Nucleic Acids Research, 1992, Vol. 20, No. 24 6535–6541

Prediction of the structure of the $Y^+ \cdot R^- \cdot R^+$-type DNA triple helix by molecular modelling

C.A.Laughton and S.Neidle
CRC Biomolecular Structure Unit, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK

Received September 25, 1992; Revised and Accepted October 30, 1992

ABSTRACT

Molecular mechanics has been used to predict the structure of the $Y^+ \cdot R^- \cdot R^+$-type DNA triple helix, in which a second polypurine strand binds antiparallel to the homopurine strand of a homopurine/homopyrimidine stretch of duplex DNA. From calculations on the sequence $d(C)_{10}.d(G)_{10}.d(G)_{10}$, two likely structures emerge. One has the glycosidic torsions of the third strand bases in the *anti*conformation and Hoogsteen hydrogen-bonds to the purine strand of the duplex, the other has the third strand purines in the *syn* orientation and uses a reverse-Hoogsteen hydrogen-bonding pattern. Despite the large structural differences between these two types of triplex, calculations performed *in vacuo* with a distance-dependent dielectric constant to mimic the shielding effect of solvent show them to be energetically very similar, with the latter (*syn*) slightly preferred. However, if explicit solvent molecules are included in the calculation, the *anti* conformation is found to be much preferred. This difference in the results seems to stem from an underestimation of short-range electrostatic interactions in the *in vacuo* simulations. When TAA or TAT base triples are substituted for the sixth CGG triple in the sequence, it is found that, for the solvated model, the third strand base of the TAA triple prefers the *syn* orientation while that in the TAT triple retains a preference, though reduced, for the *anti* conformation.

INTRODUCTION

The interaction of an oligonucleotide with duplex DNA to form triple-helical structures is a phenomenon which shows great promise as a method of achieving sequence-specific DNA recognition and hence artificial gene regulation [for a comprehensive list of references, see (1) and (2)]. To date two general classes of DNA triple helix have been characterised. Both consist of three strands of alternating polarity, so strand 2 runs antiparallel to strands 1 and 3. In both types strands 1 and 2 may be thought of as originating from the duplex and maintain their conventional Watson–Crick hydrogen-bonding pattern. In the first class of triplex, strand 1 is homopurine while strands 2 and 3 are homopyrimidine (3,4), in the second, strand 1 is homopyrimidine, while strands 2 and 3 are homopurine (5,6) (Figure 1). Taking strand polarity into account by a superscript, the two classes may therefore be represented as $R^+Y^-Y^+$ and $Y^+R^-R^+$ respectively.

To date the $R^+Y^-Y^+$ triplex has been more extensively studied although there is as yet no X-ray crystal structure for it. However, it has been established by fibre-diffraction (7), nmr (8–12), and other physical methods that the structure consists of a right-hand helix in which the three strands have backbone conformations reminiscent of A-type duplex DNA, and the pyrimidine bases of the third strand interact with the purines of the first strand through Hoogsteen hydrogen-bonding. In all three strands the glycosidic torsions of the bases have the *anti* conformation. Nevertheless, there is still much about the detailed structure of the $R^+Y^-Y^+$ triplex family which is not well established, such as the nature of the DNA sugar puckers (13).

The amount of structural information on the $Y^+R^-R^+$ type of triplex is much more limited, but a basic similarity to the $R^+Y^-Y^+$ structure seems evident (14). Affinity-cleaving methods (6) have confirmed the polarity of the third (purine) strand as R^+. Two possible conformations have been hypothesised: one in which the purines of the third strand have an *anti* conformation and bind with reverse-Hoogsteen hydrogen-bonds to the second-strand purines, and another in which the third strand purines have a *syn* conformation and bind to the second strand using Hoogsteen hydrogen-bonds (6,15–17).

