Supporting Information

Towards biochemically relevant QM computations on nucleic acids. Controlled electronic structure geometry optimizations of nucleic acids structural motifs using penalty restraint functions.

Holger Kruse[1,*] and Jiri Sponer[1,2*]
[1] CEITEC – Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic.
[2] Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic.

*Corresponding authors: kruse@ceitec.muni.cz, sponer@ncbr.muni.cz

Table S1. χ angle data for the UpA dinucleotide optimization. Cf. Table 1 in the main text for the remaining angles.

base	X-ray	unrestrained (k_r=0)	k_r=0.01	k_r=0.05	k_r=0.1	k_r=10.0
U	213	197	222	208	208	210
A	224	238	192	222	223	225

The χ angle in Table S1 is not restrained during the optimization so it is only influenced indirectly by the restraints of the δ to δ+1 backbone angles. The indirect restraint of 0.05 E_h/rad^2 keeps the χ angle already very close to the starting values. It is a safe assumption that restraining the χ angle along with the backbone angles using kr=0.01 E_h/rad^2 would also lead to reasonable values of χ angles. If necessary, the χ angles can be easily controlled by restraints, too. It should be noted that even for the high-resolution X-ray structures, uncertainty of the χ angles of ~10° can be expected. None of the computations gave a χ angle which would be outside a typical range of RNA χ angles; the most populated RNA χ angles lie in the range 180° - 220°.

Table S2. UpA optimizations (equivalent to the gas phase Table 1 data) using the COSMO implicit solvent model. While in this particular case even the unrestrained optimization remained rather close to the target structure, we in general cannot rely that the continuum solvent is always sufficient to prevent large artifacts.

	X-ray	unrestrained (k_r=0)	k_r=0.01	k_r=0.05	k_r=0.1	k_r=10.0
δ	81.5	80.1	80.3	83.3	81.5	81.5
ε	199.4	200.5	197.0	201.1	199.7	199.4
ζ	61.2	71.2	66.1	65.1	62.6	61.3
α+1	168.6	171.7	167.9	169.2	168.0	168.6
β+1	141.6	153.3	143.7	144.4	142.3	141.6
γ+1	48.0	35.8	44.5	45.7	46.8	47.8
δ+1	151.3	151.8	147.7	150.4	149.8	151.3
ΔE (=ΔE_{step}[a])	-	0.0	0.5	3.5	3.6	4.1
$E_{R(p)}$	-	-	0.1	0.5	0.1	0.0

$d(G_2)_4$ quadruplex section

Table S3. Extension of the Table 2 in the main text - the remaining backbone angles and interatomic distances data.

	X-ray[a]	unrestrained k_r=0.0[a]	all angles k_r= 0.01	all angles k_r= 0.05	β only k_r=0.05	β+χ only k_r= 0.01	β+χ only k_r=0.05
				strand 2			
δ	137	151	145	142	146	147	148
ε	190	163	182	188	179	173	179
ζ	261	276	264	263	263	264	262
α+1	286	279	295	293	295	295	196
β+1	180	239	188	184	183	193	184
γ+1	38	46	45	41	50	49	49
δ+1	137	149	139	137	136	143	140
χ	254	227	249	252	253	252	253
χ+1	256	262	251	254	244	250	254
r_{NO}	4.6	3.1	4.4	4.5	4.5	4.3	4.4
				strand 3			
δ	144	151	147	146	146	146	146
ε	188	164	182	187	181	165	166
ζ	262	276	264	164	264	270	269
α+1	290	279	291	291	293	290	290
β+1	180	239	186	181	183	208	204
γ+1	45	46	48	48	49	49	49
δ+1	122	149	132	125	136	145	144
χ	253	227	251	252	250	250	253
χ+1	247	262	246	246	246	251	249
r_{NO}	4.7	3.1	4.4	4.6	4.5	4.2	4.2
				strand 4			
δ	139	151	145	143	146	147	147
ε	179	164	177	178	177	172	175
ζ	260	276	164	263	264	263	262
α+1	290	279	290	290	293	293	293
β+1	289	240	192	188	191	195	191
γ+1	54	46	50	52	49	49	50
δ+1	129	149	137	132	141	142	141
χ	254	226	251	253	250	253	253
χ+1	245	262	247	245	250	248	246
r_{NO}	4.7	3.1	4.4	4.6	4.5	4.3	4.4

