Supporting Information

Dihedral-based segment identification and classification of biopolymers II: Nucleotides

Gabor Nagy, Chris Oostenbrink[*]

University of Natural Resources and Life Sciences, Institute for Molecular Modeling and

Simulation

Muthgasse 18, 1190 Vienna, Austria

[*]: corresponding author: chris.oostenbrink@boku.ac.at

Table S1: Definitions for DISICL polynucleotide classification. Definitions of the regions in Figure 1 (on the left side) in terms of the (ε,ζ,χ) dihedral angles of the polynucleotide backbone. Angles are shown in degrees, and rectangles with identical names are merged together to follow the non-rectangular nature of the density distributions. Subregions are corrections for overlaps, and overwrite definitions in their volume.

DISICL Polynucleotide Regions						
Region Definitions (0 - 360)						
Name	εmin	εmax	ζmin	ζmax	χmin	χmax
α1	175	225	250	310	170	235
α2	165	250	310	360	160	250
α2	145	230	0	40	180	270
α3	225	290	250	310	150	260
β1	145	225	240	305	215	290
β2	225	310	120	215	180	310
β3	160	225	150	240	215	280
ζ1	150	260	230	330	20	100
ζ3	135	250	30	120	20	105
ζ2	170	320	40	120	160	290
δ1	20	100	25	135	170	290
δ2	20	120	155	340	190	290
δ3	20	100	25	100	30	100
subregions:						
ab1	175	225	250	305	215	235
ab2	225	300	215	250	160	280
α1	225	250	250	285	170	215
β3	180	225	240	250	215	270

Table S2: Superclass mapping used to calculate the overall agreement between methods (Table 1). Abbreviations for simplified classes are found in Tables 3-4.

Methods	DISICL vs. X3DNA
A-Helix match	AH/A
B-Helix match	BH/B, IB/B
Trans. match	AB/TA, TR/TA
Overall Match	A-hel. match, B-hel. match Trans. Match, UC/UC

Table S3: Scaled match scores for the comparison of secondary structure classifications by DISICL and X3DNA on the DNA dataset. For both algorithms, the occurrence of each class is displayed. In the top half of the table, simplified DISICL classes are compared to X3DNA classes. In the lower half of the table, detailed DISICL classes are compared to X3DNA classes.

Class	X3DNA		A helix	B helix	TA trans.	Unclassified
DISICL	%		2.0	34.7	0.4	62.8
B-helix	35.6		0.9	48.3	31.3	46.6
Irregular B	21.7		1.5	43.3	15.1	48.7
A-helix	2.2		64.2	6.1	1.4	23.7
Irregular A	2.1		2.8	2.9	1.1	61.8
Z-helix	1.0		0.0	0.0	0.0	68.4
Quad-loop	3.6		0.0	4.8	0.8	83.1
AB trans.	8.8		1.3	13.9	5.6	49.1
Transitory	11.1		4.4	34.4	12.8	56.6
Unclassified	14.0		3.3	13.5	11.2	51.3

Class	X3DNA		A helix	B helix	TA trans.	Unclassified
DISICL	%		2.0	34.7	0.4	62.8
BI-helix	35.6		0.9	48.3	31.3	46.6
BII-helix	3.4		0.0	46.9	1.7	42.8
BIII-helix	2.4		0.1	43.0	1.4	53.0
B-loop	15.9		1.4	42.6	12.0	49.4
A-helix	2.2		64.2	6.1	1.4	23.7
A-loop	0.4		7.0	2.6	0.0	53.0
Z-helix	1.0		0.0	0.0	0.0	68.4
Quad-loop	3.6		0.0	4.8	0.8	83.1
Sharp-turns	3.1		0.1	5.7	2.2	41.0
Tetraloop B.	1.6		1.9	2.2	1.1	63.3
AB trans.	11.1		4.4	34.4	12.8	56.6
AB2 trans.	1.4		0.8	14.4	1.4	55.6
AZ trans.	0.3		0.0	5.4	0.0	61.4
ZB trans.	0.9		0.0	8.2	0.0	71.2
AD trans.	0.3		5.4	15.8	0.0	52.2
BD trans.	2.3		0.2	30.5	2.0	45.0
ZD trans.	0.6		0.0	3.4	0.0	53.6
Unclassified	14.0		3.3	13.5	11.2	51.3

Table S4: Pearson correlation matrices for the comparison of secondary structure classifications by DISICL and X3DNA algorithms. In the top half of the table, simplified DISICL classes are compared with X3DNA classes in the DNA dataset, while the bottom half of the table shows the same comparison between detailed DISICL classes and X3DNA classes.

Pearson Score DNA (Simple DISICL/X3DNA)				
DISICL/X3DNA	A-helix	B-helix	Trans.	Unclassified
B-helix	-0.015	0.387	0.032	0.238
Irreg.B	-0.008	0.262	0.018	0.196
A-helix	0.589	-0.004	0.005	0.017
Irreg.A	0.021	-0.011	0.004	0.083
Z-helix	-0.003	-0.012	-0.001	0.065
Quad-loop	-0.005	-0.009	0.001	0.156
Transitory	-0.001	0.028	0.01	0.124
AB trans.	0.012	0.139	0.025	0.171
Unclassified	0.004	0.033	0.018	0.168