In this study we have used molecular modelling to predict if and how such structures might be formed in an energetically feasible manner, and examine them in the light of the existing experimental data. In particular we are interested in the electrostatic, steric and conformational characteristics of the $Y^+R^-R^+$ type of triplex, and how it may be combined with the $R^+Y^-Y^+$ type to extend the range of DNA sequences which can be recognised by triplex formation (1,18). Molecular mechanics calculations on this type of triplex have been reported by other workers (16,19), but using a different force-field and with a different treatment of electrostatic interactions. As shown below, we find that the results obtained depend critically on how electrostatic and solvent-associated effects are taken into account in the modelling procedure.

METHODS

General

All molecular mechanics and dynamics calculations were performed using the AMBER 4.0 suite of programs (20) on an Alliant FX40/3 computer in-house or on a Convex C3840 at the University of London Computing centre. Visualisations and interactive modelling were performed using the programs GEMINI 1.03 (21) and MidasPlus (22) on a Silicon Graphics Iris 4D-20 workstation.

Construction of the triplexes

The basic sequence used in this study was $d(C)_{10}.d(G)_{10}.d(G)_{10}$. The initial structure for strands 1 and 2 was taken from the fibre diffraction-derived model for the $R^+Y^-Y^+$ triplex (7). Strand 3 was built by hand, with the glycosidic torsions of the bases in either the *anti* or *syn* conformations, resulting in several starting models. All had the third strand parallel to the cytidine strand. When counterions were included, these were not added using the standard AMBER methodology, but by the procedure that we have previously described (23).

Minimisation and dynamics protocols

All energy minimisations and dynamics were performed using an 8.0 Å non-bonded cutoff and a pairlist which was updated after every 20 steps. In simulations without explicit water, a distance-dependent dielectric constant, $\epsilon=4r_{ij}$, was used (24); in simulations including water a constant dielectric of 1 was used. Energy minimisations were terminated at an r.m.s. gradient of 0.1 kcal/mol/Å. The first stage in the minimisation of the triplex structures was restricted to the backbone atoms of the third strand in order to remove bad interactions remaining from the interactive model-building procedure. In the second stage, torsion restraints were used to force the backbone of the third strand into an A- or B-type conformation (Table 1).

In the third stage, the whole system was allowed to relax according to the protocol shown in the flow chart (Figure 2). The average values for the DNA backbone torsion angles α-ζ and χ were calculated for each strand independently. These values were then used as restraints for the next cycle of minimisation, of a maximum of 100 steps. The new average torsion angles were then recalculated and the cycle repeated until the minimisation terminated at step 1. Typically, 15–20 such cycles were required to achieve convergence. Using this protocol the minimisation proceeded with the retention of approximate helical symmetry.

For those simulations performed with explicit solvent, after the addition of counterions the 'blob' solvation option of AMBER was used to surround the structures with a 5Å shell of Monte-Carlo TIP3P water (approx. 730 molecules). The water structure was conditioned by energy minimisation, 10 ps of dynamics at 100 K, and then reminimisation before the whole system was energy minimised according to the protocol described above. Molecular dynamics was performed with SHAKE on all bonds and a 2.0 fs time-step.

```
                    5'-RRRRRRRRRRRRRR-3'
R+Y-Y+ triplex:     3'-YYYYYYYYYYYYYY-5'
                    5'-YYYYYYYYYYYYYY-3'

                    5'-YYYYYYYYYYYYYY-3'
Y+R-R+ triplex:     3'-RRRRRRRRRRRRRR-5'
                    5'-RRRRRRRRRRRRRR-3'

                    5'-CCCCCXCCCC-3'
This study:         3'-GGGGGYGGGG-5'
                    5'-GGGGGZGGGG-3'

                XYZ = CGG,TAA or TAT
```

Figure 1. Topology of RYY and YRR triplexes, and structure of the triplexes modelled in this study.

RESULTS

Energy-minimized models for the $Y^+ \cdot R^- \cdot R^+$ triplex structure

In producing models for the YRR triplex structure we have begun with the assumption that the conformation of the duplex-derived strands is not far removed from that observed in the RYY triplex. This is reasonable, based on the observation that the two classes of triplex can be formed in contiguous sections of DNA (1,18)—it is unlikely that, for instance, the YRR triple helix is left-handed. Having manually positioned the third strand bases to produce the required Hoogsteen or reverse-Hoogsteen hydrogen-bonding patterns to strand 2, depending on whether the *syn* or *anti* conformations were being investigated (Figure 3), a suitable backbone conformation had to be established to link the bases in an energetically reasonable manner while maintaining the helical twist and rise.