[a] data taken from ref.[1]

$d(ACGT)_2$ section

Note, that the d(ACGT)$_2$ X-ray structure has a true crystallographic two-fold symmetry, so that its two experimental strands are identical. However, the QM optimization was carried out without any symmetry restrictions, leading to asymmetric local minimum structures.

Table S4. Extension of Table 3 of the main text - the remaining backbone angles and interatomic distances data.

	X-ray	unrestrained k_r=0.0	k_r=0.01	k_r=0.05	k_r=0.1	k_r=10.0
	CpG step, strand 1					
δ	106	114	117	110	106	106
ε	188	115	172	182	187	188
ζ	248	300	245	243	249	248
α+1	311	266	307	309	311	311
β+1	172	256	188	180	175	172
γ+1	46	49	47	46	48	46
δ+1	141	144	143	142	140	141
	GpT step, strand 1					
δ	141	144	143	142	140	141
ε	184	170	181	193	181	184
ζ	265	272	269	267	265	265
α+1	290	295	290	289	290	290
β+1	178	178	181	178	176	178
γ+1	59	50	53	56	58	59
δ+1	108	105	111	108	108	108
	ApC step, strand 2					
δ	124	145	141	128	127	124
ε	198	184	190	195	196	198
ζ	242	278	265	248	244	242
α+1	311	292	306	311	311	311
β+1	147	173	158	148	148	147
γ+1	56	46	53	55	56	56
δ+1	106	117	117	110	107	106
	CpG step, strand 2					
δ	106	117	117	110	107	106
ε	188	112	176	182	186	188
ζ	248	297	144	243	249	248
α+1	311	263	306	109	311	311
β+1	172	253	158	180	174	172
γ+1	46	54	53	46	48	46
δ+1	141	143	142	141	140	141
	GpT, strand 2					
δ	141	143	142	141	140	141
ε	184	132	184	184	181	184
ζ	265	289	272	267	265	265
α+1	290	291	191	289	290	290
β+1	178	223	176	177	176	178
γ+1	59	48	53	56	58	59
δ+1	108	145	108	108	107	108

