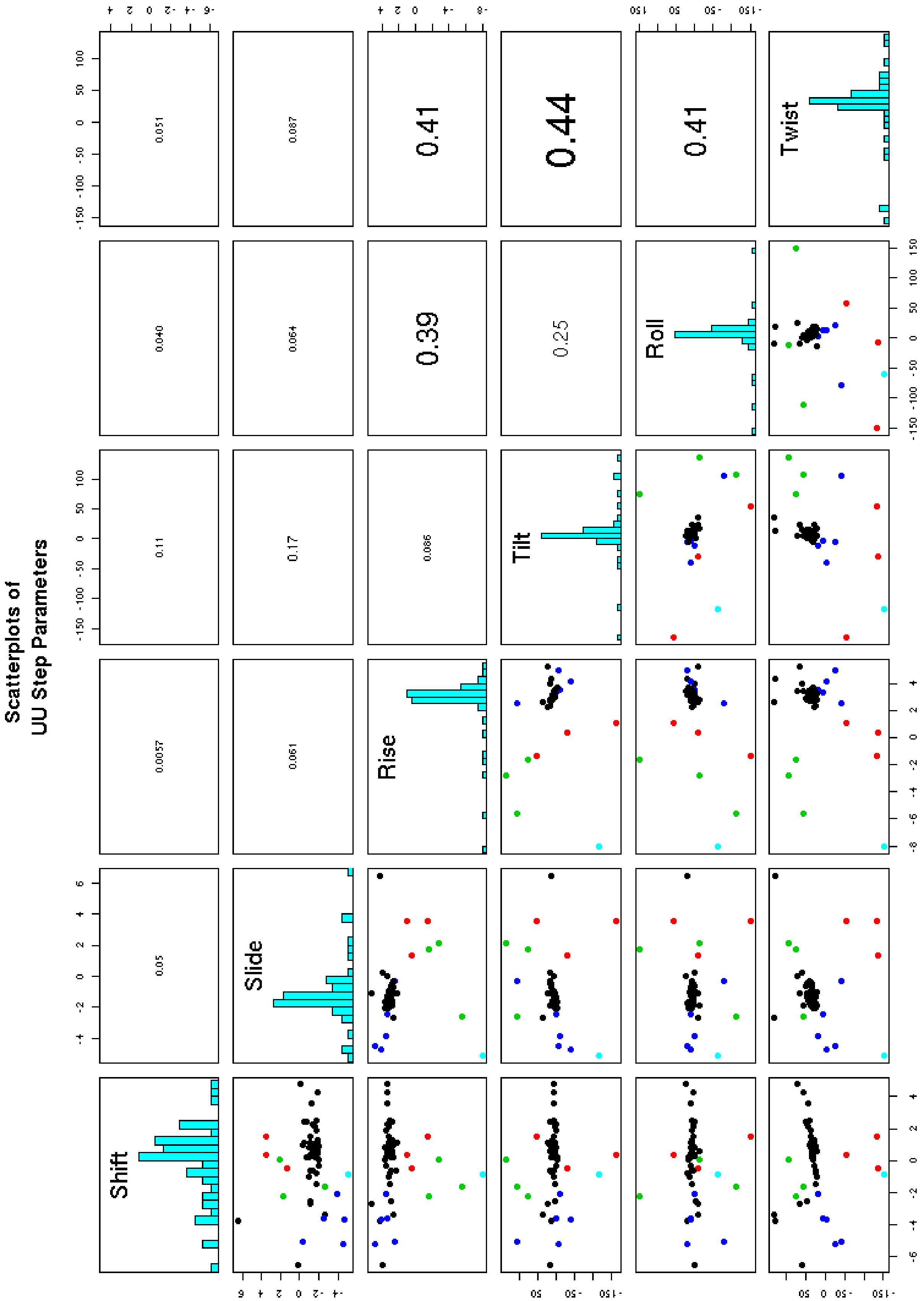

Figure 2.21: Scatterplots for step-parameters of **Helical** UU dinucleotide steps in 50S rRNA.

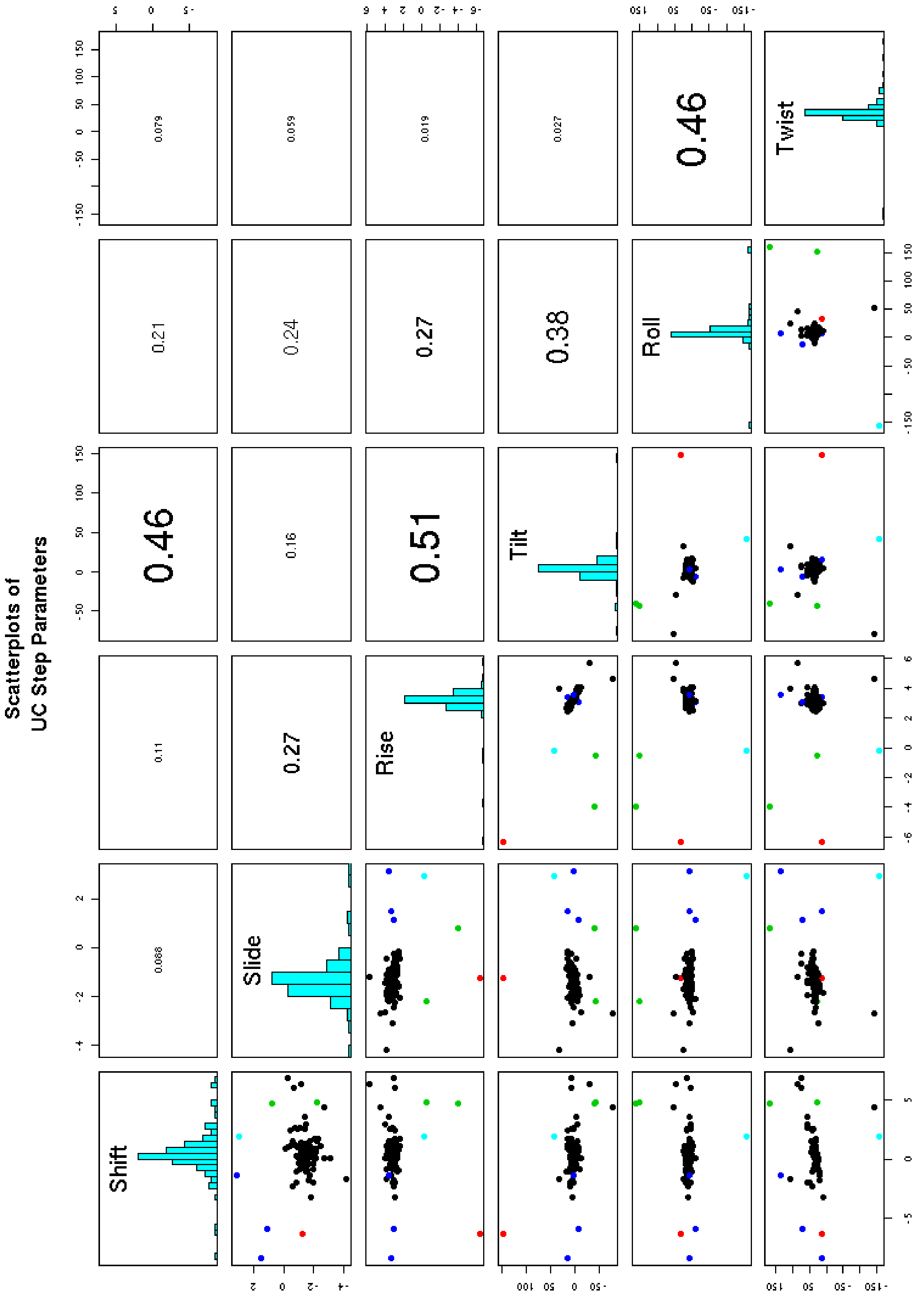

Figure 2.22: Scatterplots for step-parameters of **Helical** UC dinucleotide steps in 50S rRNA.

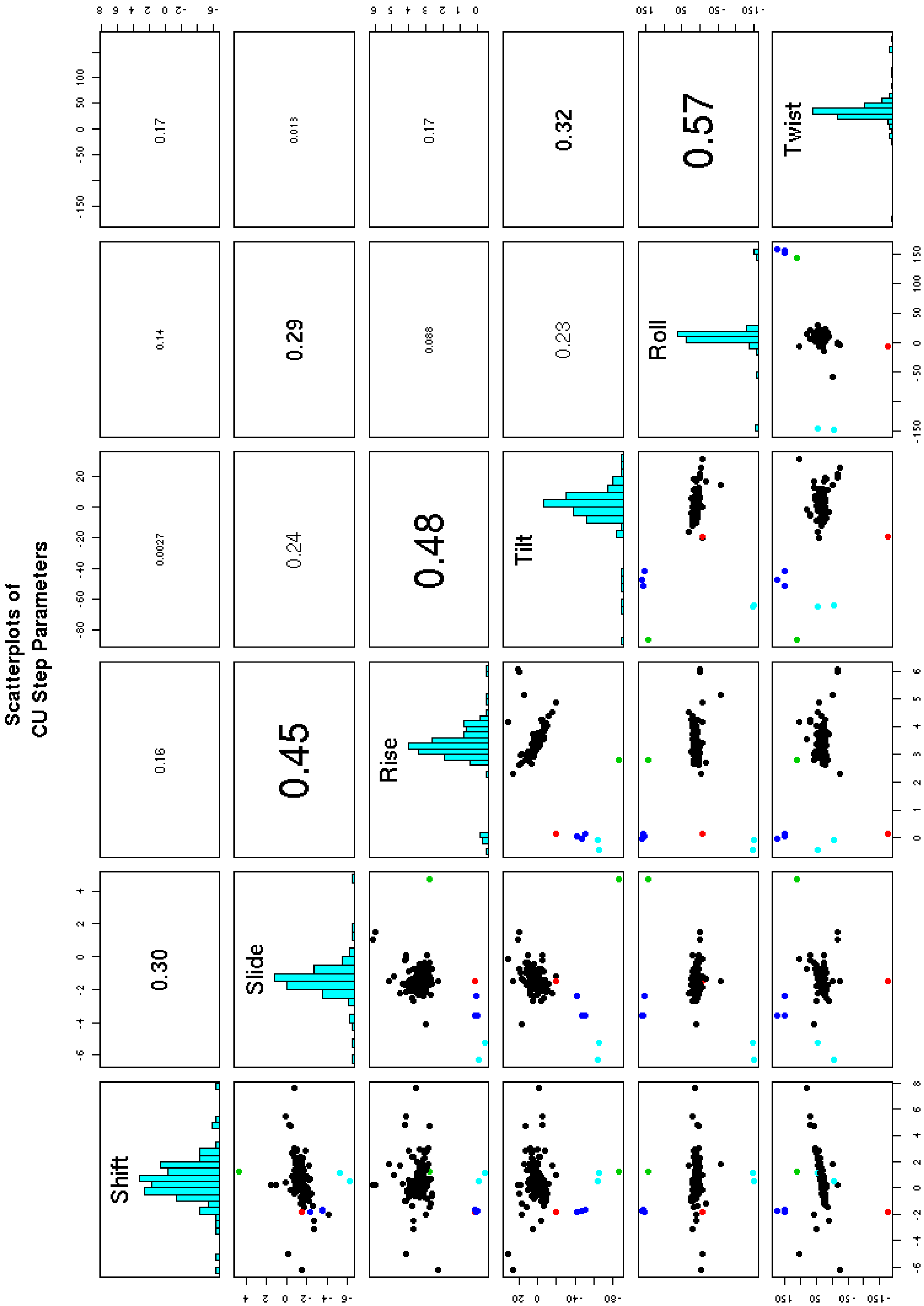

Figure 2.23: Scatterplots for step-parameters of **Helical** CU dinucleotide steps in 50S rRNA.

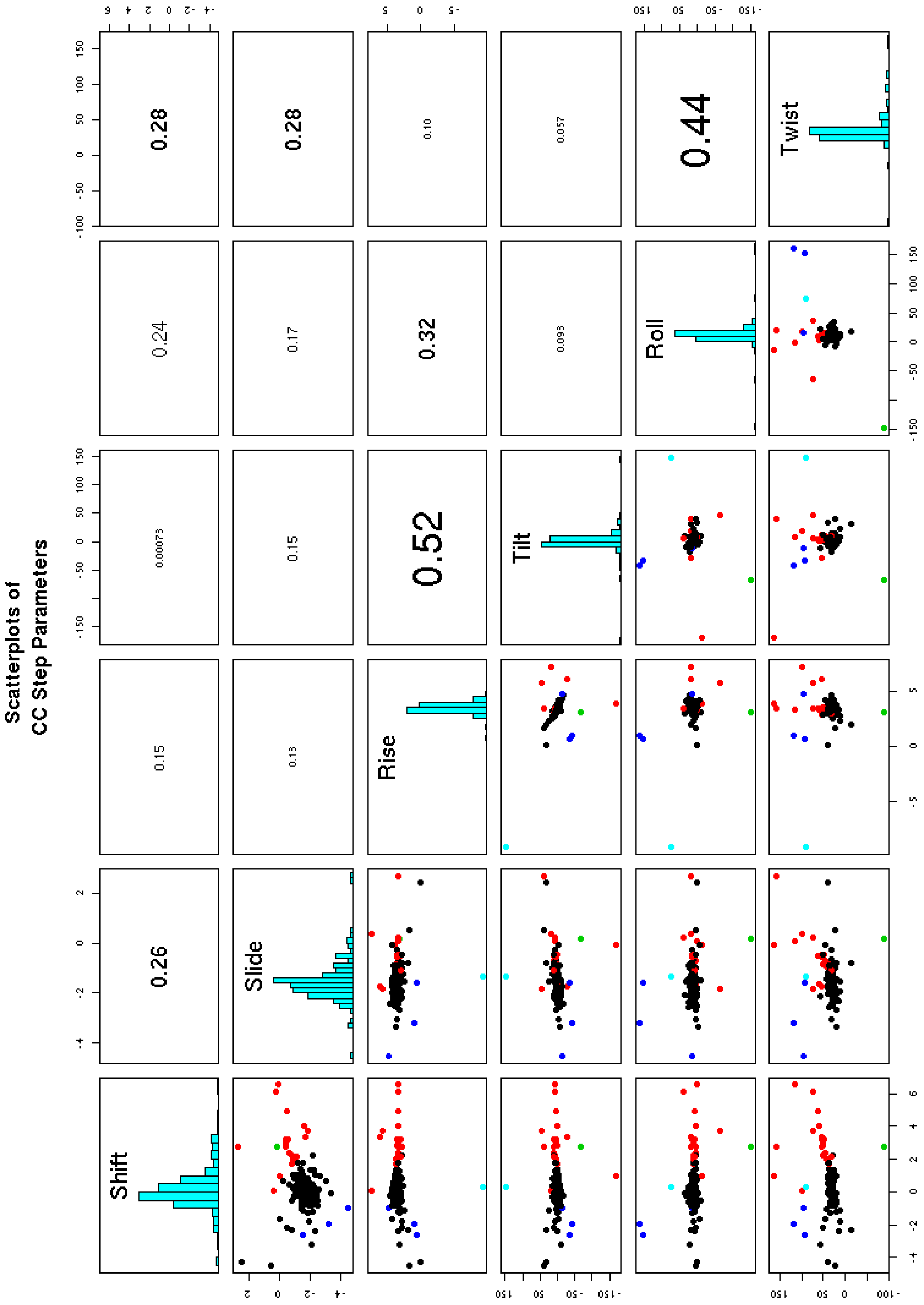

Figure 2.24: Scatterplots for step-parameters of **Helical** CC dinucleotide steps in 50S rRNA.

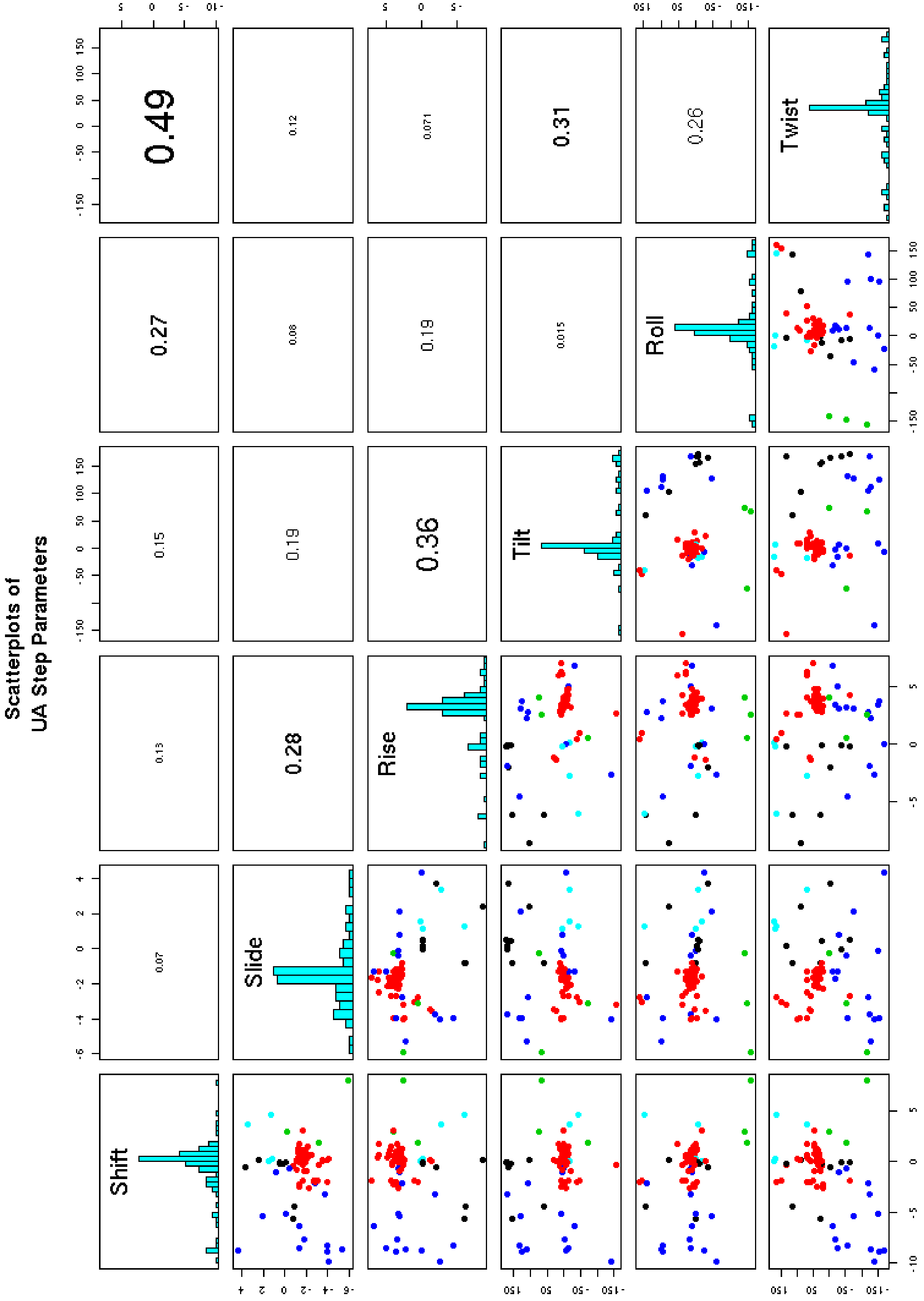

Figure 2.25: Scatterplots for step-parameters of **Helical** UA dinucleotide steps in 50S rRNA.

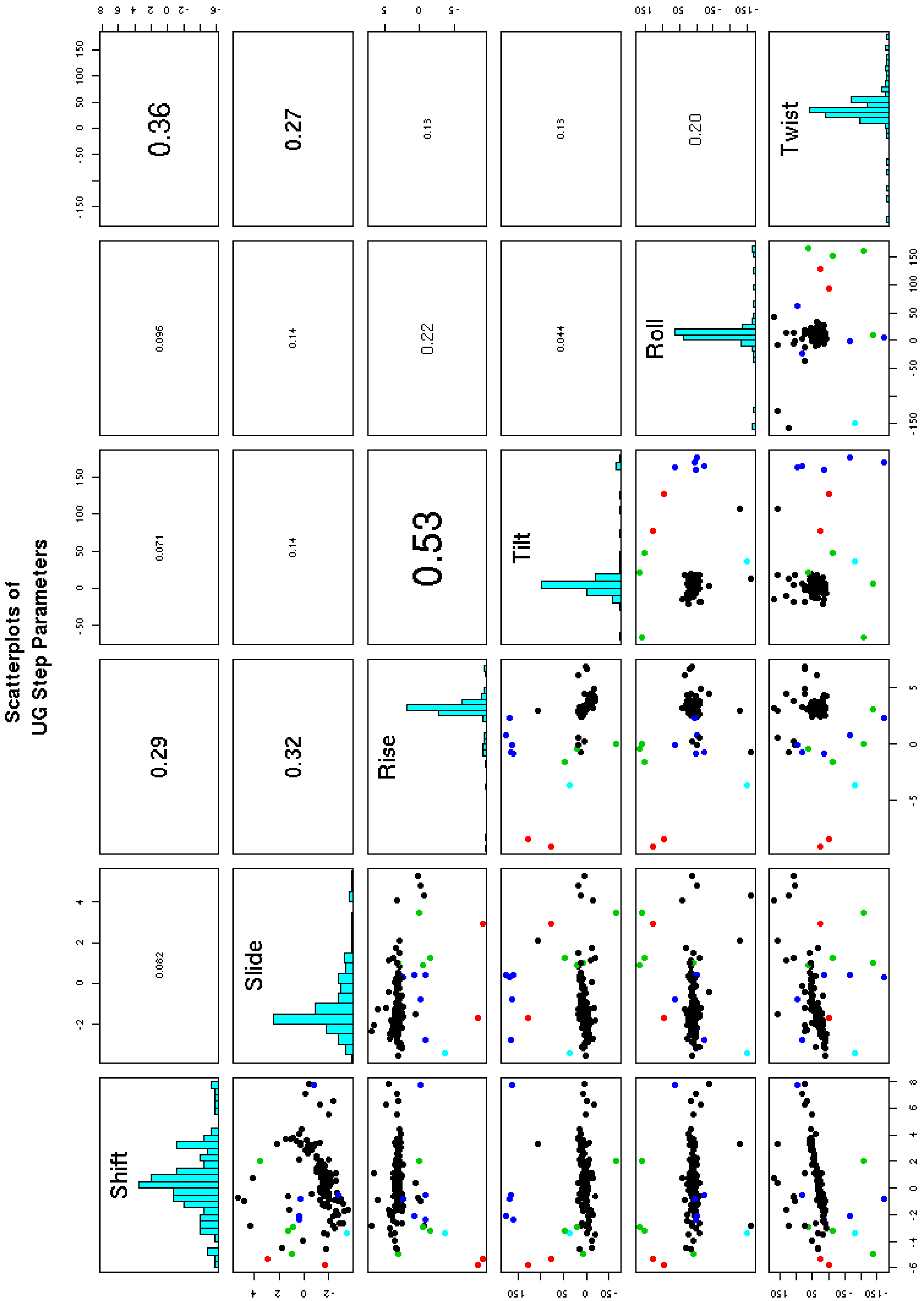

Figure 2.26: Scatterplots for step-parameters of **Helical** UG dinucleotide steps in 50S rRNA.