Pearson Score DNA (detailed DISICL/X3DNA)				
DISICL/X3DNA	A-helix	B-helix	Trans.	Unclassified
BI-helix	-0.015	0.387	0.032	0.238
BII-helix	-0.005	0.112	0.004	0.063
BIII-helix	-0.003	0.085	0.005	0.072
B-loop	-0.006	0.218	0.018	0.17
A-helix	0.589	-0.004	0.005	0.017
A-loop	0.026	-0.005	-0.001	0.03
Z-helix	-0.003	-0.012	-0.001	0.065
Quad-loop	-0.005	-0.009	0.001	0.156
Sharp-turns	-0.004	-0.006	0.007	0.056
Tetraloop B.	0.011	-0.011	0.005	0.074
AB trans.	0.012	0.139	0.025	0.171
AB2 trans.	0.002	0.012	0.007	0.058
AZ trans.	-0.002	-0.002	-0.001	0.03
ZB trans.	-0.003	0	-0.001	0.066
AD trans.	0.016	0.007	-0.001	0.024
BD trans.	-0.002	0.053	0.008	0.056
ZD trans.	-0.002	-0.005	-0.001	0.035
Unclassified	0.004	0.033	0.018	0.168

Table S5: Scaled match scores for the comparison of secondary structure classifications by DISICL and X3DNA on the RNA dataset. For both algorithms, the occurrence of each class is displayed. In the top half of the table, simplified DISICL classes are compared to X3DNA classes. In the lower half of the table, detailed DISICL classes are compared to X3DNA classes.

Class	X3DNA		A helix	B helix	TA trans.	Unclassified
DISICL	%		56.7	0.02	1.2	42.1
B-helix	0.4		2.3	0.0	10.2	42.4
Irregular B	1.0		10.6	5.9	0.4	42.2
A-helix	51.3		66.3	11.8	25.7	36.2
Irregular A	16.0		30.9	11.8	8.6	43.2
Z-helix	0.4		1.5	0.0	0.0	22.0
Quad-loop	0.4		7.8	2.9	0.0	40.0
AB trans.	8.0		15.7	23.5	1.0	42.4
Transitory	6.6		33.2	11.8	37.7	44.0
Unclassified	16.0		15.6	11.8	9.5	39.7

Class	X3DNA		A helix	B helix	TA trans.	Unclassified
DISICL	%		56.7	0.02	1.2	42.1
BI-helix	0.4		2.3	0.0	10.2	42.4
BII-helix	0.1		3.0	0.0	0.0	29.4
BIII-helix	0.1		0.0	0.0	0.0	30.4
B-loop	0.8		12.6	5.9	0.5	45.0
A-helix	51.3		66.3	11.8	25.7	36.2
A-loop	7.1		38.8	2.9	5.1	41.6
Z-helix	0.4		1.5	0.0	0.0	22.0
Quad-loop	0.4		7.8	2.9	0.0	40.0
Sharp-turns	2.1		7.5	5.9	0.2	32.4
Tetraloop B.	8.6		25.0	8.8	3.4	44.8
AB trans.	6.6		33.2	11.8	37.7	44.0
AB2 trans.	3.2		19.8	11.8	0.5	43.5
AZ trans.	1.0		4.9	0.0	0.1	65.3
ZB trans.	0.4		8.3	0.0	0.4	31.3
AD trans.	1.0		38.2	0.0	0.1	39.4
BD trans.	0.4		4.8	5.9	0.5	46.7
ZD trans.	0.2		9.4	0.0	0.0	36.1
Unclassified	16.0		15.6	11.8	9.5	39.7

Table S6: Pearson correlation matrices for the comparison of secondary structure classifications by DISICL and X3DNA algorithms. In the top half of the table simplified DISICL classes are compared with X3DNA classes in the RNA dataset, while the bottom half of the table shows the same comparison between detailed DISICL classes and X3DNA classes.

Pearson Score RNA (Simple DISICL/X3DNA)				
DISICL/X3DNA	A-helix	B-helix	Trans.	Unclassified
B-helix	-0.012	0	0.057	0.032
Irreg.B	-0.007	0.008	0.001	0.052
A-helix	0.593	-0.001	0.017	0.212
Irreg.A	0.095	0.003	0.01	0.221
Z-helix	-0.014	0	-0.002	0.012
Quad-loop	-0.007	0.007	-0.002	0.029
Transitory	0.001	0.011	-0.005	0.151
AB trans.	0.069	0.006	0.149	0.144
Unclassified	0.001	0.003	0.013	0.197

Pearson Score RNA (detailed DISICL/X3DNA)				
DISICL/X3DNA	A-helix	B-helix	Trans.	Unclassified
BI-helix	-0.012	0	0.057	0.032
BII-helix	-0.007	0	-0.001	0.011
BIII-helix	-0.006	0	-0.001	0.008
B-loop	-0.004	0.009	0.002	0.051
A-helix	0.593	-0.001	0.017	0.212
A-loop	0.094	0	0.012	0.138
Z-helix	-0.014	0	-0.002	0.012
Quad-loop	-0.007	0.007	-0.002	0.029
Sharp-turns	-0.017	0.005	-0.003	0.052
Tetraloop B.	0.043	0.003	0.003	0.169
AB trans.	0.069	0.006	0.149	0.144
AB2 trans.	0.012	0.009	-0.003	0.098
AZ trans.	-0.016	0	-0.002	0.09
ZB trans.	-0.007	0	0.001	0.022
AD trans.	0.034	0	-0.002	0.047
BD trans.	-0.01	0.014	0.001	0.036
ZD trans.	-0.004	0	-0.001	0.017
Unclassified	0.001	0.003	0.013	0.197