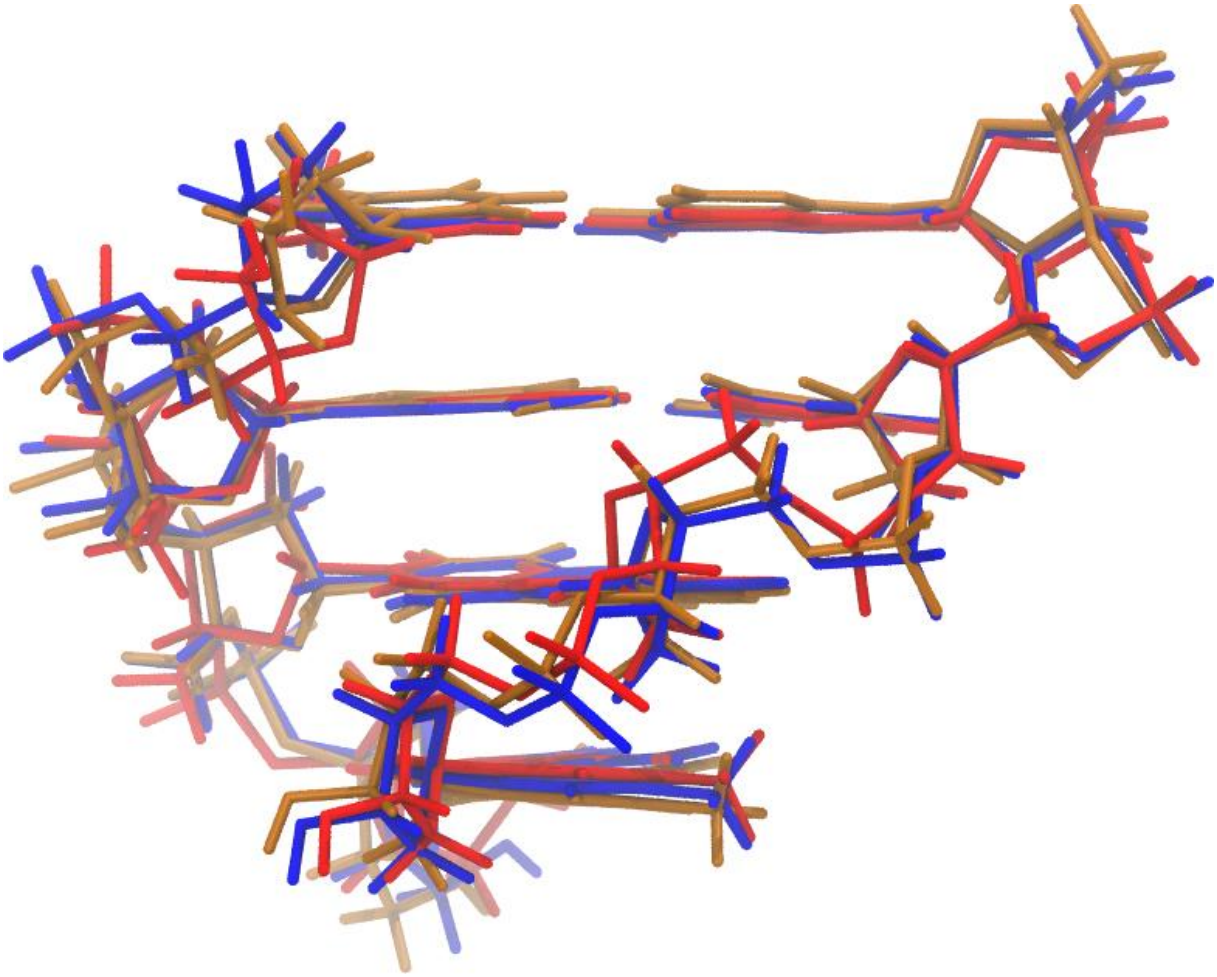

Figure S1. Color-coded side view of the X-ray (orange), unrestrained (red) and restrained (blue, kr=0.05) structures of the d(ACGT)2 B-DNA system. The restrained optimization structure clearly shows improvements for the backbone topology.

Helical parameters - the base pairs and base-pair steps. The parameter analysis shown in Table S4 was done using the 3DNA code.[2] The unrestrained optimization leads to an underestimation of the propeller twist, which is only partly improved using the backbone angle restraints. We expect further improvement if we additionally restrain the χ angles. At this point it is unclear if the underestimated propeller twist is an artefact of the HF-3c method or if this particular property is just very sensitive to the level of theory in general. The unrestrained structure further shows a reduced twist (helicity). From all restrained optimizations, the restraint with k_r=0.05 E_h/rad^2 yields the twist parameter closest to the X-ray structure. The stronger restraints (k_r>0.05 E_h/rad^2) seem to have an over-twisting tendency. The buckle and inclination show an irregular behaviour among all optimizations and no clear tendencies regarding different restraining constants. However, the middle-strong restraint with kr=0.05 E_h/rad^2 gives reasonable values for the Inclinations. The deviations in the buckle are particularly strong in the terminal bases, suggesting that the missing molecular environment might be the main cause of the error in this case ('terminal effects'). It is, however, clear that a more thorough study is needed to understand all structural deviations and parameters in this system, as well as the exact influences of the restraints. As pointed in the main text, quantitative description of details of B-DNA structure is difficult and is hampered by the lack of unambiguous experimental data. The B-DNA is a highly flexible molecule with sequence-dependent structural dynamics heterogeneity. The atomistic X-ray structures are always affected by averaging, resolution limits and crystal

packing. Therefore, it is presently not known if B-DNA optimizations should be restrained (and how tightly) according to the individual X-ray structures or rather by using some averaged or idealized structural parameters.