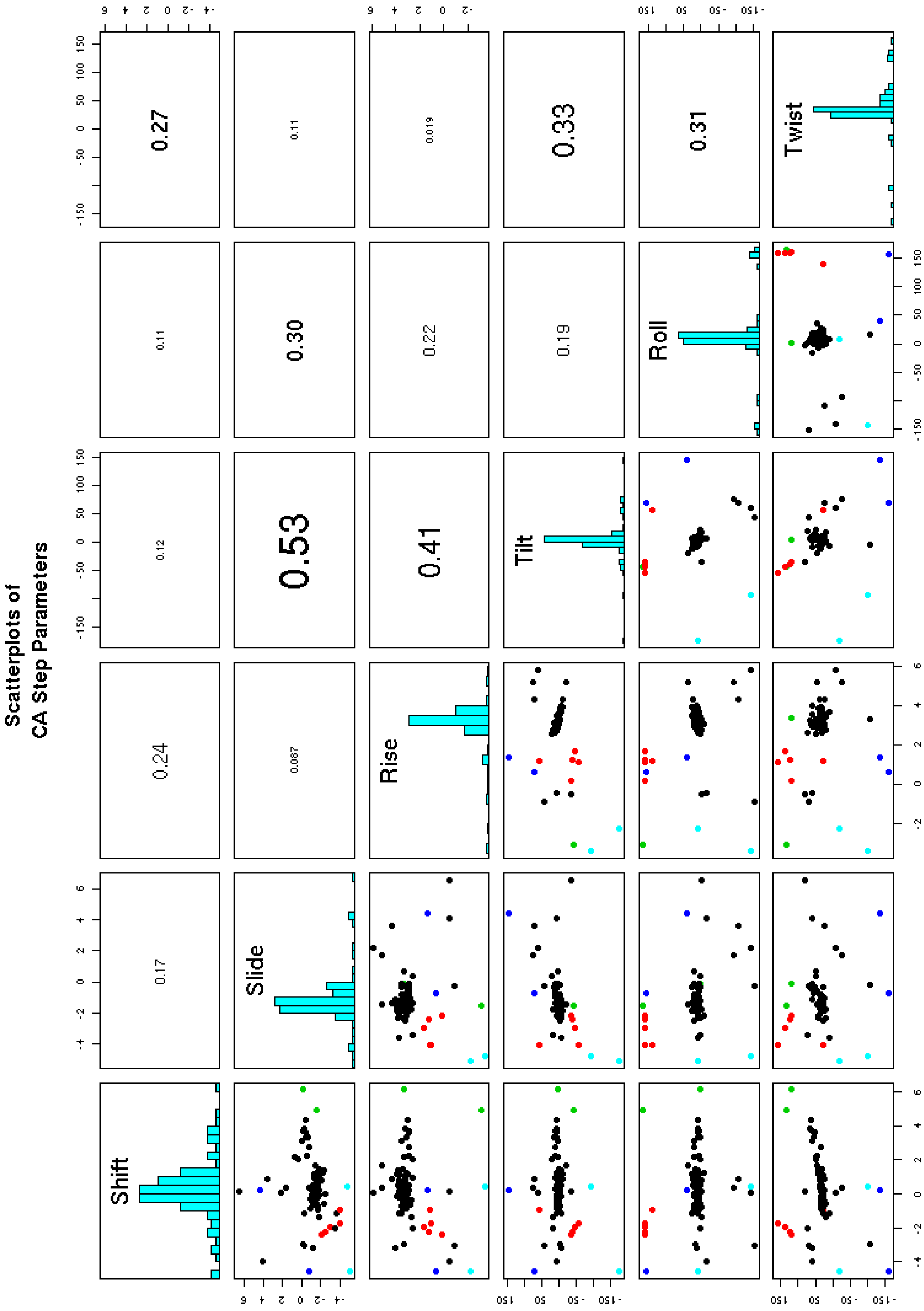

Figure 2.27: Scatterplots for step-parameters of **Helical** CA dinucleotide steps in 50S rRNA.

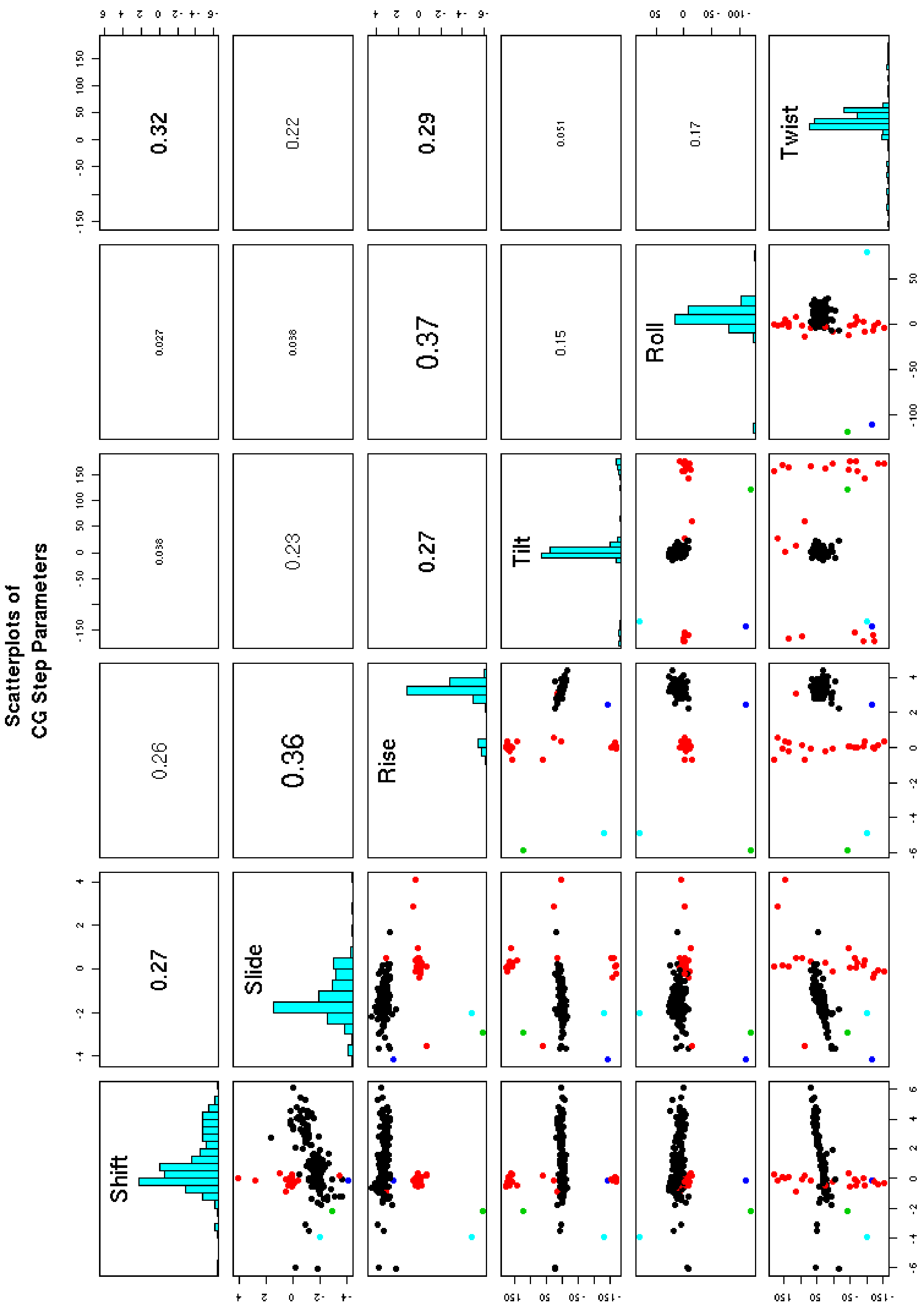

Figure 2.28: Scatterplots for step-parameters of **Helical** CG dinucleotide steps in 50S rRNA.

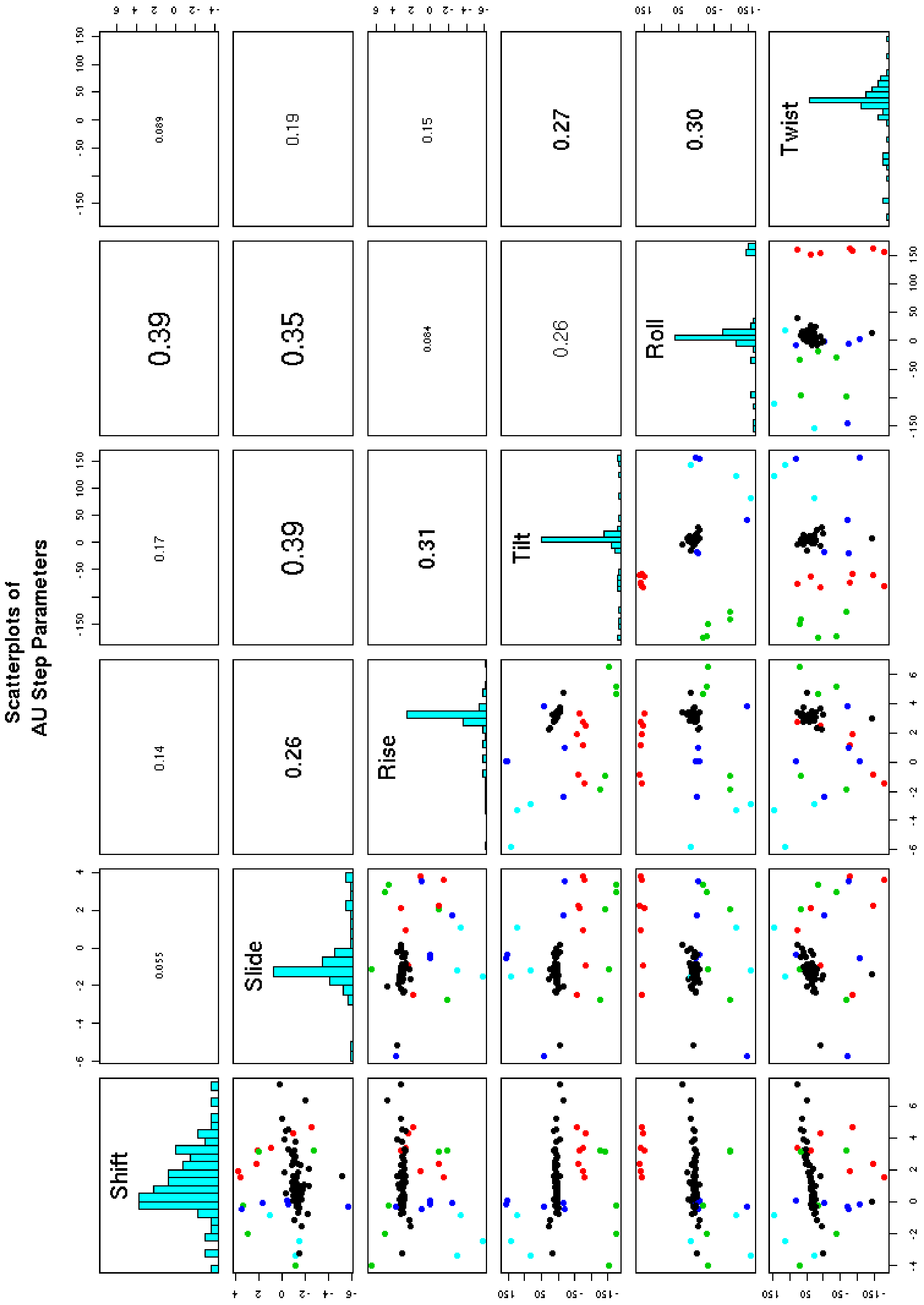

Figure 2.29: Scatterplots for step-parameters of **Helical** AU dinucleotide steps in 50S rRNA.

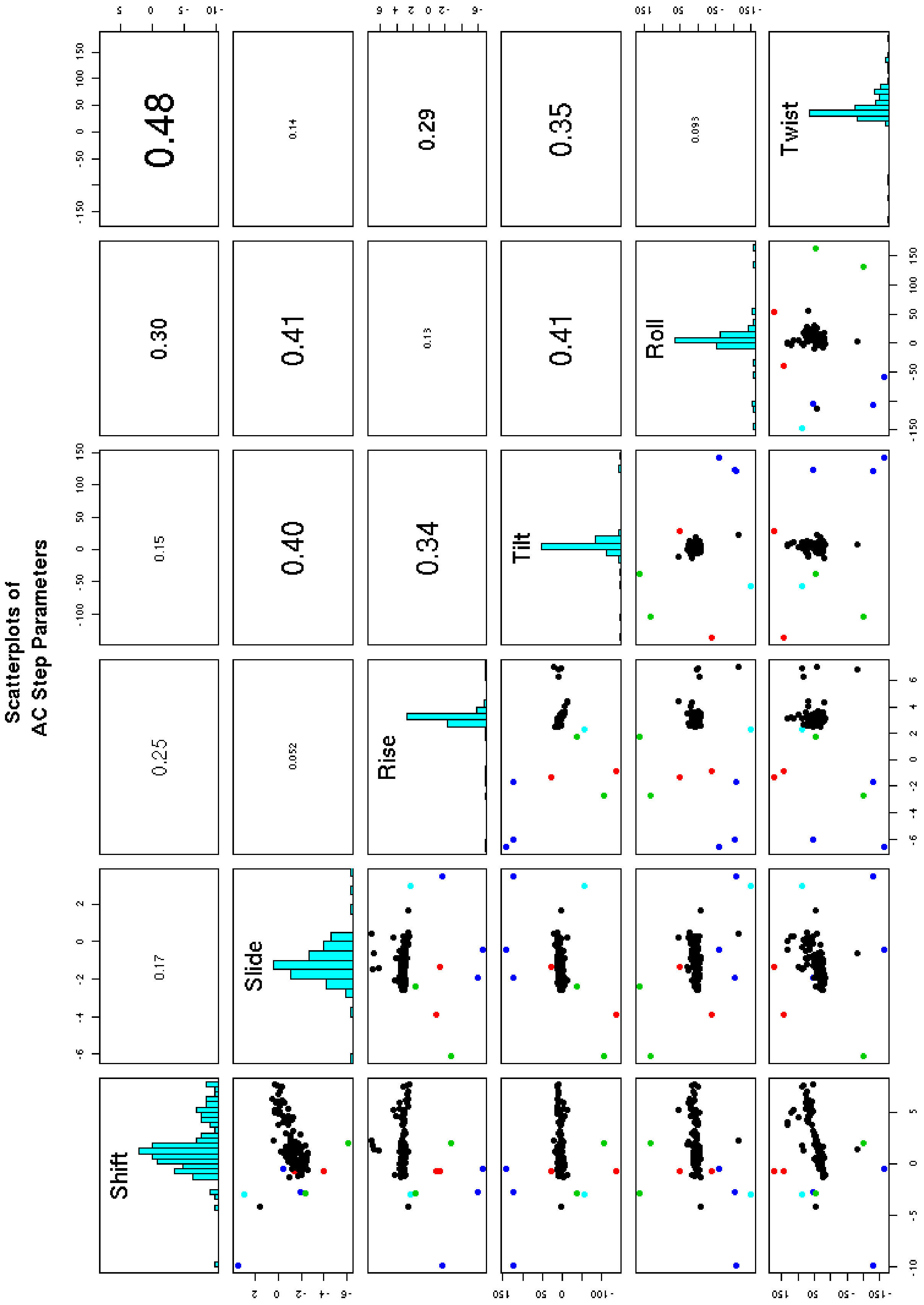

Figure 2.30: Scatterplots for step-parameters of **Helical** AC dinucleotide steps in 50S rRNA.

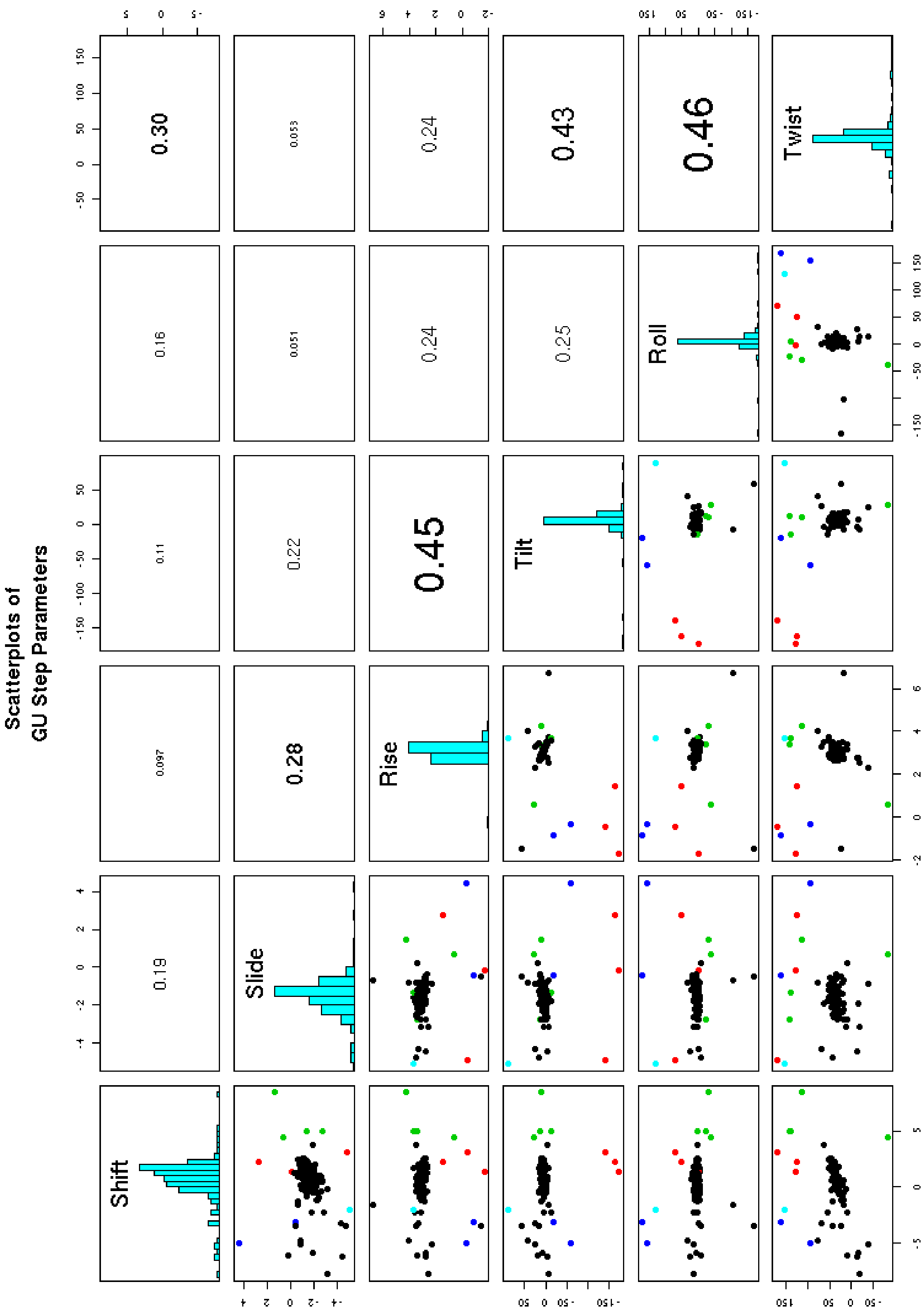

Figure 2.31: Scatterplots for step-parameters of **Helical** GU dinucleotide steps in 50S rRNA.

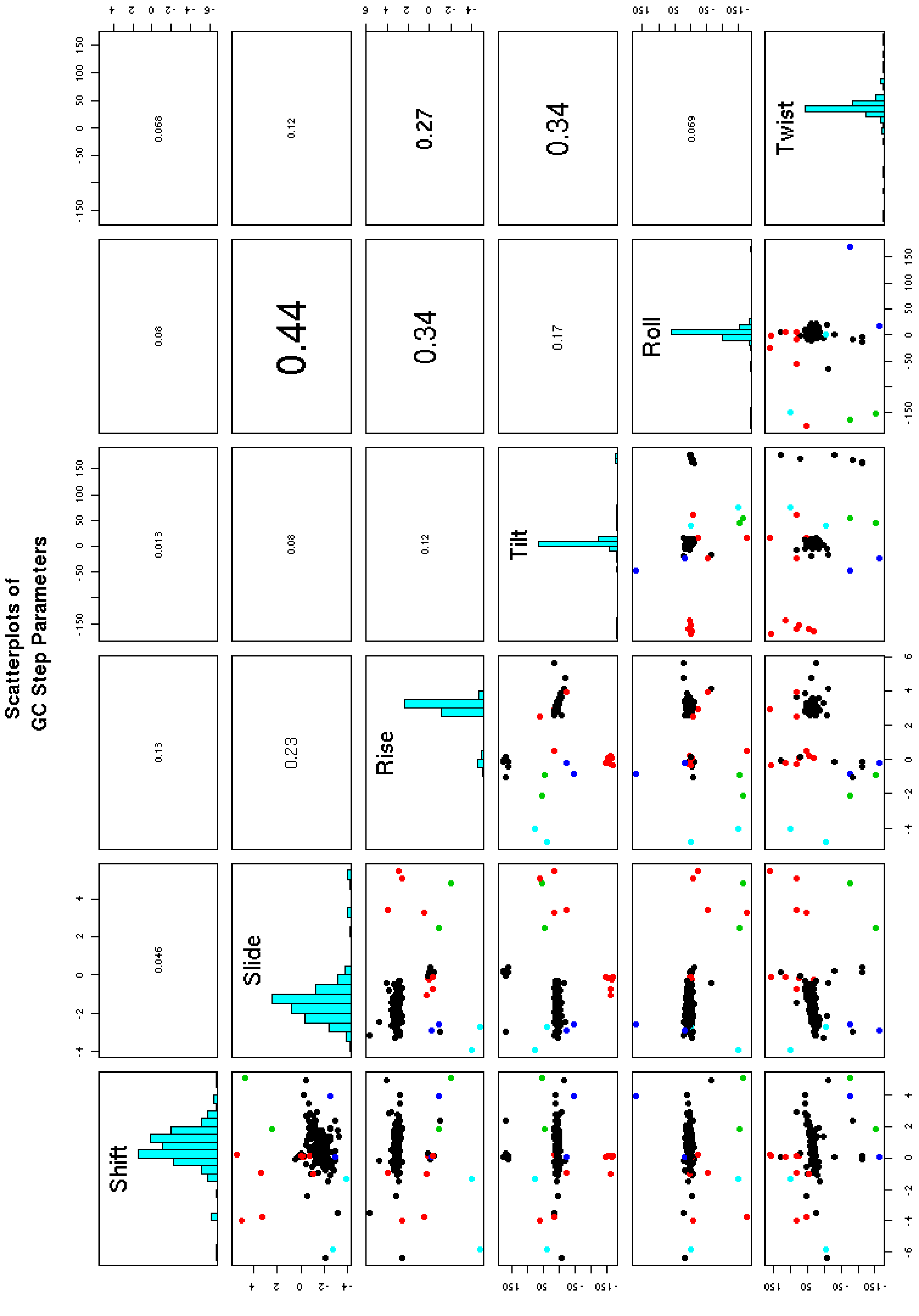

Figure 2.32: Scatterplots for step-parameters of **Helical** GC dinucleotide steps in 50S rRNA.

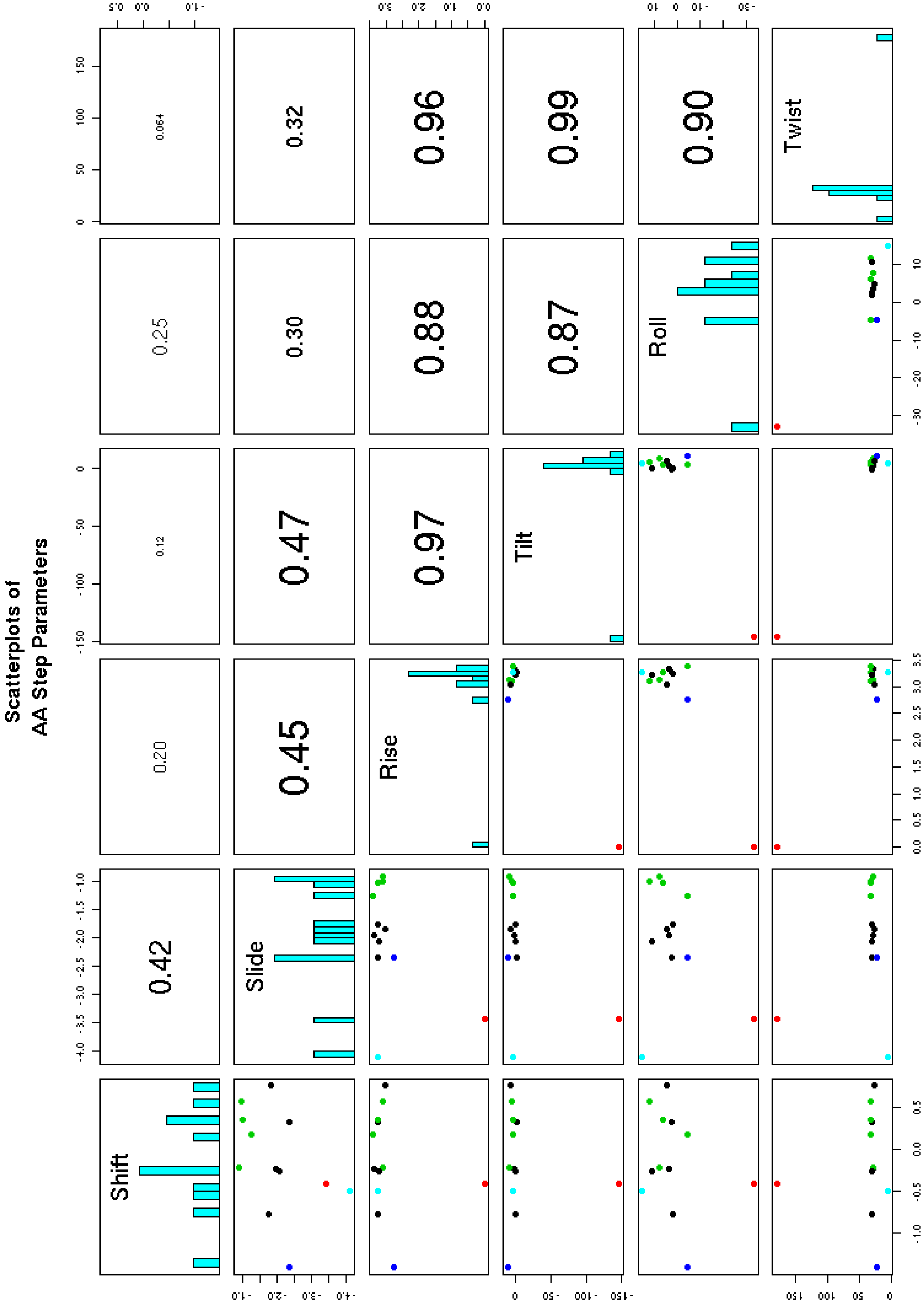

Figure 2.33: Scatterplots for step-parameters of **WC** AA dinucleotide steps in 50S rRNA.

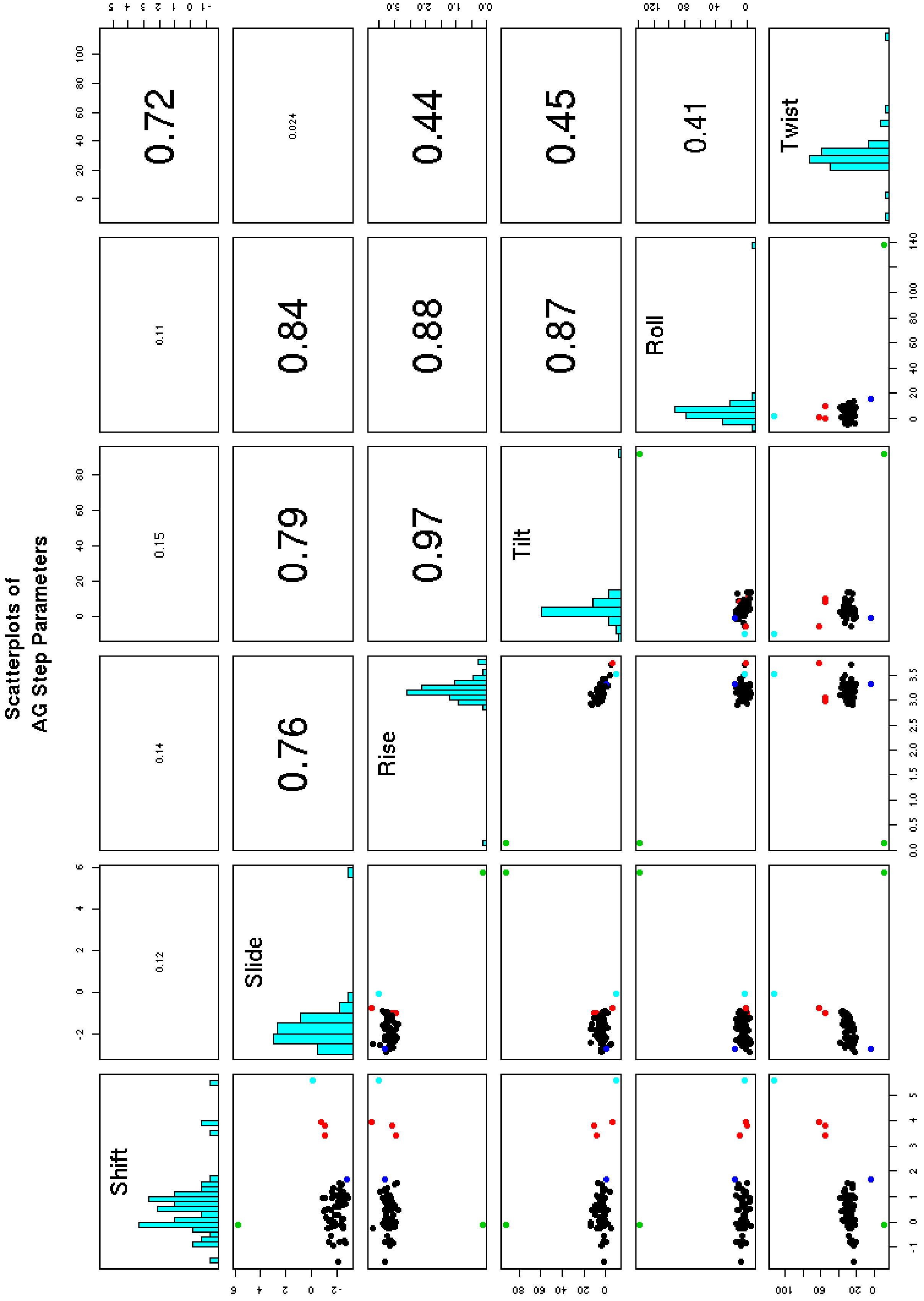

Figure 2.34: Scatterplots for step-parameters of **WC** AG dinucleotide steps in 50S rRNA.

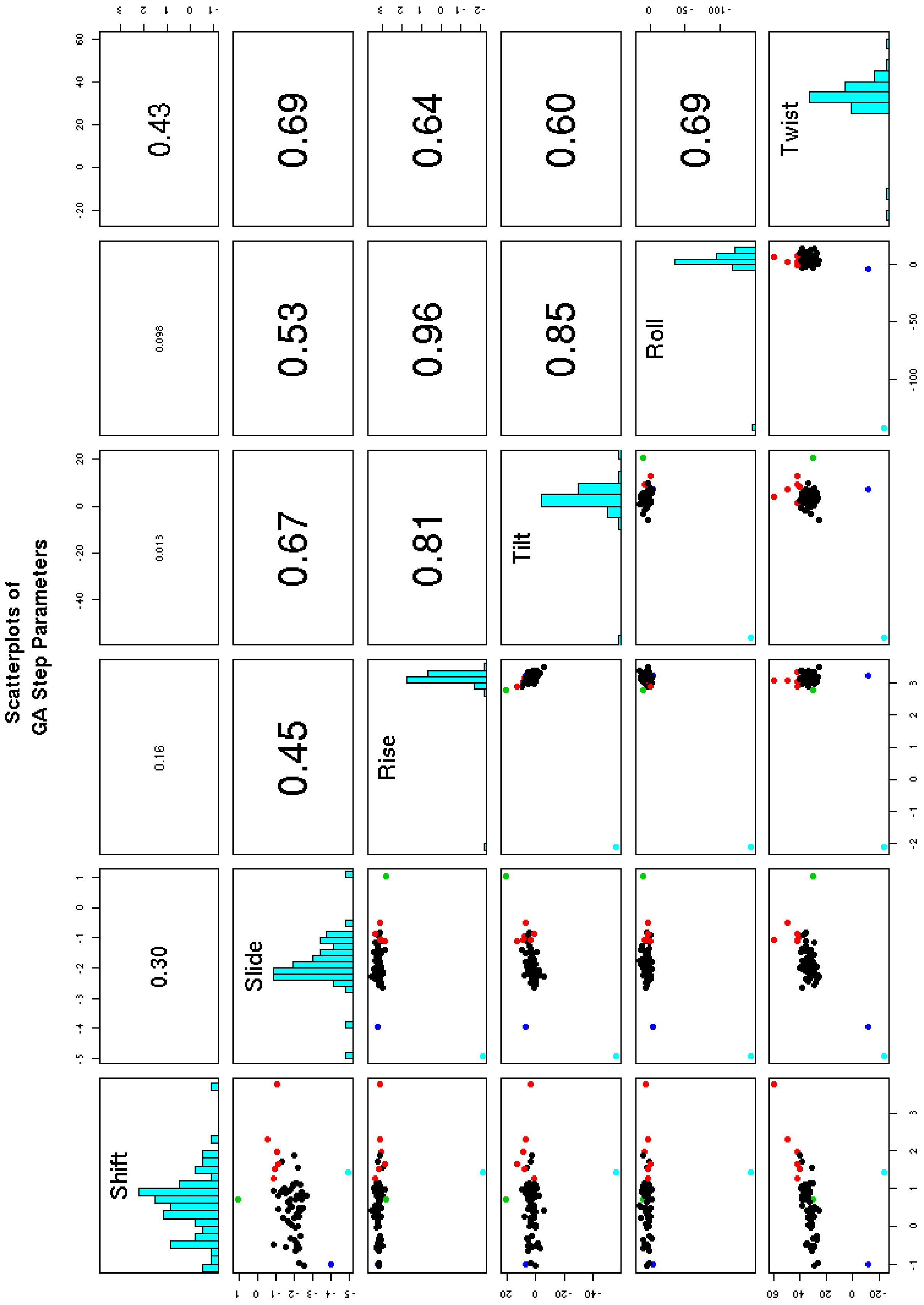

Figure 2.35: Scatterplots for step-parameters of **WC** GA dinucleotide steps in 50S rRNA.

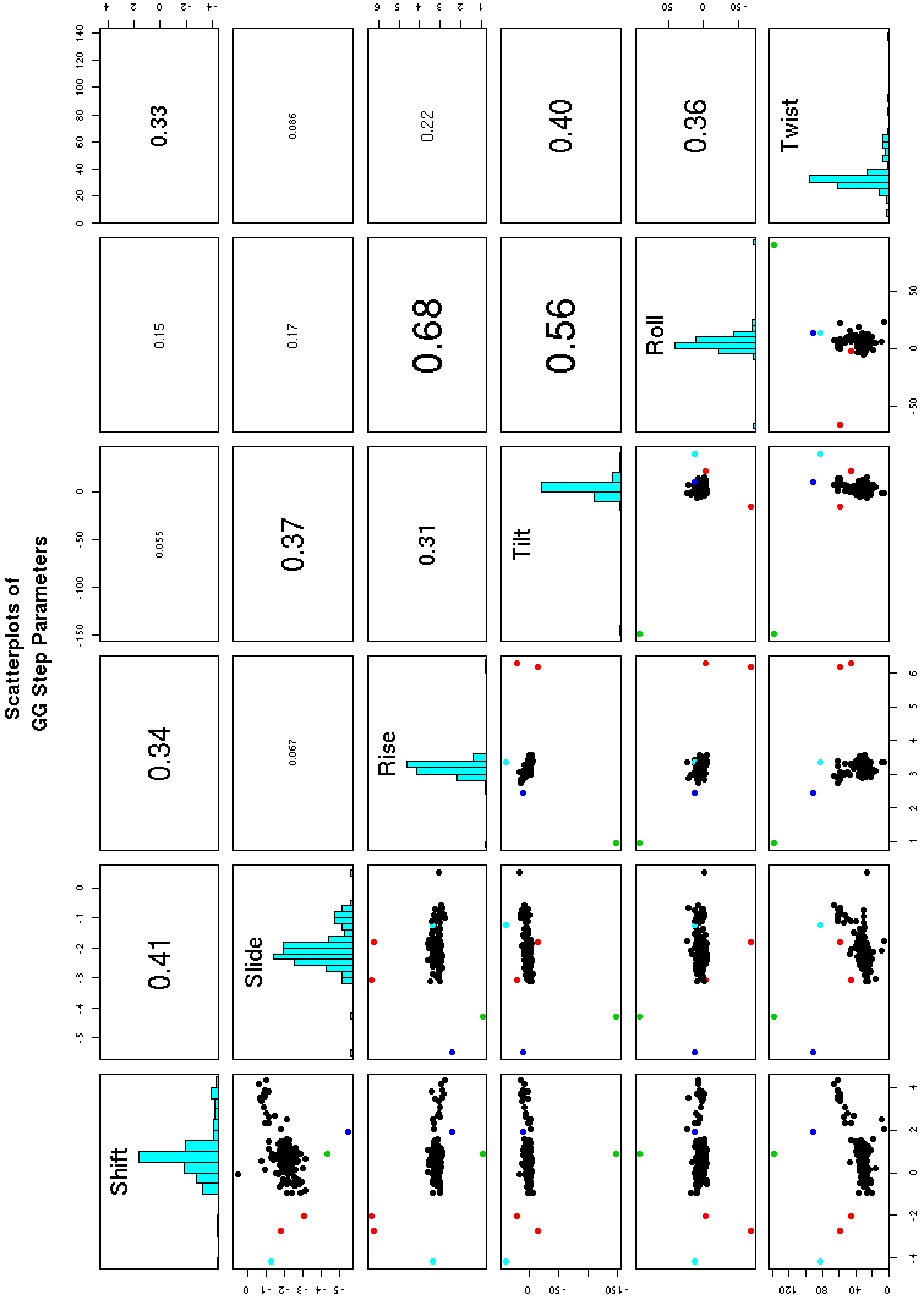

Figure 2.36: Scatterplots for step-parameters of **WC** GG dinucleotide steps in 50S rRNA.

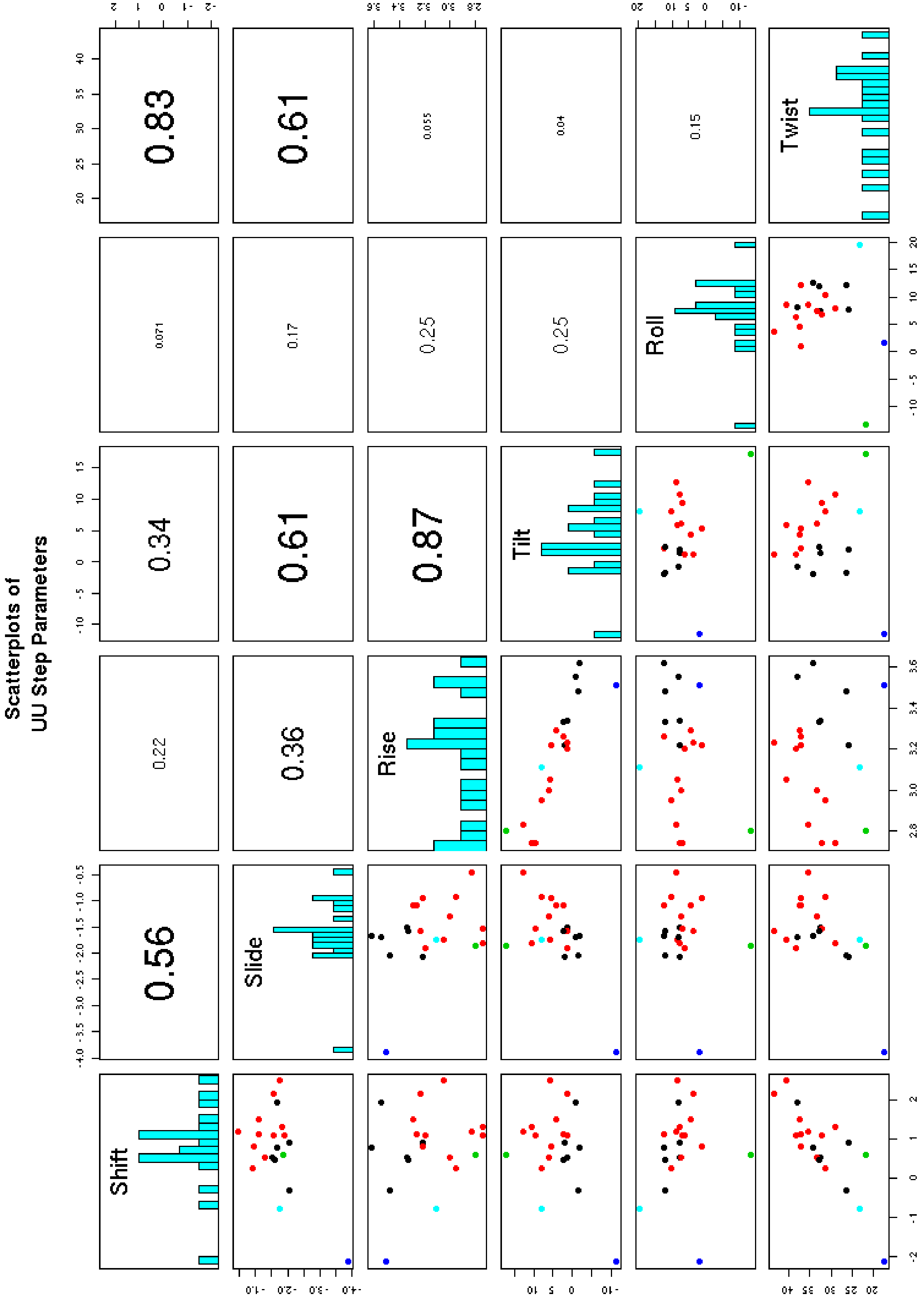

Figure 2.37: Scatterplots for step-parameters of **WC** UU dinucleotide steps in 50S rRNA.

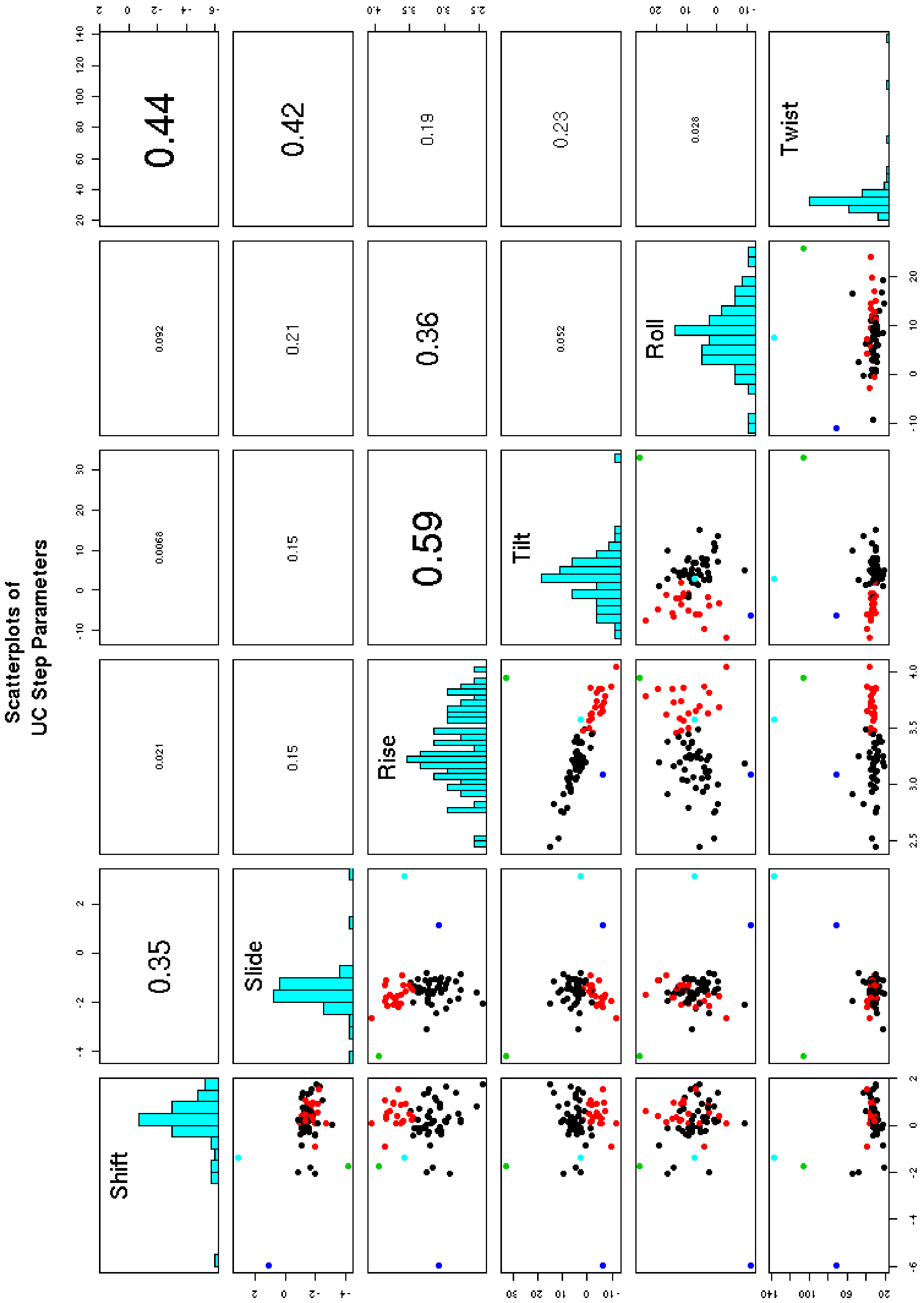

Figure 2.38: Scatterplots for step-parameters of **WC** UC dinucleotide steps in 50S rRNA.