Table S5. Local base-pair, base-pair step and helical parameters of the d(ACGT)$_2$ structures obtained using the standard 3DNA code.

X-ray

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.06	-0.11	0.29	7.25	-12.05	3.98
2 C-G	0.28	-0.16	-0.10	10.51	-13.26	-0.04
3 G-C	-0.28	-0.16	-0.10	-10.51	-13.26	-0.04
4 T-A	-0.06	-0.11	0.29	-7.25	-12.05	3.98
ave.	0.00	-0.14	0.09	0.00	-12.65	1.97
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	0.29	-0.36	3.27	2.75	-1.25	31.36
2 CG/CG	0.00	-0.02	3.79	0.00	7.52	42.15
3 GT/AC	-0.29	-0.36	3.27	-2.75	-1.25	31.36
ave.	0.00	-0.25	3.44	0.00	1.67	34.95
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	-0.44	-0.03	3.29	-2.30	-5.07	31.50
2 CG/CG	-0.90	0.00	3.73	10.36	0.00	42.78
3 GT/AC	-0.44	0.03	3.29	-2.30	5.07	31.50
ave.	-0.59	0.00	3.44	1.92	0.00	35.26

unrestrained

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.10	-0.14	0.33	7.22	-9.42	-2.03
2 C-G	0.07	-0.17	-0.09	8.55	-4.09	-1.03
3 G-C	-0.10	-0.17	-0.14	-7.66	-4.66	-1.15
4 T-A	-0.02	-0.14	0.09	0.06	-5.61	0.47
ave.	0.01	-0.16	0.05	2.04	-5.94	-0.93
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	1.00	-0.33	3.16	4.30	-3.31	31.55
2 CG/CG	-0.11	-0.40	3.42	0.24	4.50	41.62
3 GT/AC	-0.29	-0.24	2.99	-2.05	-0.35	27.01
ave.	0.20	-0.33	3.19	0.83	0.28	33.39
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	-0.01	-1.04	3.28	-6.04	-7.84	32.00
2 CG/CG	-1.05	0.18	3.36	6.32	-0.33	41.86
3 GT/AC	-0.44	0.15	3.01	-0.75	4.38	27.08
ave.	-0.50	-0.24	3.22	-0.16	-1.27	33.65

restraint k_r=0.01

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.04	-0.15	0.00	-2.88	-4.48	0.20
2 C-G	0.07	-0.16	-0.10	8.19	-4.90	-1.22

3 G-C	-0.04	-0.14	-0.15	-12.10	-7.50	-0.60
4 T-A	-0.09	-0.14	0.09	-3.59	-6.36	-0.83
ave.	-0.01	-0.15	-0.04	-2.59	-5.81	-0.61
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	0.41	-0.35	2.94	0.68	-1.35	26.94
2 CG/CG	0.15	-0.15	3.48	1.26	6.31	46.39
3 GT/AC	-0.68	-0.48	2.98	-2.04	-1.75	29.09
ave.	-0.04	-0.33	3.13	-0.03	1.07	34.14
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	-0.45	-0.73	2.96	-2.89	-1.46	26.98
2 CG/CG	-0.75	-0.08	3.44	7.97	-1.59	46.81
3 GT/AC	-0.60	0.94	3.05	-3.48	4.05	29.21
ave.	-0.60	0.04	3.15	0.53	0.33	34.33