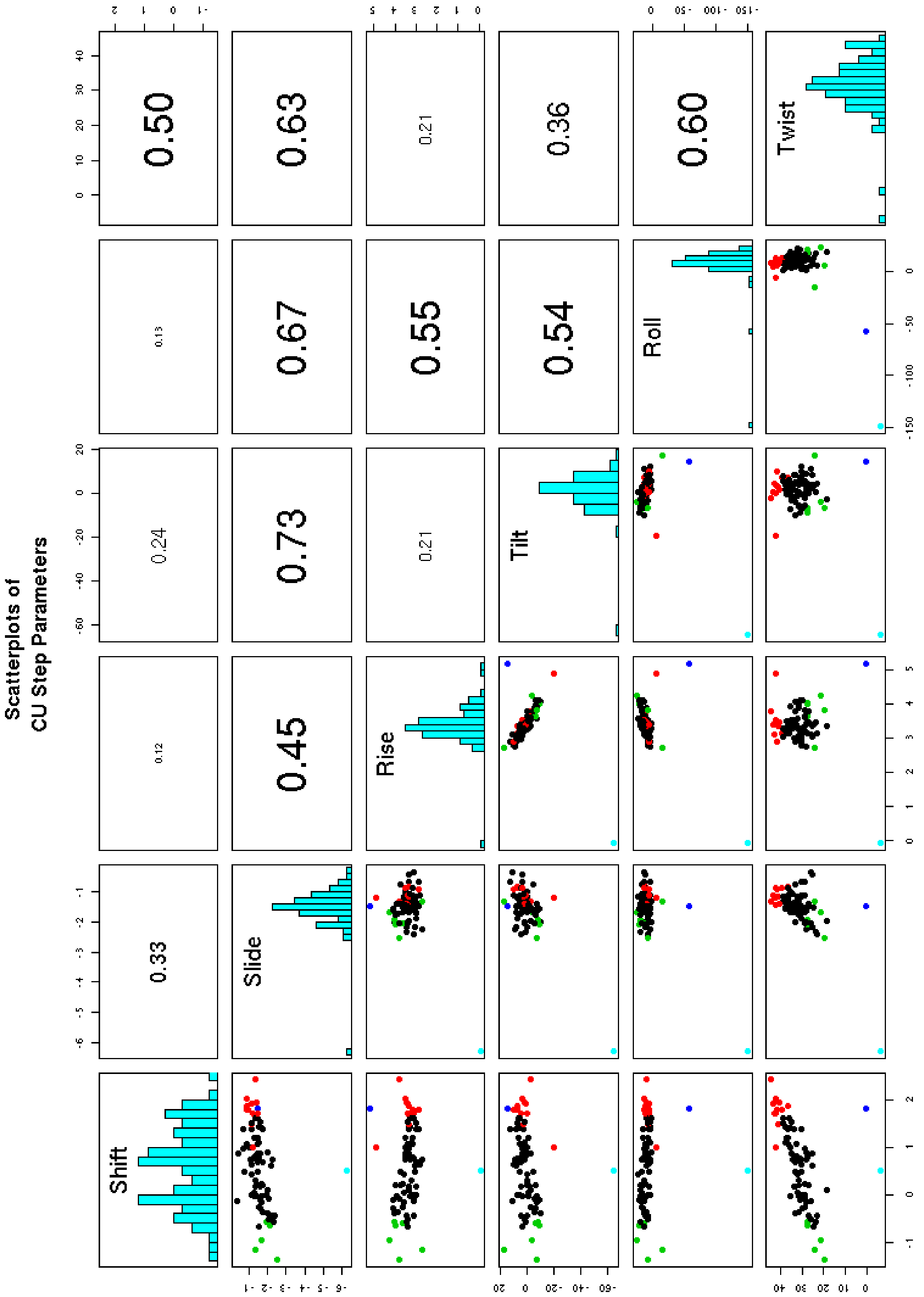

Figure 2.39: Scatterplots for step-parameters of **WC** CU dinucleotide steps in 50S rRNA.

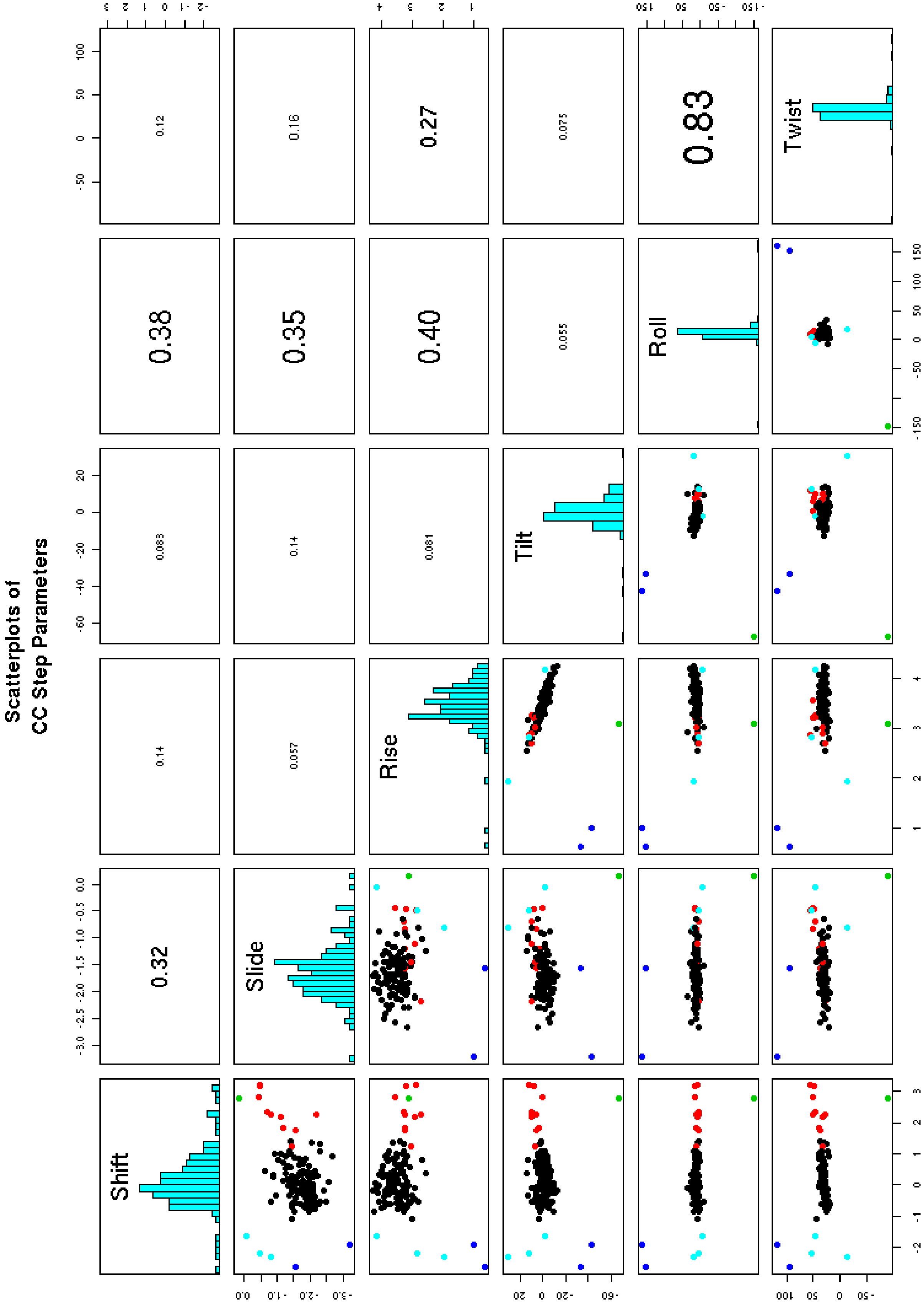

Figure 2.40: Scatterplots for step-parameters of **WC** CC dinucleotide steps in 50S rRNA.

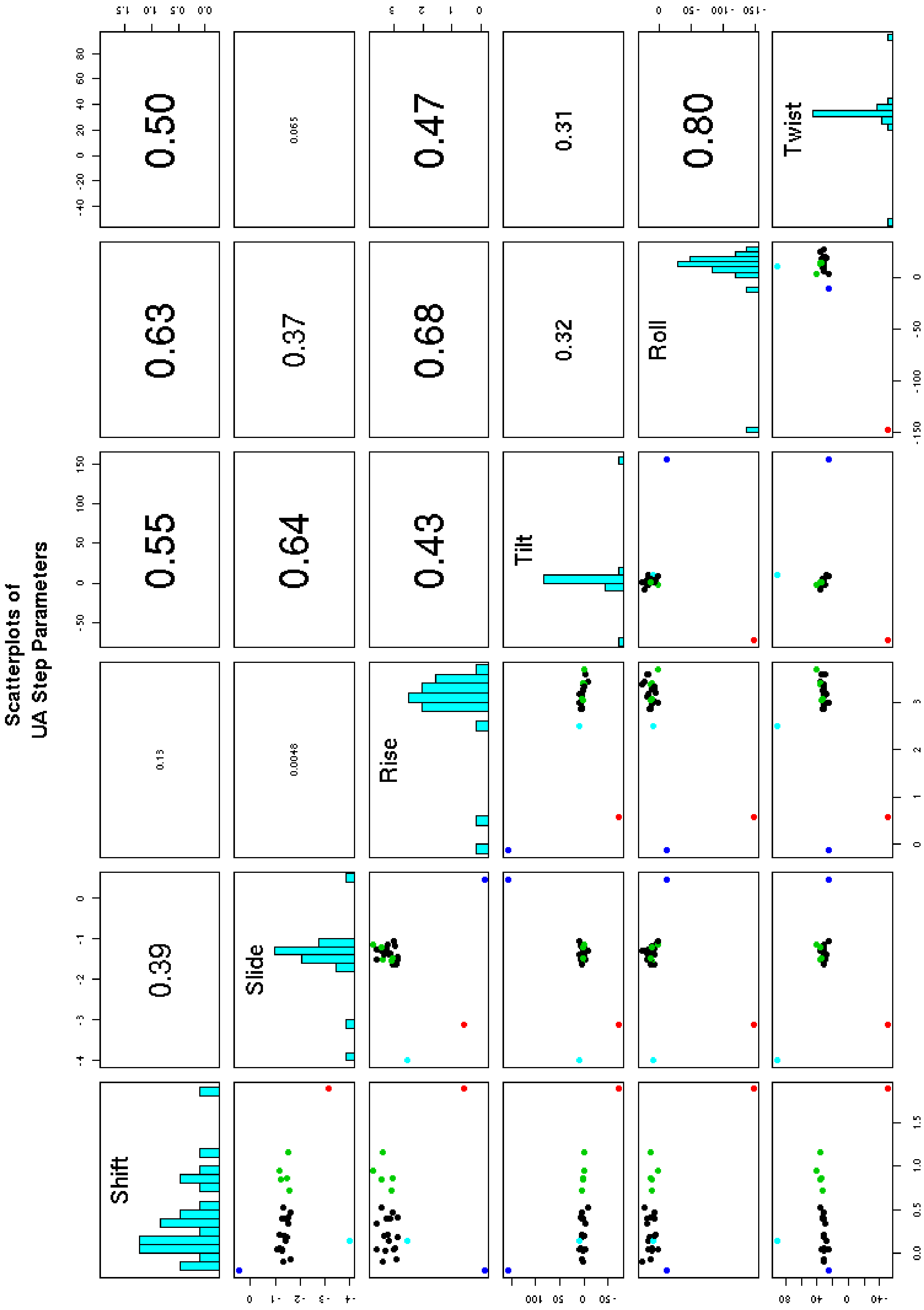

Figure 2.41: Scatterplots for step-parameters of **WC** UA dinucleotide steps in 50S rRNA.

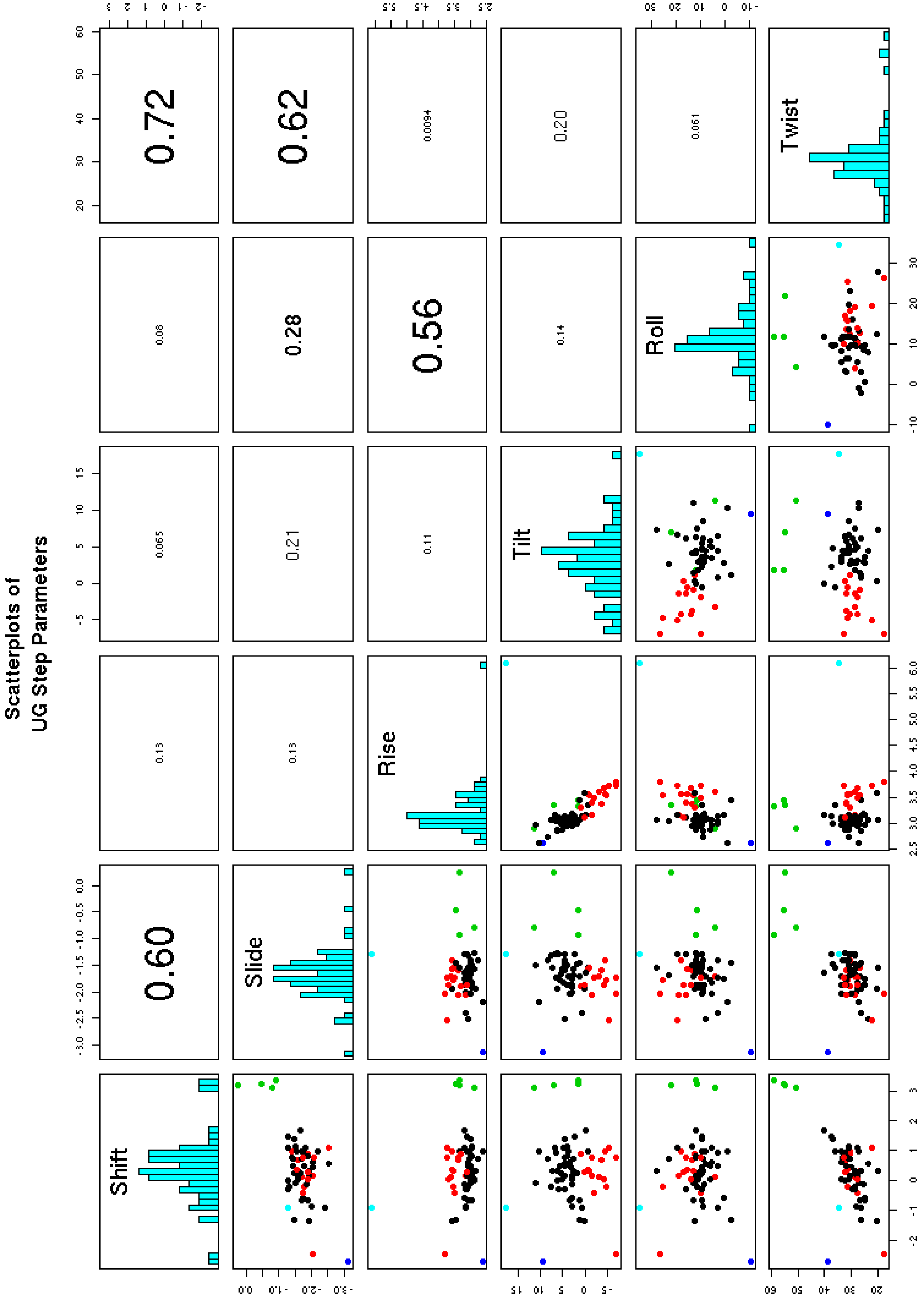

Figure 2.42: Scatterplots for step-parameters of **WC** UG dinucleotide steps in 50S rRNA.

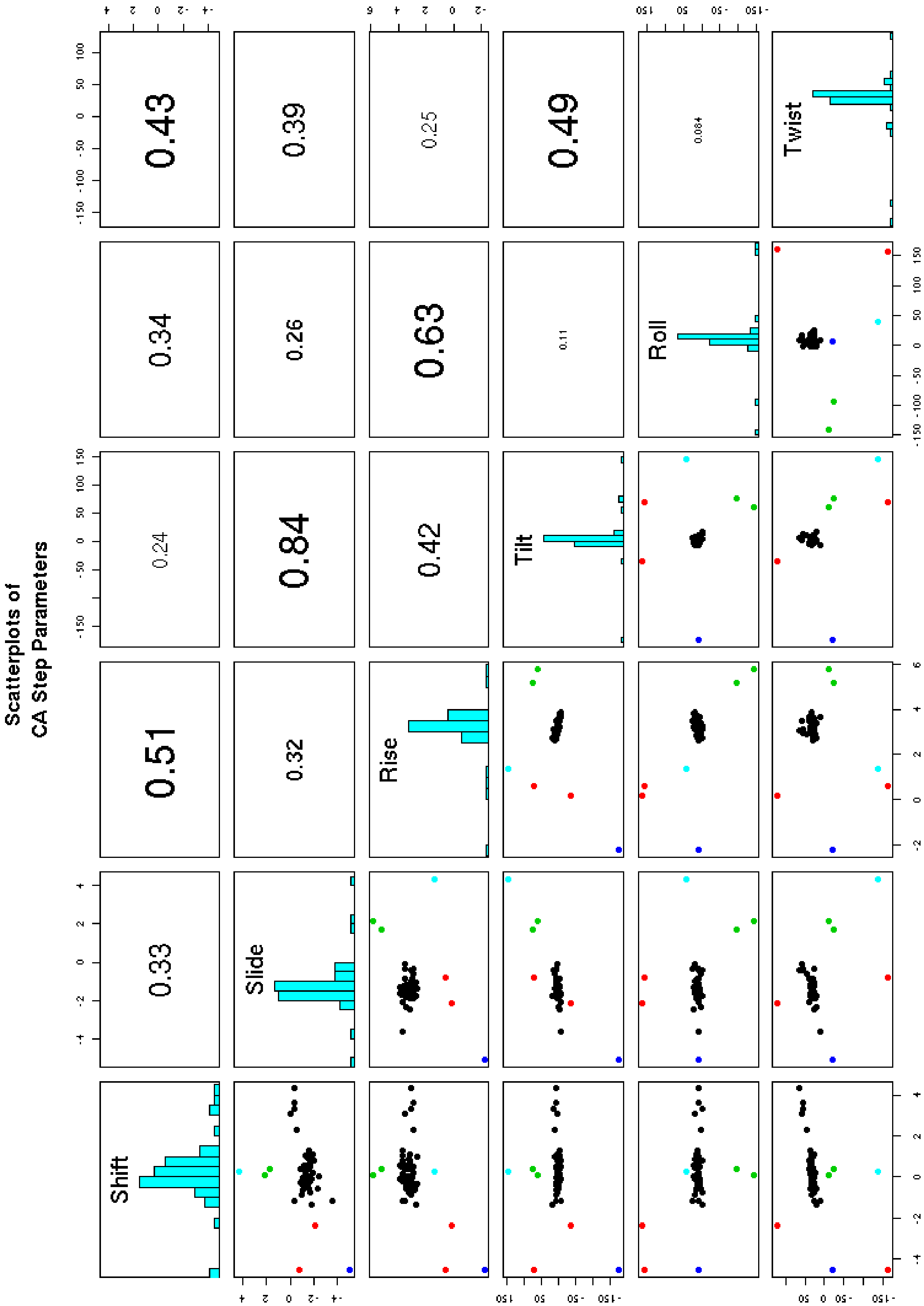

Figure 2.43: Scatterplots for step-parameters of **WC** CA dinucleotide steps in 50S rRNA.

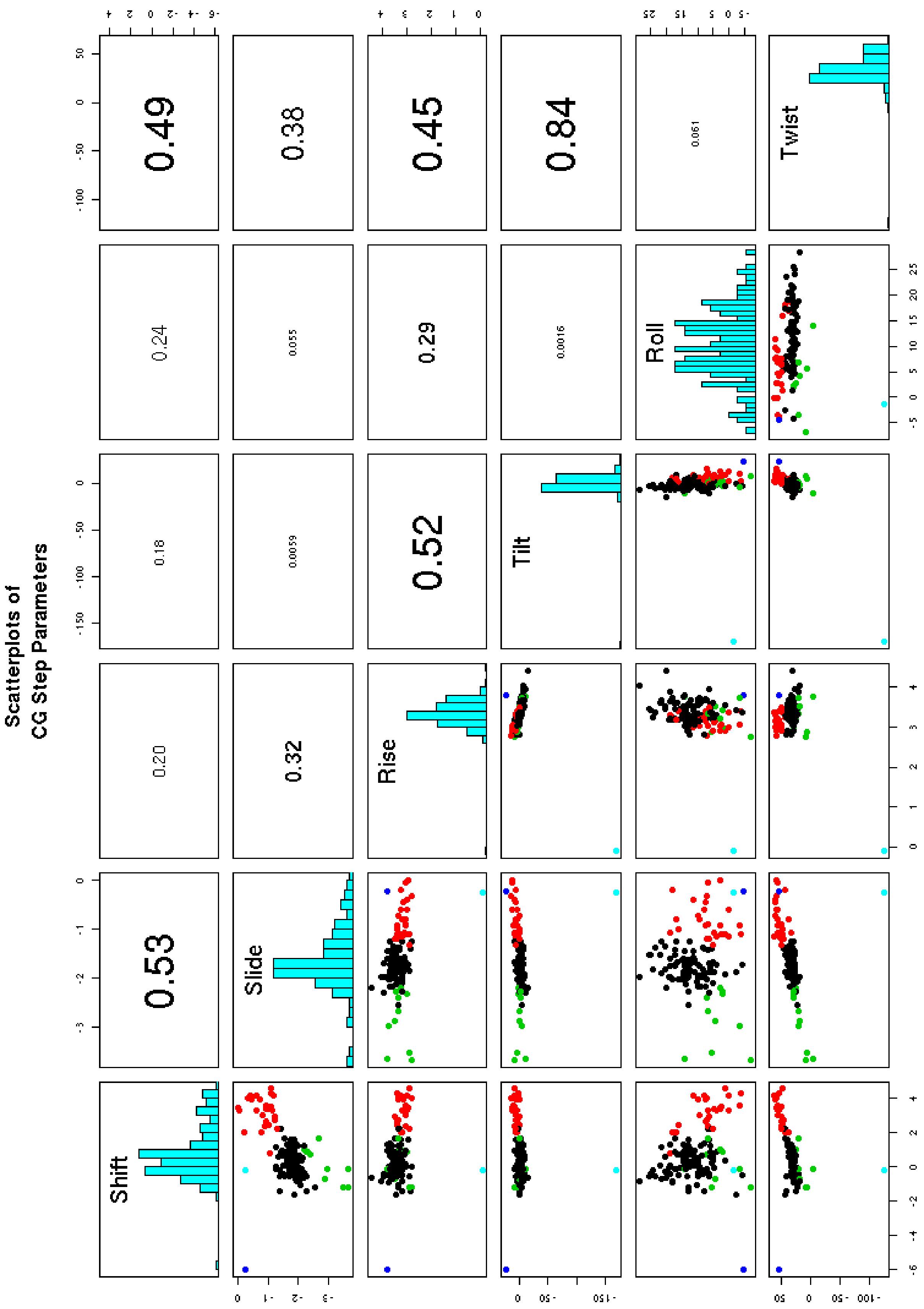

Figure 2.44: Scatterplots for step-parameters of **WC** CG dinucleotide steps in 50S rRNA.

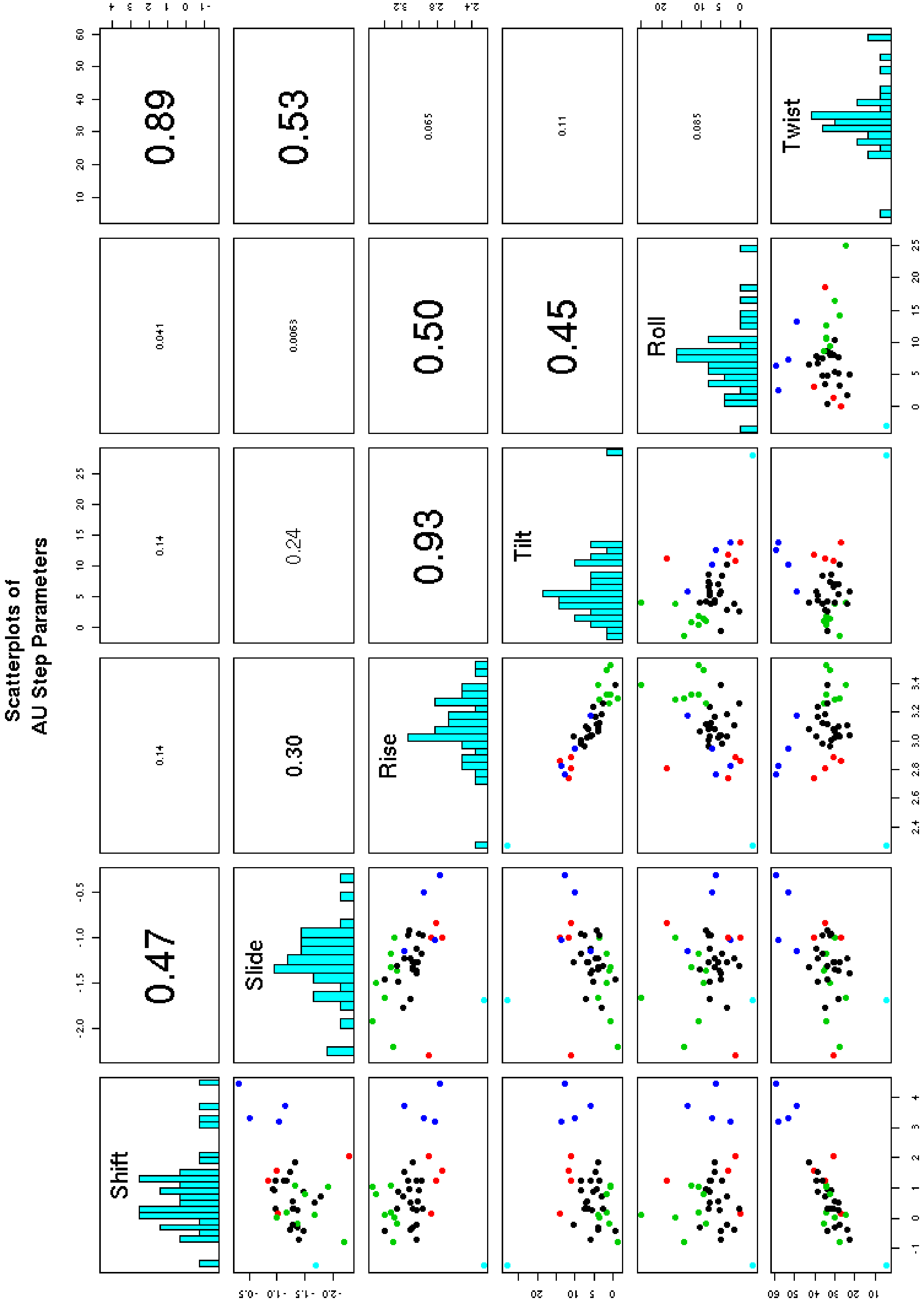

Figure 2.45: Scatterplots for step-parameters of **WC** AU dinucleotide steps in 50S rRNA.

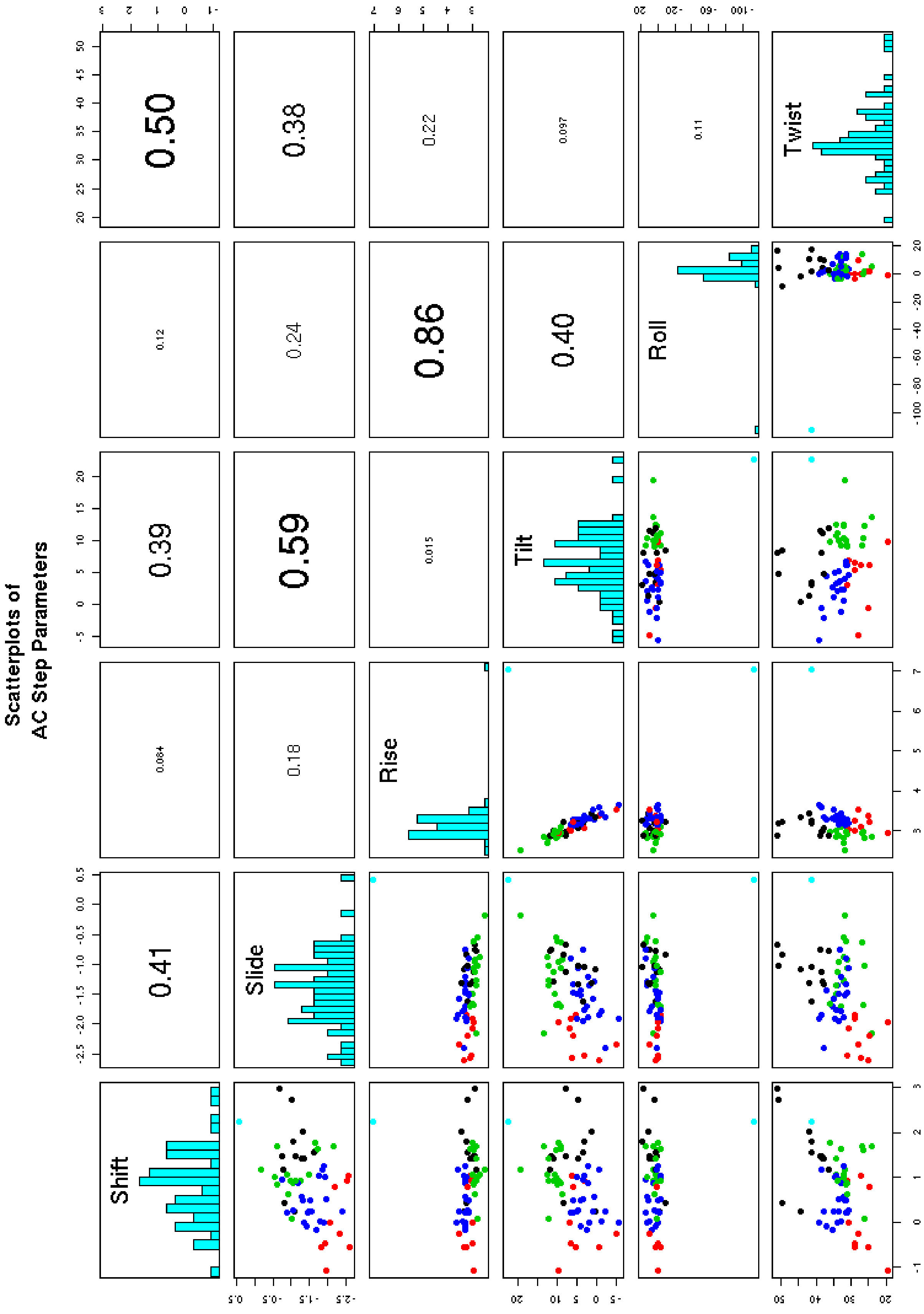

Figure 2.46: Scatterplots for step-parameters of **WC** AC dinucleotide steps in 50S rRNA.

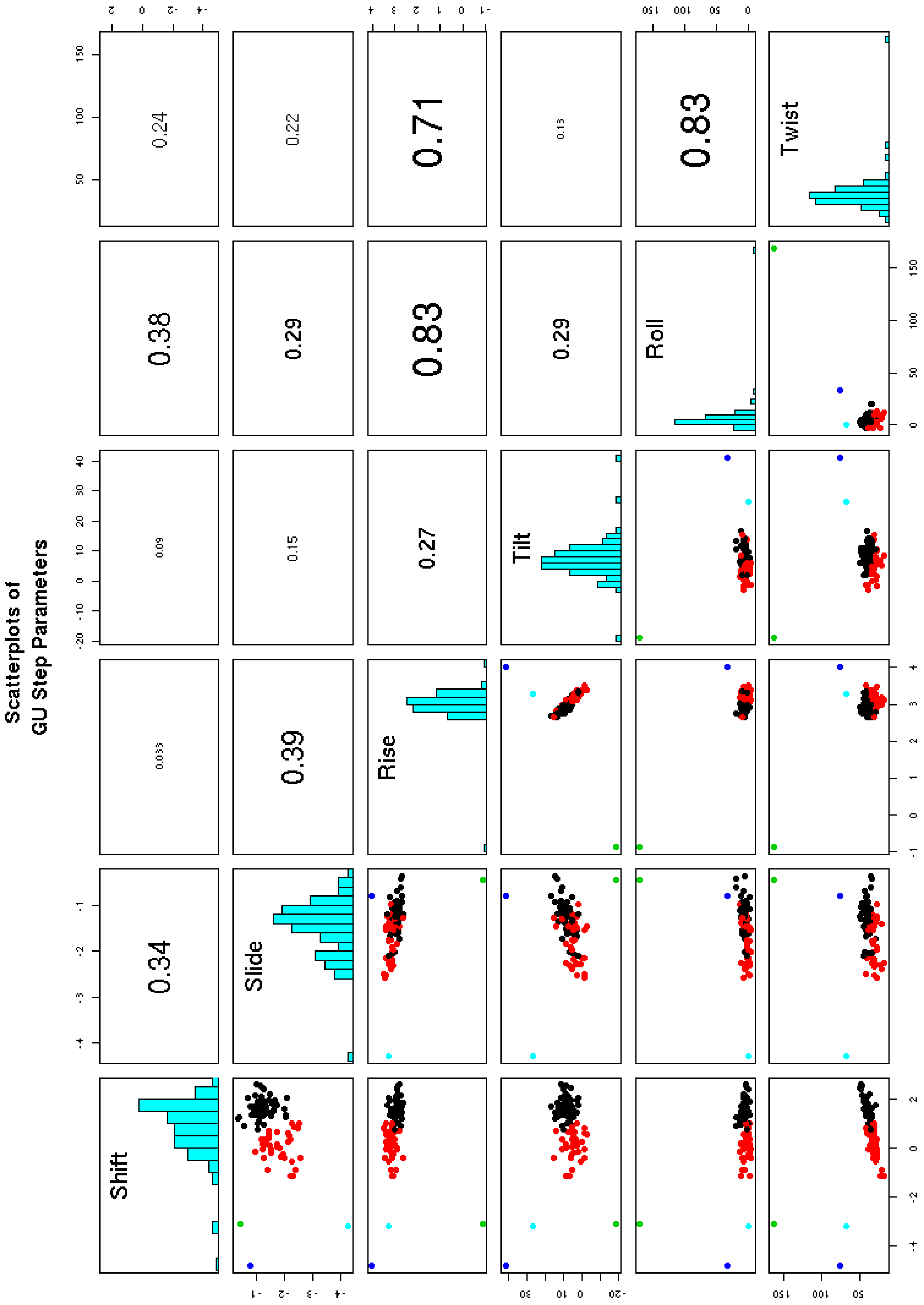

Figure 2.47: Scatterplots for step-parameters of **WC** GU dinucleotide steps in 50S rRNA.

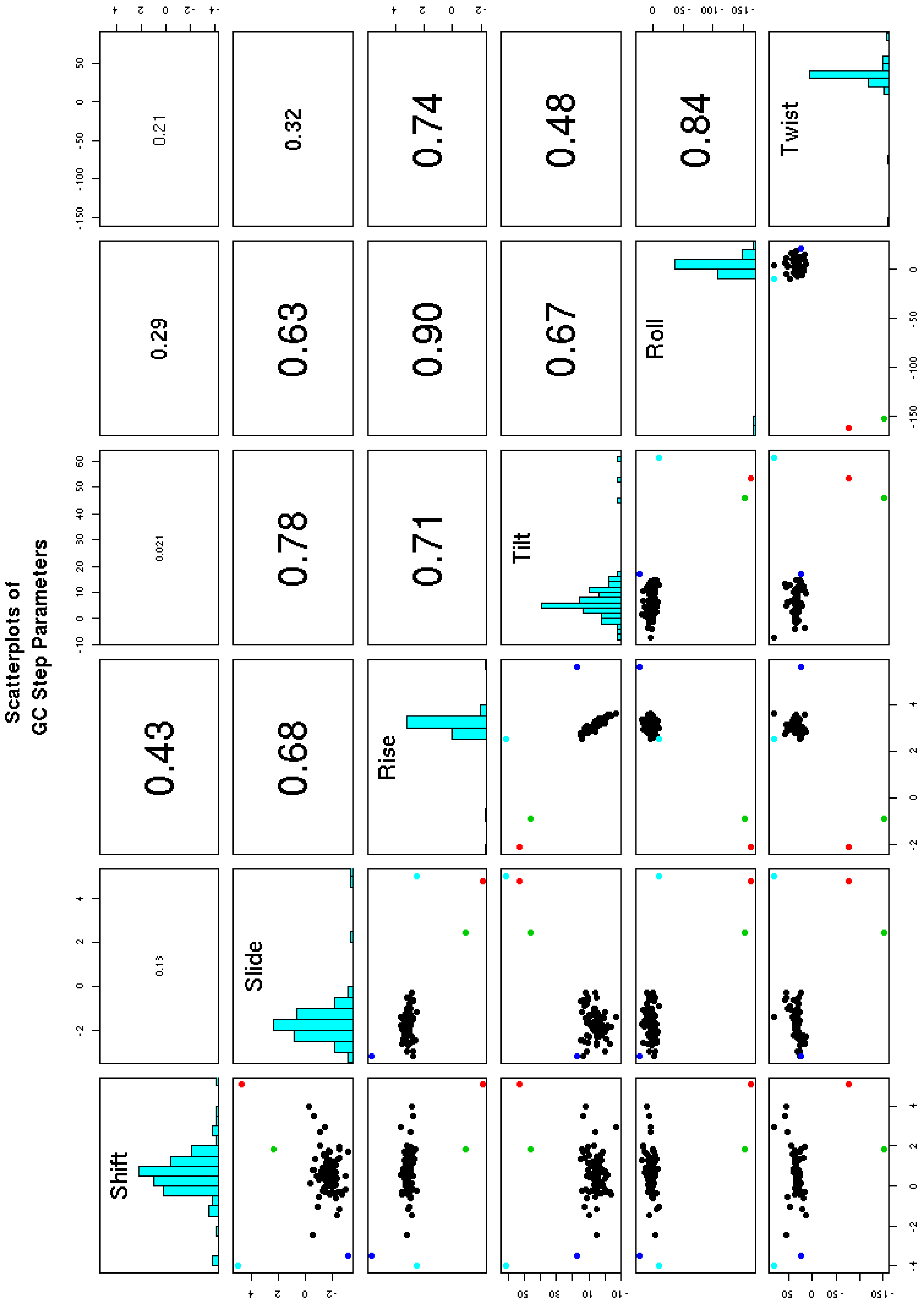

Figure 2.48: Scatterplots for step-parameters of **WC** GC dinucleotide steps in 50S rRNA.

Chapter 3
RNA Dataset

MicroRNAs (miRNA) are small (22 nucleotide) helical RNA's which down-regulate translation in Eukaryotes forming partial duplex structures with mRNA [1]. Motivated by the goal of understanding the structural role non-canonical base pairs play in the regulatory mechanisms in which miRNAs take part, we have assembled a data set of 785 RNA crystal structures taken from the Protein Data Bank (PDB). The collected structures are composed of a minimum of 3 base pairs, have been parsed through 3DNA to detect hydrogen bonds with a distance no longer than 3.4 Å, Stagger ≤ 1.5 Å, Buckle $\leq 30°$, and other 3DNA parameters in their default values.

3.1 Base Pair Counting and Classification

Table 3.1: Distribution of BP (Base Pairs) in RNA Database with 785 Crystal Structures.

Total number of base pairs	131934
Total number of WC base pairs	87735(66.5%)
Total number of non-WC base pairs	44199(33.5%)

In Figure 3.1 we show a scatterplot for the base pair parameters of WC base pairs in red, and non-WC base pairs in blue. From this Figure it's clear that WC base pairs have a narrow distribution for Shear, Stretch and Opening, but for Stagger, Buckle and Propeller, the distribution is almost as broad as that of non-cannonical WC base pairs. Taking this fact into account we can now split our dataset in two parts, a WC one, and a non-canonical one.

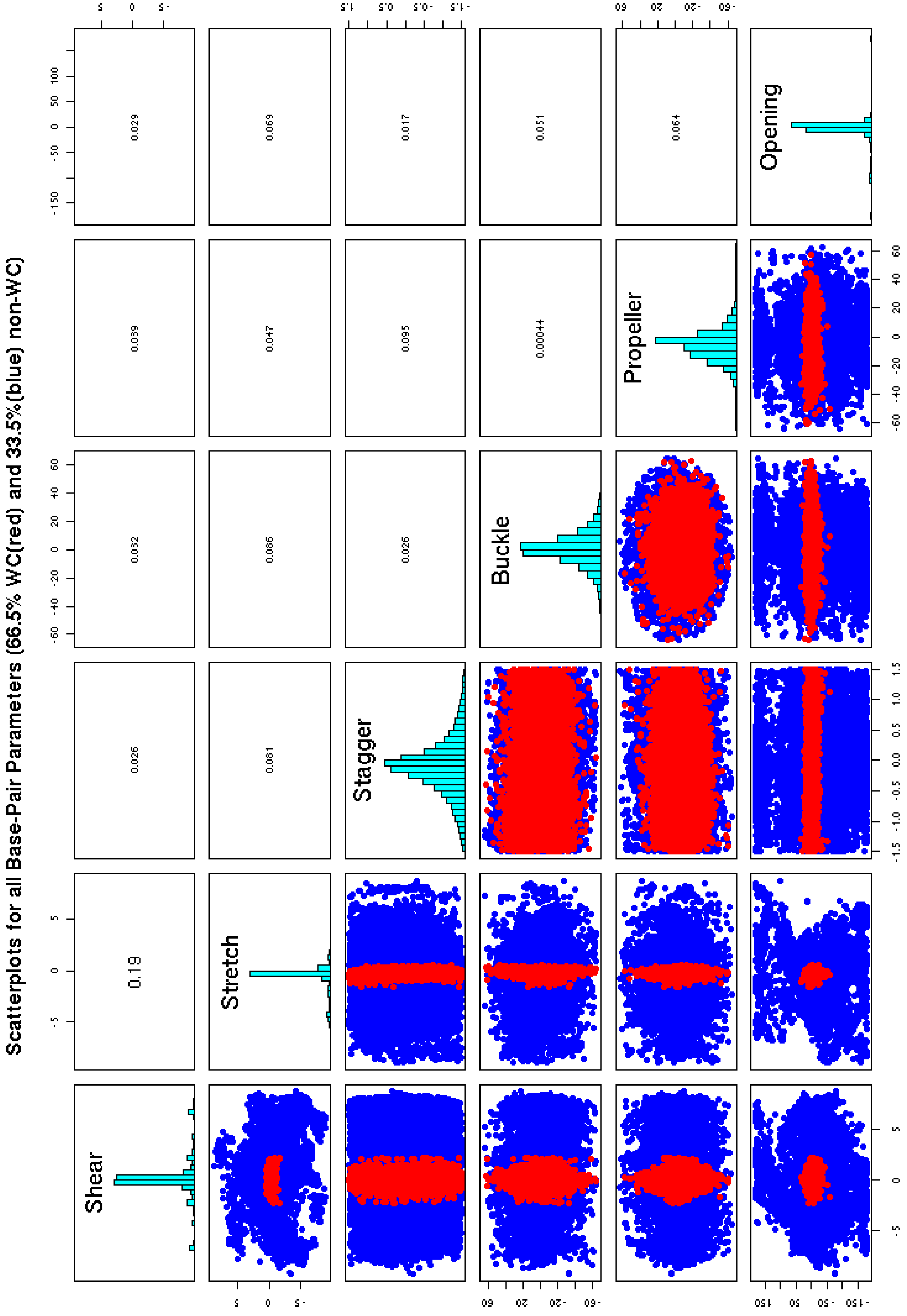

Figure 3.1: Scatterplots showing the distribution of Watson-Crick (WC) and Non-WC base pair in dataset.

3.1.1 All Base Pairs (WC and non-WC)

Table 3.2: Distribution of base pairs according to Leontis-Westhof (LW) classification.

Number of Base Pairs	LW Clasification
105703	cis-W/W
9570	NA
4158	trans-S/H
3101	trans-H/S
2593	trans-W/H
1926	trans-H/W
1050	trans-W/W
691	trans-H/H
677	trans-S/W
566	cis-W/H
431	cis-W/S
371	cis-S/W
333	cis-H/W
195	trans-W/S
178	cis-S/H
132	trans-S/S
117	cis-H/S
82	cis-S/S
23	cis-H/H
11	trans-W/.
5	cis-X/X
5	cis-W/.
5	cis-H/.
3	trans-./W
3	trans-S/.
3	trans-./S
1	trans-./H
1	cis-./.

Table 3.3: Distribution of cis-W/W Base Pairs with LW, HB, and Helical Classifications

# Base Pairs	Base Pair Type	LW Classification	# HB	Helical Classification
21061	C:G	cis-W/W	3	H/H
19828	G:C	cis-W/W	3	H/H
8165	A:U	cis-W/W	2	H/H
7939	U:A	cis-W/W	2	H/H
4028	C:G	cis-W/W	3	H/H_qe
3080	G:C	cis-W/W	3	H_qe/H

Continued on Next Page...

Table 3.3 – Continued

# Base Pairs	Base Pair Type	LW Classification	# HB	Helical Classification
3013	G:C	cis-W/W	3	H/H_qe
2204	G:U	cis-W/W	2	H/H
2164	U:G	cis-W/W	2	H/H
2055	C:G	cis-W/W	3	H_qe/H
1998	C:G	cis-W/W	3	H_e/H_e
1701	G:C	cis-W/W	3	H_e/H_e
1535	C:G	cis-W/W	3	H_qe/H_qe
1250	G:C	cis-W/W	3	H_qe/H_qe
1200	G:U	cis-W/W	3	H/H
1086	A:U	cis-W/W	2	H/H_qe
1065	U:G	cis-W/W	3	H/H
1000	U:A	cis-W/W	2	H/H_qe
865	C:G	cis-W/W	2	H/H
795	G:C	cis-W/W	2	H/H
709	A:U	cis-W/W	2	H_qe/H
681	U:A	cis-W/W	2	H_qe/H
588	U:A	cis-W/W	2	H_qe/H_qe
534	U:G	cis-W/W	2	H_qe/H
501	U:U	cis-W/W	2	H/H
490	A:U	cis-W/W	2	H_qe/H_qe
456	G:U	cis-W/W	2	H/H_qe
407	C:G	cis-W/W	3	H_e/H_ie
386	G:C	cis-W/W	4	H/H
362	G:U	cis-W/W	2	H_qe/H
362	C:G	cis-W/W	4	H/H
354	U:A	cis-W/W	2	H_e/H_e
331	U:A	cis-W/W	3	H/H
315	U:G	cis-W/W	2	H_e/H_e
305	A:U	cis-W/W	3	H/H
305	A:U	cis-W/W	2	H_e/H_e
302	U:G	cis-W/W	2	H/H_qe
293	U:A	cis-W/W	1	H/H
290	A:U	cis-W/W	1	H/H
229	A:G	cis-W/W	2	H/H_qe
207	G:C	cis-W/W	3	H_qe/H_i
205	G:U	cis-W/W	3	H_qe/H
202	G:C	cis-W/W	1	H/H
188	G:U	cis-W/W	3	H/H_qe
186	U:G	cis-W/W	3	H/H_qe
179	C:C	cis-W/W	2	H/H

3.1.2 Watson-Crick Base Pairs

In addition to Figure 3.1 we have also colored the most abundant WC base pairs in Figure 3.2 that is, CG, GC, AU, and UA, base pairs. The base pairs counts for every possible Watson-Crick pair can be seen in Table 3.4, and the average base pair parameters and their standard deviations can be seen in Table 3.5

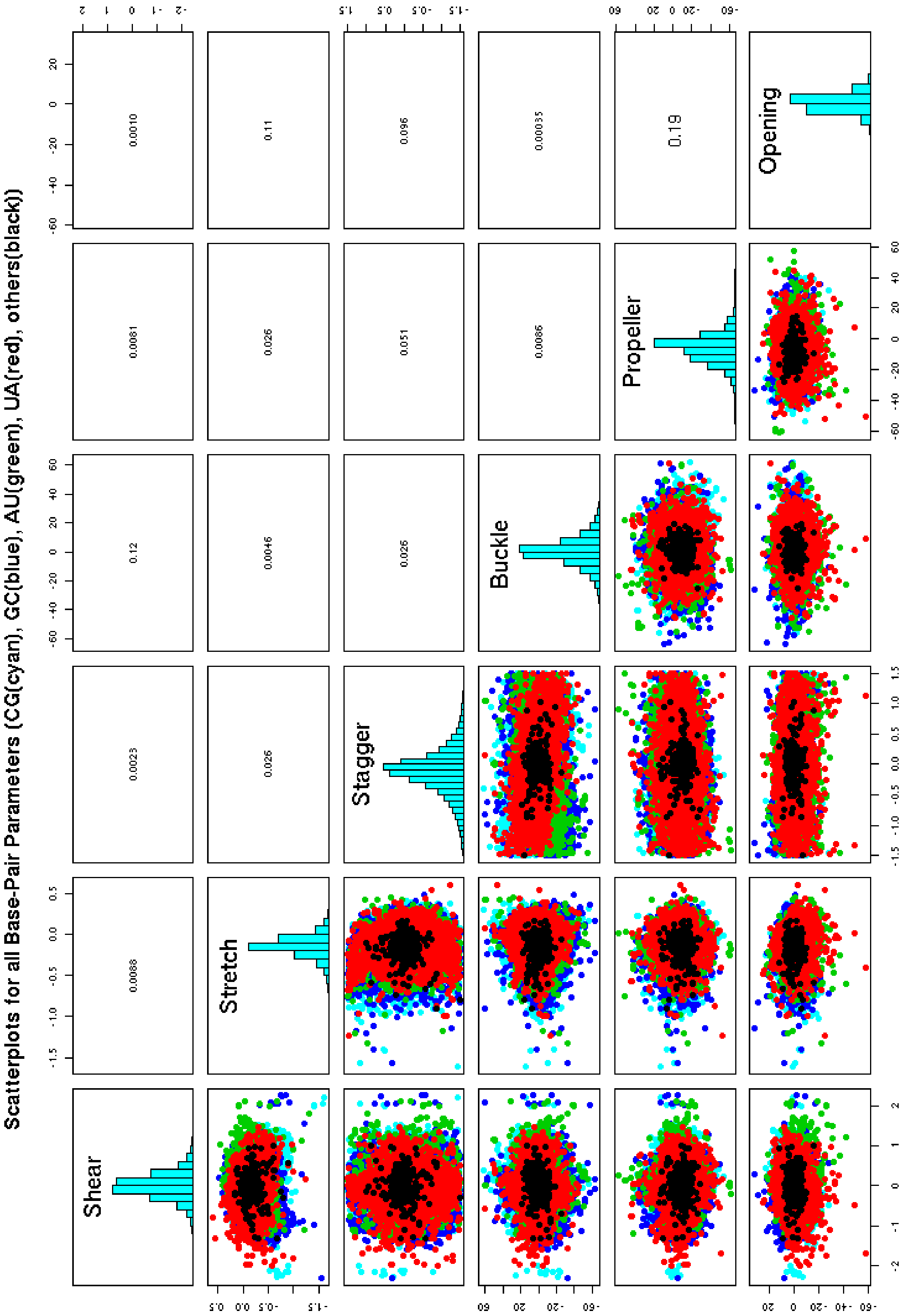

Figure 3.2: Scatterplots showing the distribution of Watson-Crick (WC) base pairs colored as CG(Cyan), GC(Green), AU(Blue), UA(Red), others(black)

Table 3.4: Number of WC base pairs sorted by base pair ID.