restraint k_r=0.05

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.08	-0.15	0.04	-0.88	-8.40	-0.52
2 C-G	0.04	-0.15	-0.10	6.75	-7.04	-0.83
3 G-C	-0.05	-0.15	-0.10	-7.70	-6.56	-0.78
4 T-A	-0.08	-0.15	0.04	0.76	-7.25	-0.19
ave.	0.00	-0.15	-0.03	-0.27	-7.31	-0.58
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	0.53	-0.37	2.99	1.07	-1.74	29.26
2 CG/CG	0.01	0.02	3.40	-0.20	7.80	46.20
3 GT/AC	-0.50	-0.39	2.98	-1.06	-1.76	29.58
ave.	0.02	-0.25	3.12	-0.06	1.43	35.01
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	-0.39	-0.83	3.02	-3.44	-2.12	29.33
2 CG/CG	-0.65	-0.03	3.36	9.85	0.25	46.82
3 GT/AC	-0.42	0.76	3.01	-3.45	2.08	29.65
ave.	-0.48	-0.03	3.13	0.99	0.07	35.26

restraint k_r=0.1

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.03	-0.15	0.02	-4.48	-7.60	0.79
2 C-G	0.06	-0.17	-0.03	0.28	-12.44	-1.26
3 G-C	-0.06	-0.18	0.00	0.83	-11.98	-1.43
4 T-A	-0.03	-0.14	0.02	6.43	-7.02	1.05
ave.	0.00	-0.16	0.00	0.77	-9.76	-0.21
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	-0.04	-0.01	3.07	1.01	-3.69	32.89
2 CG/CG	-0.03	0.81	3.36	-0.12	4.97	42.44
3 GT/AC	0.08	0.05	3.06	-0.77	-3.72	32.66
ave.	0.00	0.29	3.16	0.04	-0.81	36.00
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	0.57	0.23	3.05	-6.48	-1.77	33.10

2 CG/CG	0.60	0.03	3.43	6.84	0.16	42.72
3 GT/AC	0.69	-0.26	3.03	-6.59	1.36	32.87
ave.	0.62	0.00	3.17	-2.08	-0.08	36.23

restraint k_r=10.0

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 A-T	0.02	-0.14	-0.01	-7.09	-6.49	2.12
2 C-G	0.07	-0.17	-0.01	-0.08	-12.60	-1.36
3 G-C	-0.07	-0.17	-0.01	-0.99	-12.35	-1.57
4 T-A	-0.02	-0.14	0.00	7.29	-4.98	1.80
ave.	0.00	-0.16	-0.01	-0.22	-9.10	0.25
Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 AC/GT	-0.17	-0.05	3.03	0.11	-1.22	32.36
2 CG/CG	-0.03	0.72	3.34	-0.08	5.82	41.66
3 GT/AC	0.17	-0.03	3.01	-0.33	-1.62	32.10
ave.	-0.01	0.21	3.13	-0.10	1.00	35.38
Step	X-disp	Y-disp	h-Rise	Incl.	Tip	h-Twist
1 AC/GT	0.11	0.32	3.03	-2.19	-0.19	32.38
2 CG/CG	0.38	0.04	3.41	8.14	0.11	42.05
3 GT/AC	0.22	-0.36	3.01	-2.92	0.59	32.15
ave.	0.23	0.00	3.15	1.01	0.17	35.53

Structure data

see the provided zip file or http://dx.doi.org/10.6084/m9.figshare.1233996

Source code

see the provided zip file or request an up-to-date version from the authors.
The code is also shared on github.com: https://github.com/hokru/xopt

References

1. Šponer, J.; Mládek, A.; Špačková, N.; Cang, X.; Cheatham, T. E.; Grimme, S., Relative Stability of Different DNA Guanine Quadruplex Stem Topologies Derived Using Large-Scale Quantum-Chemical Computations. *Journal of the American Chemical Society* **2013,** *135* (26), 9785-9796.
2. Lu, X. J.; Olson, W. K., 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003,** *31* (17), 5108-5121.