Number of Base Pairs	Base Pair Type
33036	C:G
31216	G:C
11756	A:U
11424	U:A
61	5MC:G
35	5BU:A
24	2MG:C
20	A:5BU
16	G48:C43
12	C43:G48
11	OMG:OMC
9	OMC:OMG
8	U36:A44
8	DC:G
8	A44:U36
7	LC:LG
7	G:CBV
6	G:5MC
6	CBV:G
5	CBR:G
4	UMS:A
4	LG:LC
4	GTP:C
4	G:CBR
4	CSL:G
4	52:22
3	IU:A
3	C:2MG
2	G:5IC
2	C:GTP
2	A:IU
2	5CG:C
1	U:RIA
1	SSU:A
1	S4C:G
1	OMU:A2M
1	N5C:G
1	G:S4C
1	G:CSL
1	G:CB2
1	G:A5M
1	G:74
1	DG:C
1	DA:U

Continued on Next Page. . .

Table 3.4 – Continued

Number of Base Pairs	Base Pair Type
1	C:XUG
1	CTP:G
1	C:DG
1	CB2:G
1	C:7MG
1	ADE:74
1	A5M:G
1	A2M:OMU
1	47:56
1	1SC:G

Table 3.5: This table summarizes for Watson-Crick base pairs, the base pair parameters average values with standard deviations in parentheses.

Base Pair Parameter	CG	GC	AU	UA	others	ARNA
Shear	0.169(0.311)	-0.184(0.315)	0.035(0.276)	-0.034(0.292)	-0.070(0.381)	0.01
Stretch	-0.149(0.135)	-0.148(0.132)	-0.122(0.118)	-0.121(0.120)	-0.175(0.151)	-0.08
Stagger	-0.145(0.427)	-0.138(0.407)	-0.042(0.394)	-0.046(0.391)	0.001(0.285)	0.01
Buckle	3.888(9.589)	-3.631(9.446)	-0.817(8.750)	1.209(8.700)	0.688(6.590)	-0.00
Propeller	-6.568(8.841)	-6.309(8.721)	-4.724(10.537)	-6.350(8.672)	-9.678(7.052)	-2.07
Opening	0.454(3.600)	0.509(3.643)	0.709(5.349)	1.043(5.248)	3.554(0.313)	-1.67

As can be seen in Table 3.6 all CG, GC, AU, and UA cannonical base pairs are given the cis-W/W LW classification, or are classified as NA as is expected.

Table 3.6: Canonical WC Base-Pairs in dataset.

Number of Base Pairs	Base Pair Type	LW Classification
31680	C:G	cis-W/W
29908	G:C	cis-W/W
11435	A:U	cis-W/W
11089	U:A	cis-W/W
1356	C:G	NA
1308	G:C	NA
335	U:A	NA
321	A:U	NA

Number of Base Pairs	Base Pair Type	Helical Classification	LW Classification

21077	C:G	H/H	cis-W/W
4034	C:G	H/H_qe	cis-W/W
2056	C:G	H_qe/H	cis-W/W
2012	C:G	H_e/H_e	cis-W/W
1551	C:G	H_qe/H_qe	cis-W/W
415	C:G	H_e/H_ie	cis-W/W
151	C:G	H/H_i	cis-W/W
124	C:G	H_qe/H_i	cis-W/W
100	C:G	H_i/H_i	cis-W/W
75	C:G	H_ie/H_ie	cis-W/W
51	C:G	H_ie/H_e	cis-W/W
27	C:G	H_i/H_qe	cis-W/W
7	C:G	H_i/H	cis-W/W
560	C:G	H/H	NA
232	C:G	H_e/H_e	NA
173	C:G	H/H_qe	NA
142	C:G	H_qe/H_qe	NA
132	C:G	H_qe/H	NA
39	C:G	H_e/H_ie	NA
24	C:G	H_ie/H_ie	NA
23	C:G	H_i/H_i	NA
12	C:G	H_ie/H_e	NA
10	C:G	H_qe/H_i	NA
5	C:G	H/H_i	NA
4	C:G	H_i/H_qe	NA

Table 3.8: Distribution of base pairs according to LW Classification and WC Classification

Number of Base Pairs	**LW Classification**	**WC or not**
84410	cis-W/W	WC
21293	cis-W/W	non-WC
6245	NA	non-WC
4158	trans-S/H	non-WC
3325	NA	WC
3101	trans-H/S	non-WC
2593	trans-W/H	non-WC
1926	trans-H/W	non-WC
1050	trans-W/W	non-WC
691	trans-H/H	non-WC
677	trans-S/W	non-WC
566	cis-W/H	non-WC
431	cis-W/S	non-WC
371	cis-S/W	non-WC
333	cis-H/W	non-WC

195	trans-W/S	non-WC
178	cis-S/H	non-WC
132	trans-S/S	non-WC
117	cis-H/S	non-WC
82	cis-S/S	non-WC
23	cis-H/H	non-WC
11	trans-W/.	non-WC
5	cis-X/X	non-WC
5	cis-W/.	non-WC
5	cis-H/.	non-WC
3	trans-./W	non-WC
3	trans-S/.	non-WC
3	trans-./S	non-WC
1	trans-./H	non-WC
1	cis-./.	non-WC

3.1.3 non-WC Base Pairs

Table 3.9: non-WC base pairs classified by type, helical region and LW.

Base Pair Type	Helical Region Type	LW Classification	Counts
G:U	H/H	cis-W/W	3636
U:G	H/H	cis-W/W	3484
C:G	H/H	cis-W/W	1511
G:C	H/H	cis-W/W	1443
G:A	H/H	trans-S/H	1321
A:G	H/H	trans-H/S	1221
G:A	H_e/H_e	trans-S/H	1000
G:U	H/H_qe	cis-W/W	697
U:G	H_qe/H	cis-W/W	680
U:U	H/H	cis-W/W	647
G:U	H_qe/H	cis-W/W	605
U:G	H/H_qe	cis-W/W	509
U:G	H_e/H_e	cis-W/W	492
A:U	H/H_qe	trans-H/W	484
G:A	H_e/H_e	trans-W/H	429
A:G	H/H_qe	trans-H/S	408
A:U	H/H	cis-W/W	386
U:A	H/H	cis-W/W	370
A:G	H_qe/H_qe	trans-H/S	357
U:A	H_qe/H	trans-W/H	349
C:C	H/H	cis-W/W	342
G:A	H_qe/H_qe	trans-S/H	339
G:A	H_qe/H	trans-S/H	325

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:G	H/H_qe	cis-W/W	312
G:A	H_e/H_e	NA	306
C:G	H/H_qe	cis-W/W	280
A:G	H_qe/H	trans-H/S	276
A:U	H_qe/H_qe	trans-H/W	273
U:G	H_qe/H_qe	cis-W/W	269
U:A	H_qe/H_qe	trans-W/H	260
C:G	H_e/H_e	cis-W/W	253
G:C	H_qe/H	cis-W/W	251
U:G	H_e/H_e	trans-S/W	234
A:A	H_qe/H	trans-H/H	231
U:A	H/H_i	trans-W/H	222
G:C	H/H_qe	cis-W/W	220
G:A	H_e/H_ie	NA	219
G:C	H_e/H_e	cis-W/W	209
G:U	H_qe/H_qe	cis-W/W	208
A:G	H/H	cis-W/W	204
U:A	H_e/H_ie	trans-W/H	199
A:C	H/H	cis-W/W	185
U:U	H/H_qe	cis-W/W	175
G:A	H/H_qe	trans-S/H	170
C:G	H/H	NA	168
A:A	H/H_qe	trans-H/H	166
G:C	H_qe/H_qe	cis-W/W	162
A:U	H_i/H	trans-H/W	160
A:G	H_qe/H	cis-W/W	156
C:G	H_qe/H_qe	cis-W/W	155
C:G	H_qe/H	cis-W/W	149
G:A	H/H	cis-W/W	147
A:G	H_e/H_e	NA	145
G:C	H/H	NA	126
U:A	H_e/H_ie	NA	125
G:A	H/H	NA	124
A:U	H/H	trans-H/W	121
U:A	H_e/H_ie	trans-W/W	120
G:U	H/H	NA	120
A:A	H/H	trans-S/H	118
A:A	H/H	trans-H/S	117
U:G	H_e/H_ie	trans-S/W	114
C:U	H/H	cis-W/W	113
A:A	H/H_qe	trans-S/H	112
G:A	H_qe/H_qe	cis-W/W	108
U:C	H/H	cis-W/W	107
A:G	H_qe/H	NA	106
A:A	H/H	cis-W/W	105

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:A	H/H_qe	cis-W/W	104
A:U	H/H_qe	cis-W/W	102
G:G	H/H	trans-H/S	101
G:A	H_qe/H	cis-W/W	101
A:C	H/H	trans-H/W	101
A:G	H_qe/H_qe	cis-W/W	98
A:C	H/H_qe	cis-W/W	98
G:A	H_e/H_ie	trans-S/W	96
C:A	H/H	cis-W/W	94
A:A	H_ie/H_e	cis-S/W	91
A:G	H_i/H	trans-H/S	90
A:G	H/H_qe	NA	89
A:U	H_qe/H	trans-H/W	88
A:A	H_qe/H_i	NA	88
U:A	H/H_qe	trans-W/H	86
C:C	H_qe/H_qe	cis-W/W	86
C:A	H/H	NA	86
A:A	H/H_qe	NA	85
U:U	H_e/H_e	cis-W/W	84
U:A	H/H	NA	84
A:U	H/H	NA	84
A:A	H_e/H_ie	NA	83
U:A	H_e/H_e	trans-S/H	82
G:U	H_e/H_e	cis-W/W	82
C:A	H_e/H_e	NA	82
A:U	H_qe/H	cis-W/W	81
U:G	H/H	NA	78
C:G	H_qe/H_qe	NA	78
U:A	H/H	trans-W/H	76
G:C	H_e/H_ie	trans-W/W	76
A:G	H_qe/H_qe	NA	76
U:A	H_e/H_e	cis-W/W	74
U:U	H_qe/H	cis-W/W	73
A:A	H/H	trans-H/H	73
G:A	H_qe/H	NA	72
A:U	H/H_i	trans-H/W	72
U:A	H_i/H_i	trans-W/H	71
A:A	H_qe/H	NA	71
U:A	H_qe/H_i	trans-W/H	70
U:A	H/H_qe	cis-W/W	69
C:A	H/H	trans-S/H	68
A:A	H_i/H_qe	cis-W/S	68
A:A	H/H	NA	68
U:A	H/H_i	cis-W/H	67
G:A	H_qe/H_qe	NA	67

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:G	H/H_i	NA	67
C:A	H_e/H_e	trans-W/H	67
A:U	H/H_i	cis-H/W	67
A:A	H_i/H_qe	trans-W/W	67
C:A	H_e/H_e	cis-W/W	66
A:A	H_qe/H_i	trans-W/W	66
G:G	H_e/H_ie	NA	65
G:C	H_e/H_e	NA	65
U:A	H_i/H	cis-W/S	63
G:G	H_qe/H	cis-W/W	63
A:U	H_i/H_i	trans-H/W	63
A:A	H/H_qe	trans-H/S	63
G:C	H/H_qe	cis-W/S	62
G:A	H_qe/H_qe	cis-H/W	62
A:A	H_i/H	trans-H/H	61
G:A	H_ie/H_ie	trans-S/H	60
A:G	H_i/H	cis-S/S	60
A:A	H_i/H	trans-W/H	60
U:A	H/H_i	trans-W/S	59
C:G	H_ie/H_ie	NA	59
A:U	H_i/H_i	NA	59
U:G	H_e/H_e	NA	58
U:A	H_e/H_ie	cis-W/H	57
G:A	H_e/H_ie	trans-S/H	57
C:U	H/H	trans-H/S	57
C:C	H_e/H_e	cis-W/W	57
C:A	H_qe/H_qe	trans-W/H	57
A:A	H/H	trans-H/W	57
U:A	H_i/H	cis-W/H	56
G:G	H_e/H_e	NA	56
G:C	H_qe/H	NA	56
G:C	H/H_qe	NA	56
C:G	H/H_qe	NA	56
U:U	H_e/H_e	NA	55
G:G	H/H	trans-S/H	55
A:U	H_qe/H	NA	55
C:U	H/H_i	cis-W/W	54
A:U	H_i/H	cis-W/W	54
U:U	H_qe/H_qe	cis-W/W	52
U:A	H_ie/H_e	trans-W/W	52
G:U	H/H_qe	NA	52
G:A	H_qe/H_i	cis-W/H	52
C:G	H_e/H_e	NA	52
A:U	H/H_qe	NA	52
A:C	H_i/H	trans-H/W	52

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:A	H_i/H_qe	NA	52
U:G	H/H	cis-S/W	51
U:A	H_qe/H_qe	NA	51
G:A	H/H_i	trans-S/H	50
A:G	H_i/H	NA	49
A:U	H_qe/H_qe	cis-W/W	48
U:U	H/H	NA	47
U:G	H_qe/H_i	cis-H/S	47
A:U	H_i/H_qe	trans-H/W	47
A:U	H_i/H_i	cis-H/W	47
G:A	H/H	trans-W/H	46
G:A	H/H_i	trans-S/S	46
A:A	H_qe/H_qe	trans-H/H	46
U:G	H_qe/H_i	cis-W/W	45
C:C	H/H	trans-H/S	45
A:C	H/H	trans-H/S	45
G:A	H_qe/H	trans-W/H	44
G:G	H_i/H	cis-W/H	43
C:C	H/H_qe	cis-W/W	43
C:A	H_e/H_ie	trans-S/H	43
A:U	H_i/H_qe	trans-W/W	43
A:U	H/H	trans-H/S	43
A:G	H/H_i	cis-W/H	43
A:C	H_e/H_e	trans-S/W	43
A:A	H_qe/H_qe	cis-W/W	43
U:C	H/H_qe	cis-W/W	42
G:C	H_qe/H_qe	NA	42
G:C	H_ie/H_ie	NA	42
C:G	H_qe/H	NA	42
A:U	H_qe/H_i	NA	42
A:C	H/H_qe	trans-H/W	42
U:U	H_qe/H_i	cis-W/W	41
G:G	H/H_i	trans-W/H	41
C:C	H_e/H_e	NA	41
A:U	H_e/H_ie	cis-S/W	41
PSU:G	H_e/H_ie	trans-S/W	40
C:A	H_i/H_qe	NA	40
A:U	H_ie/H_ie	cis-H/W	40
A:A	H/H_qe	trans-W/H	40
A:A	H_e/H_e	trans-H/W	40
G:C	H_ie/H_ie	trans-W/W	39
C:A	H_ie/H_e	NA	39
C:A	H_e/H_ie	NA	39
A:G	H_e/H_e	trans-H/S	39
A:C	H_qe/H_i	trans-H/W	39

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:A	H_qe/H	trans-H/W	39
A:A	H/H_qe	trans-W/W	39
U:C	H_i/H	NA	38
U:C	H/H_qe	trans-W/W	38
G:G	H/H_i	cis-W/H	38
A:G	H_i/H_qe	NA	38
A:G	H/H	NA	38
A:C	H_qe/H_i	cis-W/H	38
U:A	H_i/H	trans-W/H	37
U:A	H_e/H_e	trans-W/H	37
A:A	H_e/H_e	trans-S/H	37
A:A	H_ie/H_ie	NA	36
U:G	H_qe/H	NA	35
G:U	H/H	trans-S/H	35
G:G	H_e/H_ie	trans-W/H	35
C:A	H/H_qe	NA	35
U:A	H_e/H_e	NA	34
G:U	H/H_i	cis-W/S	34
C:C	H_i/H	cis-W/W	34
C:A	H/H	trans-H/S	34
A:A	H_i/H	NA	34
U:A	H_qe/H_qe	cis-W/W	33
U:A	H_qe/H	cis-W/W	33
C:U	H/H_i	NA	33
C:A	H_qe/H_qe	cis-W/W	33
A:U	H_i/H	NA	33
G:C	H_qe/H_i	NA	32
G:A	H_ie/H_ie	NA	32
G:A	H_e/H_ie	cis-S/W	32
A:G	H_qe/H_i	NA	32
A:U	H_ie/H_ie	NA	31
A:C	H_qe/H_qe	NA	31
A:C	H/H	NA	31
A:A	H_qe/H_i	cis-S/H	31
A:G	H_i/H_qe	trans-W/S	30
A:G	H_ie/H_e	NA	30
A:G	H/H_i	NA	30
A:A	H_qe/H_qe	NA	30
A:A	H_e/H_e	NA	30
U:G	H_qe/H_i	NA	29
U:C	H_e/H_e	NA	29
G:U	H_qe/H_qe	NA	29
G:U	H_i/H_qe	NA	29
A:A	H/H_qe	cis-W/W	29
U:G	H_qe/H	trans-H/S	28

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:A	H_e/H_ie	trans-S/S	28
U:A	H_e/H_e	trans-S/S	28
G:G	H_qe/H_i	NA	28
G:A	H_qe/H_i	NA	28
C:G	H_e/H_ie	NA	28
A:U	H_e/H_ie	trans-W/W	28
A:C	H_qe/H_qe	cis-W/W	28
U:U	H_i/H	cis-W/W	27
U:A	H_ie/H_ie	NA	27
G:U	H_i/H_qe	cis-S/H	27
G:A	H_i/H	trans-S/H	27
C:A	H_i/H_qe	trans-W/H	27
A:U	H_e/H_e	trans-H/W	27
A:A	H_qe/H_qe	trans-H/S	27
U:A	H_qe/H	NA	26
G:G	H/H_qe	trans-H/W	26
G:C	H_ie/H_e	cis-W/W	26
G:A	H_qe/H	trans-H/H	26
A:U	H_qe/H	trans-W/W	26
A:U	H_i/H_qe	cis-W/W	26
A:A	H_qe/H_qe	trans-W/W	26
A:A	H/H_qe	trans-H/W	26
5MU:A	H/H_i	trans-W/H	26
U:A	H_ie/H_e	NA	25
C:A	H/H	trans-W/H	25
U:U	H_ie/H_ie	trans-W/W	24
U:A	H_i/H_qe	NA	24
G:U	H_qe/H	NA	24
G:A	H_i/H_qe	NA	24
G:A	H_i/H_i	NA	24
A:A	H_i/H_i	trans-H/H	24
5MU:1MA	H/H_i	trans-W/H	24
G:G	H/H	cis-H/W	23
G:C	H_i/H_qe	NA	23
G:A	H/H_qe	NA	23
C:G	H_ie/H_ie	cis-W/W	23
C:G	H_e/H_ie	trans-W/W	23
A:U	H_e/H_e	cis-W/W	23
A:A	H_ie/H_e	NA	23
U:G	H_ie/H_ie	NA	22
U:C	H_qe/H_qe	cis-W/W	22
U:A	H_i/H_i	cis-W/W	22
G:U	H/H	trans-W/H	22
G:U	H_e/H_e	NA	22
G:G	H_ie/H_e	NA	22

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:A	H_ie/H_ie	trans-S/W	22
C:A	H_qe/H	cis-W/W	22
C:A	H/H_qe	cis-W/W	22
A:G	H_i/H_qe	trans-H/S	22
A:C	H_qe/H_i	NA	22
A:C	H_e/H_ie	NA	22
A:A	H/H_qe	cis-W/S	22
U:G	H_e/H_ie	NA	21
U:C	H_qe/H	trans-S/H	21
G:U	H_qe/H_i	cis-S/H	21
G:G	H/H_qe	trans-W/W	21
G:C	H_i/H	cis-W/W	21
G:A	H_qe/H_qe	trans-W/H	21
C:A	H_qe/H	trans-W/H	21
A:M2G	H_qe/H	NA	21
A:G	H_ie/H_ie	NA	21
A:C	H_qe/H	cis-W/S	21
U:U	H_ie/H_ie	NA	20
U:A	H/H_i	NA	20
G:U	H/H_qe	cis-W/S	20
G:C	H_ie/H_ie	cis-W/W	20
C:U	H/H	NA	20
A:U	H_qe/H_qe	NA	20
A:U	H_qe/H	cis-W/S	20
A:U	H_e/H_e	NA	20
A:G	H_i/H	cis-W/W	20
A:C	H/H_qe	cis-W/S	20
A:A	H/H	trans-W/H	20
U:U	H_qe/H_qe	NA	19
U:G	H_e/H_ie	cis-W/W	19
U:C	H_e/H_e	trans-S/H	19
U:A	H/H_qe	NA	19
G:U	H_ie/H_ie	NA	19
C:U	H/H	trans-W/S	19
A:U	H_qe/H_i	cis-H/S	19
A:U	H_i/H	trans-W/W	19
A:A	H/H_i	NA	19
PSU:A	H/H	cis-W/W	18
G:C	H_i/H_i	NA	18
A:U	H_e/H_ie	NA	18
A:G	H_ie/H_e	trans-H/S	18
A:A	H_i/H_i	trans-W/W	18
A:A	H/H_i	trans-W/H	18
U:G	H_qe/H	trans-H/W	17
U:G	H_qe/H_qe	NA	17

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:G	H_i/H_qe	NA	17
U:A	H_e/H_ie	cis-W/W	17
G:U	H_qe/H_qe	trans-S/H	17
G:C	H_i/H_i	cis-W/W	17
C:U	H/H_qe	cis-W/W	17
A:G	H_i/H_i	trans-H/S	17
A:A	H_i/H_i	trans-H/W	17
U:U	H_e/H_ie	NA	16
U:G	H_i/H	NA	16
U:C	H_qe/H	NA	16
G:C	H_ie/H_e	NA	16
C:A	H_e/H_e	trans-S/H	16
A:G	H_e/H_ie	NA	16
A:A	H_qe/H_i	trans-H/W	16
U:C	H_qe/H	trans-S/W	15
U:C	H_e/H_ie	NA	15
U:A	H_e/H_e	trans-H/H	15
G:G	H/H_i	NA	15
C:G	H_ie/H_e	NA	15
C:A	H/H	trans-W/S	15
A:A	H/H_i	trans-W/W	15
U:U	H_ie/H_e	NA	14
U:A	H_i/H_i	NA	14
C:G	H/H_qe	cis-H/W	14
C:G	H_e/H_ie	cis-W/H	14
C:A	H_qe/H_qe	NA	14
A:A	H_ie/H_ie	trans-H/H	14
U:U	H_qe/H	NA	13
U:G	H_e/H_e	trans-S/H	13
U:A	H_i/H	NA	13
G:C	H_i/H	NA	13
G:C	H_e/H_ie	NA	13
C:G	H_e/H_ie	cis-W/W	13
A:U	H_i/H_i	cis-W/W	13
A:G	H/H_i	trans-H/S	13
A:A	H_qe/H_qe	trans-S/H	13
A:A	H_e/H_e	trans-H/H	13
U:G	H_e/H_e	trans-W/W	12
U:C	H_qe/H	cis-W/W	12
U:A	H_e/H_e	cis-W/H	12
G:G	H/H_qe	cis-W/H	12
G:A	H_ie/H_e	NA	12
G:A	H/H_i	NA	12
G:A	H_e/H_ie	trans-W/H	12
C:U	H_qe/H_qe	cis-W/W	12

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:G	H_i/H_qe	cis-W/W	12
C:A	H_e/H_e	trans-W/W	12
A:G	H_ie/H_ie	trans-H/S	12
A:C	H_ie/H_e	NA	12
U:U	H_i/H_i	NA	11
U:G	H/H_qe	NA	11
U:G	H_e/H_ie	trans-W/H	11
U:C	H/H	NA	11
U:C	H_e/H_e	cis-W/W	11
U:A	H_qe/H_i	NA	11
U:A	H_e/H_e	cis-S/W	11
G:U	H_qe/H_i	NA	11
G:U	H_i/H_qe	trans-W/W	11
G:U	H_e/H_ie	trans-W/W	11
G:U	H_e/H_ie	cis-W/W	11
G:G	H_qe/H_qe	NA	11
G:G	H/H_qe	NA	11
G:G	H_e/H_e	trans-W/H	11
G:A	H/H_qe	trans-S/W	11
G:A	H/H_qe	cis-W/S	11
C:C	H/H	trans-W/W	11
C:C	H/H	NA	11
C:A	H_ie/H_ie	NA	11
A:U	H_qe/H_i	trans-H/W	11
A:U	H_ie/H_e	cis-H/W	11
A:G	H/H_qe	trans-H/W	11
A:G	H_e/H_e	cis-W/W	11
A:A	H_i/H	trans-W/W	11
A:A	H_i/H_qe	cis-S/H	11
A:A	H_i/H_qe	cis-H/S	11
A:A	H/H	trans-W/S	11
A:A	H/H_qe	cis-W/H	11
U:U	H_qe/H_qe	trans-W/W	10
U:G	H/H	trans-W/W	10
G:U	H_ie/H_e	cis-S/H	10
G:G	H_e/H_ie	cis-S/H	10
G:G	H_e/H_e	cis-W/W	10
G:A	H_i/H_i	trans-S/W	10
G:A	H_ie/H_ie	trans-W/H	10
C:G	H_i/H_i	cis-W/W	10
C:C	H/H_qe	NA	10
C:A	H_i/H_i	NA	10
A:U	H_i/H_qe	NA	10
A:C	H_qe/H	trans-H/S	10
A:C	H_e/H_e	NA	10

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:A	H_qe/H	trans-W/H	10
A:A	H_qe/H_qe	trans-W/H	10
A:A	H_e/H_ie	trans-H/H	10
U:U	H/H_qe	NA	9
U:U	H/H_qe	cis-W/H	9
U:C	H_qe/H_qe	NA	9
U:C	H_ie/H_ie	NA	9
U:C	H/H_qe	NA	9
U:A	H_qe/H_i	trans-W/W	9
U:A	H_qe/H_i	cis-S/H	9
U:A	H_ie/H_ie	trans-W/W	9
U:A	H/H_qe	cis-W/H	9
H2U:U	H_e/H_ie	trans-W/W	9
G:U	H_ie/H_e	NA	9
G:U	H_e/H_ie	NA	9
G:G	H/H	NA	9
G:C	H/H_i	trans-W/W	9
G:A	H_qe/H_i	trans-S/H	9
G:A	H_i/H_i	trans-S/H	9
C:G	H_qe/H_i	NA	9
C:G	H_i/H_i	NA	9
C:C	H/H	trans-S/H	9
A:U	H_qe/H	trans-W/S	9
A:U	H_ie/H_e	trans-H/W	9
A:C	H_qe/H	trans-W/S	9
A:C	H/H	trans-S/H	9
A:C	H/H_qe	NA	9
A:A	H_qe/H_qe	cis-W/H	9
A:A	H_i/H_i	cis-H/H	9
U:G	H_e/H_e	cis-S/W	8
U:A	H_i/H_qe	trans-W/H	8
U:A	H/H_qe	trans-S/H	8
G:U	H/H	cis-W/S	8
G:G	H_qe/H_qe	cis-W/S	8
G:G	H_qe/H_i	trans-S/S	8
G:G	H_i/H_qe	trans-H/W	8
G:G	H/H	cis-W/H	8
G:C	H_i/H_qe	cis-W/W	8
G:C	H_e/H_ie	cis-W/W	8
G:A	H_qe/H_i	trans-S/W	8
G:A	H_ie/H_e	trans-S/H	8
C:G	H_qe/H_i	cis-W/W	8
C:G	H_i/H	NA	8
C:G	H_ie/H_ie	trans-W/W	8
C:C	H_i/H	NA	8

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:C	H_ie/H_e	cis-W/W	8
C:A	H_qe/H	trans-S/H	8
C:A	H_qe/H_qe	trans-W/W	8
C:A	H_qe/H_i	NA	8
C:A	H/H_qe	trans-W/W	8
C:A	H/H_qe	trans-S/H	8
C:A	H_e/H_ie	cis-W/W	8
C:A	H_e/H_e	cis-S/W	8
A:U	H_i/H_qe	cis-H/S	8
A:U	H_ie/H_e	NA	8
A:G	H_qe/H_qe	trans-H/W	8
A:C	H_i/H_qe	cis-W/S	8
A:C	H_i/H	cis-W/S	8
A:C	H/H_qe	trans-S/W	8
A:A	H_qe/H_qe	cis-S/W	8
A:A	H_qe/H_i	cis-S/W	8
A:A	H_i/H_i	NA	8
U:G	H_i/H	cis-S/W	7
U:C	H_ie/H_e	NA	7
U:C	H/H	trans-W/W	7
U:A	H_qe/H_i	cis-W/H	7
U:A	H_ie/H_ie	trans-W/H	7
U:A	H/H_i	trans-W/W	7
PSU:G	H/H_qe	cis-W/W	7
PSU:G	H_e/H_ie	NA	7
G:U	H_qe/H_i	cis-W/W	7
G:G	H_qe/H	NA	7
G:G	H/H	trans-W/H	7
G:G	H/H_qe	trans-H/S	7
G:G	H/H_qe	cis-W/W	7
C:C	H/H	trans-W/S	7
C:C	H_e/H_ie	NA	7
A:U	H_qe/H	trans-H/S	7
A:U	H/H_i	cis-H/S	7
A:G	H_qe/H	cis-S/W	7
A:G	H_i/H_i	NA	7
A:C	H_qe/H	cis-W/W	7
A:C	H_ie/H_ie	trans-H/S	7
A:C	H_e/H_e	trans-S/H	7
A:A	H_qe/H_qe	trans-H/W	7
A:A	H/H_i	cis-W/S	7
A:A	H_e/H_e	cis-W/W	7
U:U	H_i/H	NA	6
U:G	H_i/H_qe	cis-W/W	6
U:G	H_i/H_qe	cis-H/S	6

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:C	H_qe/H	cis-S/W	6
U:A	H_ie/H_ie	trans-S/W	6
G:U	H_i/H_i	cis-W/W	6
G:U	H_ie/H_ie	cis-S/H	6
G:G	H_i/H_i	NA	6
G:G	H_ie/H_e	cis-H/W	6
G:G	H/H_i	cis-H/W	6
G:G	H_e/H_e	trans-S/H	6
G:C	H_qe/H	cis-W/H	6
G:A	H/H_i	trans-S/W	6
G:A	H_e/H_e	trans-W/S	6
C:G	H_qe/H	cis-S/W	6
C:G	H_ie/H_e	cis-W/W	6
C:G	H/H	cis-S/W	6
C:C	H_i/H_qe	NA	6
C:A	H/H_i	NA	6
A:G	H_qe/H_i	trans-H/S	6
A:C	H_qe/H	trans-W/W	6
A:C	H/H_qe	trans-S/H	6
A:C	H_e/H_e	trans-H/W	6
A:C	H_e/H_e	cis-W/W	6
A:A	H_i/H	trans-H/W	6
A:A	H_i/H_qe	trans-H/W	6
A:A	H_i/H_i	trans-W/H	6
A:A	H_ie/H_ie	trans-S/W	6
A:A	H_ie/H_e	trans-W/W	6
A:A	H_e/H_ie	trans-S/W	6
U:U	H/H_i	NA	5
U:U	H/H_i	cis-W/W	5
U:G	H_qe/H	trans-W/W	5
U:C	H_e/H_e	trans-W/.	5
U:A	H_e/H_ie	trans-S/H	5
U:A	H_e/H_ie	cis-S/H	5
U:A	H_e/H_e	trans-S/W	5
G:U	H/H_qe	trans-S/W	5
G:U	H_e/H_e	trans-W/W	5
G:G	H_qe/H_qe	cis-W/W	5
G:G	H_i/H_qe	NA	5
G:G	H/H	trans-W/W	5
G:G	H/H	trans-H/W	5
G:G	H/H_qe	trans-S/W	5
G:C	H_qe/H_i	cis-W/W	5
G:A	H_ie/H_ie	cis-S/W	5
G:A	H_ie/H_ie	cis-S/H	5
C:U	H_ie/H_e	cis-W/W	5

Continued on Next Page...

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:U	H/H_qe	NA	5
C:U	H_e/H_ie	NA	5
C:G	H/H_i	cis-W/W	5
C:A	H_i/H_qe	trans-W/W	5
C:A	H_ie/H_ie	trans-S/H	5
C:A	H/H_qe	cis-S/W	5
C:A	H/H_i	cis-S/S	5
C:A	H_e/H_ie	trans-W/H	5
A:U	H_qe/H_i	cis-W/W	5
A:U	H/H_i	trans-W/W	5
A:U	H/H_i	NA	5
A:G	H_i/H	trans-W/S	5
A:G	H_ie/H_e	trans-H/W	5
A:G	H/H_qe	cis-W/H	5
A:C	H_qe/H_qe	trans-H/W	5
A:C	H_ie/H_ie	NA	5
A:C	H_e/H_ie	trans-H/W	5
A:A	H_qe/H_i	trans-H/H	5
A:A	H_i/H_i	cis-S/H	5
A:A	H_ie/H_e	cis-H/H	5
A:4SU	H/H_i	trans-H/W	5
U:U	H_qe/H_qe	cis-W/S	4
U:U	H_qe/H_i	cis-W/H	4
U:U	H_e/H_ie	trans-W/H	4
U:U	H_e/H_ie	cis-W/W	4
U:G	H_qe/H_qe	cis-S/W	4
U:G	H_i/H_qe	cis-W/H	4
U:G	H_i/H_i	cis-W/W	4
U:G	H_ie/H_ie	cis-W/W	4
U:G	H_ie/H_e	NA	4
U:G	H/H_i	cis-W/W	4
U:G	H_e/H_ie	trans-S/H	4
U:C	H_qe/H_qe	trans-W/W	4
U:C	H_i/H_i	NA	4
U:C	H_e/H_e	trans-W/W	4
U:C	H_e/H_e	trans-S/W	4
U:A	H_qe/H	trans-W/W	4
U:A	H_i/H_qe	cis-W/H	4
U:A	H_ie/H_ie	trans-S/H	4
U:A	H/H	cis-W/H	4
PSU:G	H/H	cis-W/W	4
HPA:C	H_i/H_qe	cis-W/W	4
G:U	H_ie/H_e	cis-S/W	4
G:U	H/H_qe	trans-S/H	4
G:U	H_e/H_e	trans-S/H	4

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:PPU	H_e/H_ie	NA	4
G:G	H_qe/H_qe	trans-W/W	4
G:G	H_qe/H_qe	trans-S/S	4
G:G	H_qe/H_qe	trans-S/H	4
G:G	H_qe/H_qe	trans-H/S	4
G:G	H_qe/H_qe	cis-W/H	4
G:G	H_qe/H_qe	cis-H/W	4
G:G	H_qe/H_i	cis-H/W	4
G:G	H_qe/H	cis-S/W	4
G:G	H_i/H_qe	trans-W/W	4
G:G	H_ie/H_ie	trans-S/H	4
G:G	H_e/H_ie	trans-W/W	4
G:C	H_qe/H	trans-W/W	4
G:A	H_qe/H	trans-H/W	4
G:A	H_qe/H	cis-H/W	4
G:A	H_e/H_ie	cis-W/H	4
G:A	H_e/H_ie	cis-S/H	4
G:5BU	H/H	cis-W/W	4
C:U	H_i/H_qe	NA	4
C:G	H_qe/H_qe	cis-H/W	4
C:G	H_i/H_qe	NA	4
C:G	H/H_qe	trans-W/W	4
C:G	H/H_qe	cis-H/.	4
C:C	H_qe/H_qe	trans-W/S	4
C:C	H_qe/H_qe	cis-W/S	4
C:C	H_qe/H_i	NA	4
C:A	H_i/H_qe	cis-S/H	4
C:A	H/H	trans-W/W	4
C:A	H/H_qe	trans-W/H	4
C:A	H_e/H_ie	trans-W/W	4
A:U	H_qe/H_i	cis-H/W	4
A:U	H_i/H_i	cis-W/S	4
A:G	H/H_i	cis-W/W	4
A:C	H_i/H	trans-W/W	4
A:C	H_ie/H_e	cis-W/S	4
A:A	H_qe/H	trans-S/H	4
A:A	H_qe/H	trans-H/S	4
A:A	H_qe/H_i	cis-H/S	4
A:A	H_qe/H	cis-W/H	4
A:A	H_i/H	cis-S/W	4
A:A	H_ie/H_ie	trans-W/H	4
A:A	H_e/H_ie	trans-W/H	4
A:A	H_e/H_ie	cis-S/W	4
5MU:MAD	H/H_i	trans-W/H	4
U:U	H_qe/H	trans-W/W	3

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:U	H_i/H	trans-H/W	3
U:U	H_i/H_i	cis-W/W	3
U:U	H/H_qe	trans-W/H	3
U:U	H/H	cis-W/S	3
U:U	H_e/H_e	cis-S/W	3
U:I	H/H	cis-W/W	3
U:G	H_i/H_i	NA	3
U:G	H_ie/H_ie	trans-S/W	3
U:G	H_ie/H_ie	cis-H/S	3
U:C	H_ie/H_e	cis-S/W	3
U:C	H_e/H_ie	cis-W/W	3
U:A	H_qe/H_qe	cis-W/H	3
U:A	H_qe/H_i	cis-W/W	3
U:A	H_i/H_i	cis-W/H	3
U:A	H_ie/H_ie	cis-W/H	3
U:A	H/H	trans-S/H	3
U:A	H/H_i	cis-W/W	3
PSU:G	H_qe/H_i	trans-S/W	3
PSU:G	H_e/H_ie	trans-W/W	3
I:U	H/H	cis-W/W	3
G:U	H_i/H_qe	cis-W/W	3
G:U	H/H_i	NA	3
G:U	H/H	cis-H/W	3
G:U	H_e/H_ie	trans-W/.	3
G:PSU	H_qe/H	cis-W/W	3
G:G	H_qe/H	trans-H/S	3
G:G	H_qe/H_qe	trans-H/W	3
G:G	H_ie/H_ie	trans-S/S	3
G:G	H_ie/H_ie	NA	3
G:G	H/H_qe	trans-W/H	3
G:G	H_e/H_e	trans-W/W	3
G:C	H_qe/H_i	trans-W/W	3
G:C	H_qe/H	cis-W/S	3
G:C	H_i/H_qe	trans-W/W	3
G:C	H_i/H_qe	cis-S/H	3
G:C	H/H_i	cis-W/W	3
G:C	H/H	cis-H/W	3
G:A	H_qe/H_qe	trans-S/W	3
G:A	H_qe/H_qe	cis-H/H	3
G:A	H_i/H_qe	cis-W/W	3
G:A	H_i/H_i	cis-H/S	3
G:A	H/H_i	cis-S/S	3
G:A	H_e/H_ie	trans-S/S	3
G:A	H_e/H_ie	cis-W/W	3
G:A	H_e/H_e	cis-W/W	3

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:A	H_e/H_e	cis-W/S	3
G:5MC	H_e/H_ie	trans-W/W	3
G:5AA	H_e/H_ie	NA	3
C:U	H_qe/H_qe	NA	3
C:U	H_qe/H_i	trans-H/S	3
C:U	H_qe/H_i	NA	3
C:U	H_i/H	NA	3
C:U	H_i/H_i	NA	3
C:U	H_e/H_e	NA	3
C:G	H_i/H_qe	trans-W/S	3
C:G	H_i/H	cis-W/W	3
C:C	H_qe/H_i	cis-S/W	3
C:C	H_ie/H_ie	NA	3
C:C	H_ie/H_ie	cis-W/W	3
C:C	H_e/H_e	trans-W/W	3
C:A	H_qe/H_qe	trans-S/H	3
A:U	H_qe/H_qe	trans-W/W	3
A:U	H_qe/H_qe	trans-H/S	3
A:U	H/H_i	cis-W/W	3
A:PSU	H_e/H_e	cis-W/W	3
A:OMC	H_qe/H	trans-W/S	3
A:G	H_qe/H_i	cis-H/S	3
A:G	H_i/H	trans-S/S	3
A:G	H_i/H_qe	cis-W/W	3
A:C	H_qe/H_i	cis-W/W	3
A:C	H_i/H_qe	cis-S/W	3
A:C	H_i/H	NA	3
A:C	H_i/H_i	NA	3
A:A	H_qe/H	trans-W/W	3
A:A	H_qe/H_qe	cis-H/H	3
A:A	H_qe/H_i	trans-W/H	3
A:A	H_qe/H	cis-W/W	3
A:A	H_i/H_qe	trans-H/H	3
A:A	H_ie/H_ie	trans-W/S	3
5BU:G	H/H	cis-W/W	3
U:U	H_qe/H_qe	cis-H/W	2
U:U	H_i/H_qe	NA	2
U:U	H_ie/H_ie	cis-W/W	2
U:U	H_ie/H_e	cis-W/W	2
U:U	H_e/H_ie	trans-W/W	2
U:U	H_e/H_ie	cis-S/W	2
U:U	H_e/H_e	trans-W/W	2
U:G	H_qe/H_qe	trans-W/H	2
U:G	H_qe/H_i	cis-W/S	2
U:G	H_i/H	cis-W/W	2

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:G	H_e/H_ie	cis-H/W	2
U:C	H_i/H	cis-W/W	2
U:C	H/H	trans-S/H	2
U:A	H_qe/H	trans-S/H	2
U:A	H_qe/H_qe	trans-W/W	2
U:A	H_i/H_qe	trans-S/H	2
U:A	H_ie/H_e	cis-W/W	2
U:A	H_ie/H_e	cis-W/H	2
U:A	H_ie/H_e	cis-S/W	2
U:A	H/H	cis-S/W	2
LHU:LG	H/H	cis-W/W	2
LG:LHU	H/H	cis-W/W	2
I:I	H/H	cis-W/H	2
I:C	H/H	cis-X/X	2
H2U:U	H_e/H_ie	NA	2
G:U	H_qe/H_qe	cis-W/S	2
G:U	H_i/H_i	NA	2
G:U	H_i/H_i	cis-S/H	2
G:U	H_ie/H_e	cis-W/W	2
G:U	H_e/H_ie	trans-S/W	2
G:IU	H/H	cis-W/W	2
G:G	H_qe/H_i	cis-W/H	2
G:G	H_i/H	trans-W/H	2
G:G	H_i/H_qe	cis-W/W	2
G:G	H_i/H_qe	cis-W/H	2
G:G	H_i/H_i	trans-S/S	2
G:G	H_ie/H_e	cis-S/W	2
G:G	H/H_i	trans-W/W	2
G:G	H/H	cis-W/S	2
G:G	H_e/H_ie	cis-W/W	2
G:G	H_e/H_ie	cis-S/W	2
G:G	H_e/H_e	cis-H/W	2
G:C	H_qe/H_qe	trans-S/W	2
G:C	H_i/H	trans-W/W	2
G:C	H_ie/H_e	cis-W/S	2
G:C	H_ie/H_e	cis-S/W	2
G:C	H/H_i	NA	2
G:C	H_e/H_e	cis-W/H	2
G:A	H_qe/H	cis-S/W	2
G:A	H_i/H_qe	trans-S/H	2
G:A	H/H_qe	trans-W/H	2
G:A	H/H	cis-H/W	2
G:5MU	H/H_qe	cis-W/W	2
G:5BU	H_qe/H_qe	cis-W/W	2
C:U	H_ie/H_ie	NA	2

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:U	H/H_qe	trans-H/S	2
C:U	H/H	cis-W/S	2
C:I	H/H	cis-X/X	2
C:G	H_qe/H_qe	trans-W/W	2
C:G	H_i/H	cis-W/.	2
C:C	H_qe/H_qe	trans-H/S	2
C:C	H_qe/H_qe	NA	2
C:C	H_qe/H_i	cis-S/H	2
C:C	H_ie/H_e	trans-W/W	2
C:C	H_ie/H_e	NA	2
C:C	H/H_qe	trans-W/W	2
C:C	H_e/H_e	trans-S/W	2
C:A	H_qe/H_i	trans-W/H	2
C:A	H_qe/H_i	cis-W/H	2
C:A	H_i/H	trans-W/W	2
C:A	H_i/H_i	trans-W/H	2
C:A	H_ie/H_ie	trans-W/W	2
C:A	H_ie/H_e	cis-S/H	2
C:A	H/H	trans-S/W	2
C:A	H/H_i	cis-S/W	2
C:A	H_e/H_ie	cis-S/W	2
C:A	H_e/H_ie	cis-S/H	2
C:A	H_e/H_e	trans-S/W	2
C:A	H_e/H_e	trans-H/H	2
AVC:A	H_qe/H_qe	trans-S/H	2
A:U	H_ie/H_ie	trans-H/W	2
A:U	H_ie/H_ie	cis-W/W	2
A:U	H_ie/H_e	cis-W/W	2
A:U	H/H_qe	trans-W/W	2
A:U	H/H	cis-H/W	2
A:U	H_e/H_ie	trans-H/W	2
A:U	H_e/H_ie	cis-H/W	2
A:PSU	H/H	cis-W/W	2
A:OMC	H/H_i	trans-H/W	2
A:M2G	H_qe/H	cis-W/W	2
A:G	H_qe/H_qe	cis-W/H	2
A:G	H_qe/H_i	cis-W/W	2
A:G	H_i/H_qe	cis-S/H	2
A:G	H_i/H	cis-W/.	2
A:G	H_ie/H_ie	trans-W/S	2
A:G	H_ie/H_ie	cis-W/W	2
A:G	H/H	cis-W/H	2
A:G	H/H	cis-H/W	2
A:G	H_e/H_ie	trans-H/S	2
A:G	H_e/H_e	cis-S/S	2

Continued on Next Page...

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:C	H_i/H_qe	cis-W/W	2
A:C	H_i/H_i	cis-W/W	2
A:C	H_i/H	cis-S/W	2
A:C	H_ie/H_ie	cis-H/S	2
A:C	H_ie/H_e	trans-W/W	2
A:C	H_ie/H_e	cis-S/W	2
A:A	H_qe/H	cis-H/W	2
A:A	H_i/H_qe	trans-W/H	2
A:A	H_ie/H_ie	cis-W/S	2
A:A	H_ie/H_e	cis-S/H	2
A:A	H/H_qe	cis-S/W	2
A:A	H/H_i	cis-W/W	2
A:A	H_e/H_ie	cis-S/H	2
A:A	H_e/H_e	trans-W/H	2
A5M:A	H/H	cis-W/W	2
5MU:G	H/H_i	trans-W/H	2
5MU:A	H_e/H_ie	trans-W/H	2
5MC:G	H/H	cis-W/W	2
5BU:G	H_qe/H_qe	cis-W/W	2
5BU:5BU	H/H	cis-W/W	2
XUG:C	H/H	cis-W/W	1
U:U	H_qe/H_qe	cis-W/H	1
U:U	H_qe/H_qe	cis-S/W	1
U:U	H_qe/H_i	trans-W/H	1
U:U	H_qe/H_i	NA	1
U:U	H_i/H	trans-W/S	1
U:U	H_i/H_qe	trans-S/W	1
U:U	H_i/H_qe	cis-S/W	1
U:U	H_i/H_i	cis-W/H	1
U:U	H_i/H	cis-W/H	1
U:U	H_ie/H_ie	cis-S/W	1
U:U	H_ie/H_e	trans-W/W	1
U:U	H_e/H_ie	trans-S/H	1
U:U	H_e/H_ie	cis-W/H	1
U:U	H_e/H_e	trans-S/W	1
U:N6G	H_qe/H	trans-H/S	1
U:MTU	H_ie/H_e	trans-H/S	1
U:MAD	H/H_i	trans-W/H	1
U:G	H_qe/H	trans-W/S	1
U:G	H_qe/H	trans-./W	1
U:G	H_i/H_qe	cis-W/S	1
U:G	H_i/H_i	trans-W/S	1
U:G	H_ie/H_ie	trans-W/S	1
U:G	H_ie/H_e	cis-W/W	1
U:G	H_ie/H_e	cis-W/S	1

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
U:G	H/H_i	trans-S/H	1
U:G	H/H_i	NA	1
U:G	H/H	cis-W/S	1
U:G	H/H	cis-W/H	1
U:G	H_e/H_e	trans-./W	1
U:G	H_e/H_e	trans-S/S	1
U:G	H_e/H_e	cis-S/H	1
U:GDP	H_e/H_e	cis-W/W	1
U:C	H_qe/H_qe	trans-S/W	1
U:C	H_qe/H_qe	trans-S/H	1
U:C	H_qe/H_qe	cis-W/S	1
U:C	H_i/H_qe	NA	1
U:C	H_ie/H_ie	trans-W/.	1
U:C	H_ie/H_ie	trans-S/H	1
U:C	H_ie/H_ie	cis-S/S	1
U:C	H/H	cis-W/H	1
U:C	H_e/H_ie	trans-W/W	1
U:C	H_e/H_e	trans-W/H	1
U:A	H_qe/H_qe	trans-S/H	1
U:A	H_qe/H	cis-S/W	1
U:A	H_i/H_qe	trans-H/S	1
U:A	H_i/H_qe	cis-W/W	1
U:A	H_i/H	cis-S/H	1
U:A	H_ie/H_ie	cis-S/S	1
U:A	H_ie/H_ie	cis-S/H	1
U:A	H_ie/H_e	cis-S/S	1
U:A	H_ie/H_e	cis-H/H	1
U:A	H_e/H_ie	trans-S/W	1
U:A	H_e/H_e	trans-W/W	1
U2N:A	H_qe/H_qe	trans-W/H	1
SAM:57	NA	non-WC	1
PSU:OMG	H_e/H_ie	trans-S/W	1
PSU:G	H_ie/H_ie	trans-S/W	1
PSU:G	H/H_qe	trans-W/W	1
PSU:G	H_e/H_ie	trans-./W	1
PSU:A	H_qe/H_i	trans-W/H	1
PSU:A	H_e/H_ie	trans-W/H	1
OMC:A	H_e/H_e	cis-W/W	1
MTU:A	H_qe/H	cis-W/W	1
M2G:A	H/H_qe	cis-W/W	1
I:I	H_qe/H_qe	cis-H/W	1
HPA:74	NA	non-WC	1
G:U	H_qe/H_qe	trans-./H	1
G:U	H_qe/H_qe	cis-S/W	1
G:U	H_i/H_qe	trans-S/W	1

Continued on Next Page...

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:U	H_i/H_qe	cis-W/S	1
G:U	H_i/H_qe	cis-S/S	1
G:U	H_i/H_qe	cis-H/W	1
G:U	H_i/H	NA	1
G:U	H_ie/H_ie	trans-W/W	1
G:U	H_ie/H_e	trans-W/W	1
G:U	H/H	trans-W/W	1
G:U	H/H_i	cis-S/S	1
G:U	H_e/H_ie	cis-S/H	1
G:U	H_e/H_e	trans-W/H	1
G:G	H_qe/H	trans-W/S	1
G:G	H_qe/H	trans-W/H	1
G:G	H_qe/H	trans-S/H	1
G:G	H_qe/H_qe	trans-W/H	1
G:G	H_qe/H_i	trans-W/W	1
G:G	H_qe/H_i	trans-W/H	1
G:G	H_qe/H_i	cis-H/S	1
G:G	H_qe/H	cis-W/H	1
G:G	H_i/H	trans-H/W	1
G:G	H_i/H_qe	cis-S/H	1
G:G	H_i/H_i	trans-W/H	1
G:G	H_i/H_i	cis-W/H	1
G:G	H_i/H_i	cis-H/W	1
G:G	H_ie/H_ie	cis-W/W	1
G:G	H_ie/H_ie	cis-W/H	1
G:G	H_ie/H_e	trans-S/S	1
G:G	H/H_qe	trans-S/H	1
G:G	H/H_qe	cis-S/W	1
G:G	H/H_i	cis-W/W	1
G:G	H/H	cis-W/W	1
G:G	H/H	cis-S/W	1
G:G	H_e/H_ie	trans-S/W	1
G:G	H_e/H_ie	trans-S/H	1
G:G	H_e/H_e	trans-H/W	1
G:G	H_e/H_e	cis-S/W	1
G:CSL	H/H	cis-W/W	1
G:C	H_qe/H	cis-W/.	1
G:C	H_i/H_i	cis-H/S	1
G:C	H_i/H	cis-W/S	1
G:C	H/H	trans-W/W	1
G:C	H/H	cis-W/S	1
G:C	H_e/H_ie	cis-H/W	1
G:C	H_e/H_e	trans-W/W	1
G:C	H_e/H_e	trans-S/H	1
G:C	H_e/H_e	cis-W/S	1

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
G:C	H_e/H_e	cis-H/W	1
G:A	H_qe/H_qe	trans-W/S	1
G:A	H_qe/H_qe	cis-S/S	1
G:A	H_qe/H_i	trans-H/W	1
G:A	H_qe/H_i	cis-S/H	1
G:A	H_qe/H_i	cis-./.	1
G:A	H_qe/H	cis-H/H	1
G:A	H_i/H_qe	cis-S/H	1
G:A	H_i/H	NA	1
G:A	H_i/H	cis-W/W	1
G:A	H_ie/H_ie	trans-S/S	1
G:A	H_ie/H_e	trans-W/H	1
G:A	H_ie/H_e	cis-S/H	1
G:A	H_e/H_ie	cis-W/S	1
G:A	H_e/H_e	trans-S/W	1
G:5BU	H_i/H_i	cis-W/W	1
G:5BU	H_e/H_e	cis-W/W	1
G:2AD	H_qe/H_i	trans-S/S	1
DU:G	H_qe/H	trans-H/S	1
DG:C	H_e/H_e	cis-W/W	1
DC:G	H_e/H_e	cis-W/W	1
DC:A	H/H	cis-W/W	1
DA:A	H_qe/H_i	trans-S/H	1
C:U	H_qe/H_qe	cis-W/S	1
C:U	H_qe/H	NA	1
C:U	H_i/H	trans-S/W	1
C:U	H_i/H_i	cis-S/H	1
C:U	H_ie/H_e	NA	1
C:I	H_e/H_e	cis-X/X	1
C:G	H_qe/H_qe	cis-W/H	1
C:G	H_qe/H_qe	cis-S/W	1
C:G	H_qe/H_qe	cis-H/.	1
C:G	H_i/H_qe	trans-S/W	1
C:G	H_i/H_i	cis-W/H	1
C:G	H_ie/H_ie	cis-W/H	1
C:G	H_ie/H_e	cis-S/W	1
C:G	H/H_qe	cis-H/H	1
C:G	H/H_i	trans-W/W	1
C:G	H/H_i	trans-W/S	1
C:G	H/H_i	cis-S/W	1
C:G	H/H_i	cis-S/H	1
C:G	H_e/H_ie	cis-S/S	1
C:G	H_e/H_e	trans-S/H	1
C:G	H_e/H_e	cis-S/W	1
C:C	H_qe/H_qe	trans-W/H	1

Continued on Next Page…

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
C:C	H_qe/H	NA	1
C:C	H_qe/H_i	cis-W/W	1
C:C	H_i/H_qe	cis-W/W	1
C:C	H_i/H_qe	cis-S/H	1
C:C	H_ie/H_e	trans-./S	1
C:C	H/H_i	cis-H/W	1
C:C	H/H	cis-H/W	1
C:C	H_e/H_ie	trans-W/W	1
C:C	H_e/H_e	cis-S/H	1
C:A	H_qe/H_qe	trans-S/W	1
C:A	H_qe/H	NA	1
C:A	H_i/H	trans-W/H	1
C:A	H_i/H_qe	trans-S/H	1
C:A	H_i/H	cis-W/S	1
C:A	H_ie/H_ie	cis-S/S	1
C:A	H_ie/H_e	trans-S/S	1
C:A	H/H	trans-S/.	1
C:A	H/H_qe	trans-S/W	1
C:A	H/H	cis-S/W	1
C:A	H_e/H_ie	cis-H/S	1
C:A	H_e/H_e	trans-W/.	1
C:A	H_e/H_e	trans-S/.	1
C:A	H_e/H_e	trans-H/W	1
C:A	H_e/H_e	cis-W/H	1
A:U	H_qe/H_qe	trans-W/H	1
A:U	H_qe/H_qe	cis-W/S	1
A:U	H_qe/H_qe	cis-H/W	1
A:U	H_qe/H_i	trans-W/W	1
A:U	H_qe/H	cis-H/W	1
A:U	H_i/H_qe	cis-S/S	1
A:U	H_i/H_qe	cis-H/W	1
A:U	H_i/H_i	trans-W/W	1
A:U	H_i/H_i	cis-W/H	1
A:U	H_i/H	cis-W/S	1
A:U	H_i/H	cis-H/W	1
A:U	H_ie/H_e	cis-S/W	1
A:U	H_e/H_ie	trans-H/S	1
A:U	H_e/H_ie	cis-W/W	1
A:PSU	H/H	cis-W/H	1
A:M5M	H/H	cis-W/W	1
A:M2G	H_qe/H_qe	cis-W/W	1
A:G	H_qe/H_qe	cis-S/S	1
A:G	H_qe/H_i	trans-W/S	1
A:G	H_qe/H_i	cis-S/H	1
A:G	H_i/H	trans-W/.	1

Continued on Next Page...

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
A:G	H_i/H_qe	trans-S/S	1
A:G	H_i/H_i	cis-H/S	1
A:G	H/H	trans-S/H	1
A:G	H/H_qe	trans-W/H	1
A:G	H/H_i	trans-W/S	1
A:G	H_e/H_ie	trans-S/S	1
A:G	H_e/H_ie	cis-W/W	1
A:G	H_e/H_e	trans-H/W	1
A:G	H_e/H_e	cis-W/S	1
A:C	H_qe/H_qe	trans-W/W	1
A:C	H_qe/H_qe	trans-S/H	1
A:C	H_qe/H_qe	trans-H/S	1
A:C	H_qe/H_qe	cis-S/S	1
A:C	H_i/H	trans-W/S	1
A:C	H_i/H_qe	NA	1
A:C	H_i/H_i	trans-H/S	1
A:C	H/H	trans-./S	1
A:C	H/H_qe	cis-H/W	1
A:C	H/H_i	trans-W/W	1
A:C	H_e/H_ie	cis-W/W	1
A:C	H_e/H_e	cis-H/W	1
A:A	H_qe/H	trans-S/.	1
A:A	H_qe/H_qe	trans-./S	1
A:A	H_qe/H_i	trans-S/W	1
A:A	H_qe/H_i	trans-S/H	1
A:A	H_i/H_qe	cis-H/W	1
A:A	H_ie/H_ie	trans-S/H	1
A:A	H_ie/H_ie	trans-H/S	1
A:A	H_ie/H_ie	cis-S/W	1
A:A	H_ie/H_ie	cis-S/S	1
A:A	H_ie/H_e	trans-H/H	1
A:A	H/H_i	trans-H/H	1
A:A	H_e/H_ie	trans-H/W	1
A:A5M	H/H	cis-W/W	1
A:5BU	H/H	cis-H/W	1
A:4OC	H/H	cis-W/W	1
A2M:A	H_qe/H_qe	trans-W/H	1
A2M:A	H_qe/H_qe	trans-S/H	1
A2M:A	H_qe/H_i	trans-S/H	1
A:1MA	H/H_i	trans-W/H	1
6AP:74	NA	non-WC	1
5MC:G	H_qe/H	cis-W/W	1
5BU:G	H_i/H_i	cis-W/W	1
5BU:A	H_i/H_qe	cis-W/H	1
5BU:A	H/H	cis-W/H	1

Continued on Next Page. . .

Table 3.9 – Continued

Base Pair Type	Helical Region Type	LW Classification	Counts
3DA:A	H_qe/H_qe	trans-S/H	1
3AY:U	H_i/H_qe	NA	1
2MU:1MA	H/H_i	trans-W/H	1
2MG:U	H/H	cis-W/W	1

3.2 Base Pair-Step (BPS) Counting and Classification

Table 3.10: Number of Base Pair Steps (BPS) per BPS-ID.

BPS ID	Number of BPS
C:G:C:G	11751
G:C:G:C	11043
G:C:C:G	8541
C:G:G:C	7472
C:G:U:A	4640
A:U:C:G	4269
U:A:C:G	4171
U:A:G:C	4120
G:C:U:A	4047
A:U:G:C	3980
C:G:A:U	3940
G:C:A:U	3839
G:C:U:G	2719
G:U:C:G	2504
A:U:A:U	2328
C:G:U:G	2093
C:G:G:A	2005
G:U:G:C	1925
A:G:G:C	1835
A:U:U:A	1534
U:G:C:G	1494
U:A:A:U	1487
G:C:G:U	1487
A:G:C:G	1362
U:A:U:A	1248
U:G:G:C	1133
G:C:G:A	1082
C:G:G:U	954
G:A:A:U	938
G:U:U:A	935
U:A:A:G	886
G:A:A:G	778
U:A:G:A	697

U:G:G:U	694
A:A:G:C	685
U:G:G:A	638
A:G:G:U	623
A:U:U:G	609
U:G:A:U	588
G:A:G:C	570
A:U:A:A	566
A:A:A:G	548
U:A:G:U	540
G:C:A:G	522
U:U:C:G	518
U:G:U:A	508
G:U:A:U	498
A:U:G:U	497
U:A:U:G	494
A:G:A:U	488
G:C:A:A	478
A:A:U:A	459
C:G:A:A	456
G:A:A:A	451
A:C:G:C	410
G:C:U:U	404
A:U:G:A	391
U:U:G:C	343
G:A:C:G	341
A:A:C:G	339
A:U:A:G	331
C:G:A:G	330
G:U:G:U	311
G:C:C:A	311
A:U:C:C	296
G:A:U:A	288
A:G:U:A	286
G:A:G:A	276
A:A:A:U	267
C:G:U:U	266
C:G:A:C	264
G:C:A:C	255
C:A:C:G	252
U:A:A:A	239
U:G:A:G	233
A:G:A:G	227
C:A:G:C	209
C:G:C:A	204
A:A:A:A	197
C:G:U:C	190

G:C:G:G	189
G:G:G:C	180
C:G:C:C	180
A:C:C:G	178
C:U:G:U	172
A:U:A:C	168
U:G:A:A	158
C:G:G:G	158
A:A:U:G	157
A:U:U:U	156
C:U:G:C	154
A:A:C:A	154
A:A:G:U	147
A:G:U:U	140
A:A:G:A	140
C:C:G:A	132
A:C:U:A	132
G:G:G:U	129
A:G:U:G	125
U:C:C:G	124
C:C:G:C	124
U:U:U:A	123
C:C:A:A	121
C:C:U:A	120
A:A:C:C	119
A:C:G:A	118
U:G:U:G	115
C:G:C:U	112
A:C:G:U	112
U:G:U:U	107
A:U:C:A	107
U:A:G:G	102
G:C:C:C	102
A:A:C:U	101
C:A:U:A	100
Total	131934

References

[1] Ruvkun, G. (2001) Glimpses of a tiny rna world. *Science,* **294**, 797–799.

Appendix A
Clustering Analysis (CA)

A.1 Hierarchical methods

The hierarchical clustering methods used were:

1. *Single linkage clustering*, where the minimum distance between elements of each cluster is taken as clustering criteria.

$$D(X,Y) = min\{d(x_i, y_j) : x_i \in X, y_j \in Y\} \tag{A.1}$$

where X and Y are vectors, and $d(x_i, y_j)$ is the distance between cluster elements.

2. *Complete linkage clustering*, where the maximum distance between cluster elements is the clustering criteria.

$$D(X,Y) = max\{d(x_i, y_j) : x_i \in X, y_j \in Y\} \tag{A.2}$$

3. *Average linkage clustering*, the mean distance between elements of each cluster is taken as clustering criteria.

$$D(X,Y) = \frac{1}{N_x * N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} d(x_i, y_j) \tag{A.3}$$

where N_x and N_y are the number of elements in respective clusters.

4. *Centroid linkage clustering*, uses the distance between cluster centroids, as clustering criteria.

$$D(X,Y) = d(\overline{x}, \overline{y}) \tag{A.4}$$

$$\overline{x} = \frac{1}{N_x} \sum_{i=1}^{N_x} x_i \tag{A.5}$$

$$\overline{y} = \frac{1}{N_y} \sum_{i=1}^{N_y} y_i \tag{A.6}$$

$$\tag{A.7}$$

Structure	Property I	Property II
1	1.00	5.00
2	-2.00	6.00
3	2.00	-2.00
4	-2.00	-3.00
5	3.00	-4.00

Table A.1: Example of structures, considered as bidimensional vectors, to be clustered using the average linkage method and the Manhattan distance.

5. *Ward's Method*, uses the error sum of squares (ESS).

$$D(X,Y) = ESS(XY) - [ESS(X) + ESS(Y)] \tag{A.8}$$

$$ESS(X) = \sum_{i=1}^{N_x} \left| x_i - \frac{1}{N_x} \sum_{j=1}^{N_x} x_j \right|^2 \tag{A.9}$$

As an example lets think of a case where we have five structures. Each one of them is descibed by a bidimensional vector as illustrated in Table A.1.

The first step is to chose a distance definition. We chose Manhattan and the distance values between structures can be displayed in a lower triangular matrix as seen in equation A.10

$$d(X,Y) = \left|\begin{array}{lllll} & 1 & 2 & 3 & 4 \\ 1 & & & & \\ 2 & 4 & & & \\ 3 & 8 & 12 & & \\ 4 & 11 & 9 & 5 & \\ 5 & 11 & 15 & 3 & 6 \end{array}\right| \tag{A.10}$$

Let's calculate explicitily the Manhattan distance between structures 2 and 3,

$$d(2,3) = |-2.00 - 6.00| + |2.00 - -2.00| = 12 \tag{A.11}$$

Now that we have calculated the distances we need a clustering method, in this case, we will use the average linkage clustering method. The first step is to group whatever structures are closer, that is, structures 3 and 5 ($d(3,5) = 3$). Now we find the mean distance between the elements of this cluster and the remaining unclustered structures, that is, structures 1, 2 and 4, we obtain the following mean distances

$$D(\{3,5\},1) = \frac{1}{2*1} * (8+11) = 4.5 \tag{A.12}$$

$$D(\{3,5\},2) = \frac{1}{2*1} * (12+15) = 13.5 \tag{A.13}$$

$$D(\{3,5\},4) = \frac{1}{2*1} * (5+6) = 5.5 \tag{A.14}$$

Since the distances between {3, 5} and all remaining unclustered vectors is higher than the distance between vectors 1 and 2 ($d(1,2) = 4$) then {1, 2} are grouped. The following value, in hierarchical increasing order is 4.5 between {3, 5} and 1 (see equation A.12), but since 1 and 2 are already grouped we can't group {3, 5} with 1. The next value, following the lower to higher hierarchy, is 5 ($d(3,4) = 5$),

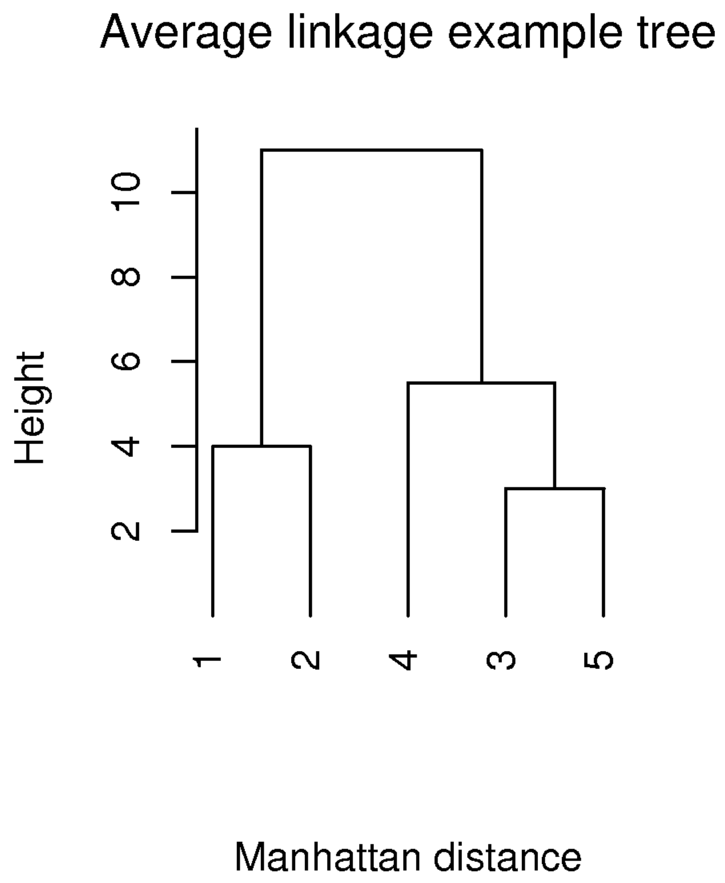

Figure A.1: Clustering tree for 5 bidimensional vectors using the Manhattan distance definition and the average linkage clustering method.

but we have already grouped 3 with 5, so we have to keep advancing in the hierarchy. The next value is 5.5, which corresponds to grouping {3, 5} with 4, so we cluster them. The only remaining possibility for grouping is, group {1, 2} and {4, 3, 5}, so we do it as illustrated in Figure A.1.