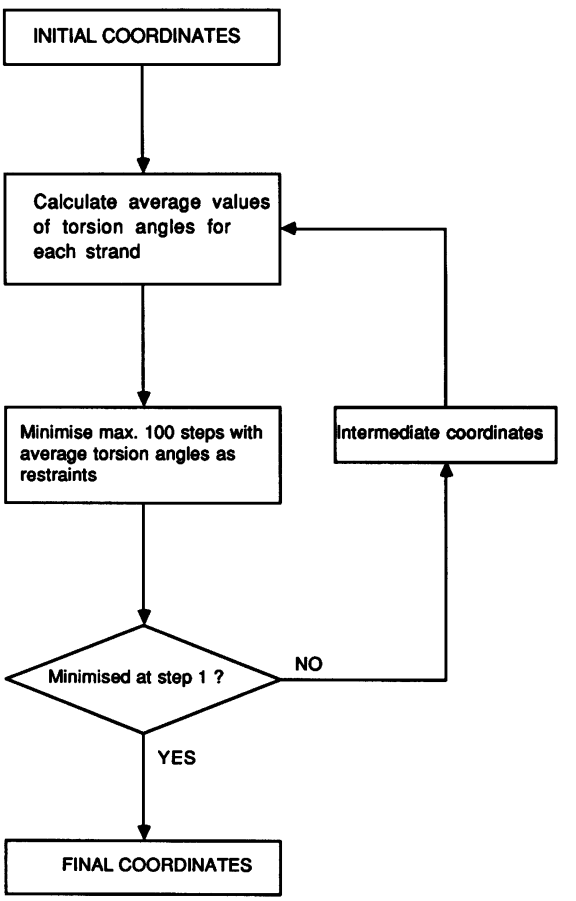

Figure 2. Flow chart of the energy minimisation protocol.

Initially we used a torsional search procedure to generate a number of possible backbone conformations, but found that, although there were several possible solutions in either case, none was as energetically favourable as a backbone in an approximately A- or B-type conformation (results not shown). Therefore only two starting conformations (A- or B- like) for each type of triplex (*syn* or *anti*) were investigated in detail (Table 1). For convenience these will be described as the A*s*, B*s*, A*a* and B*a* models.

Table 1. Initial values (in degrees) of backbone torsion angles α-ζ and χ, and sugar pseudorotations, P, in $Y^+ \cdot R^- \cdot R^+$ triplex models.

Strand	α	β	γ	δ	ϵ	ζ	χ	P
1	−73	175	60	79	−167	−66	−148	29
2	−76	174	64	80	−168	−64	−159	26
3A*a*	−75	170	67	86	−176	−73	−122	27
3B*a*	−38	137	42	130	−144	−155	−106	110
3A*s*	−74	172	66	77	−166	−61	22	52
3B*s*	−33	140	39	123	−144	−137	44	100

The results of the energy minimisations are shown in Table 2. This lists the average energy components per base triplet, including their 5′-phosphate groups. The first and last triplets are excluded, as not all bases have a 5′-phosphate attached, and the second and penultimate triplets have also been excluded to reduce end-effects. It can be seen that the B-type backbone conformation is preferred over the A-type, and model B*s* over B*a*; however, the energy differences are small.

The associated backbone parameters are shown in Table 3. We observe that for all models, the conformations of strands 1 and 2 show little change from their initial values. Model A*a* has a reduced sugar pseudorotation angle for the third strand, whereas in model B*a* it somewhat increased and there are also changes in torsions α and γ. Model A*s* shows very little change in backbone conformation, but minimisation of the B*s* model produces major changes in most backbone torsion angles.

Figure 4 shows a space-filling representation of model B*a* (right) and, for comparison, the structure of the $R^+Y^-Y^+$ triplex $d(A)_{10}.d(T)_{10}.d(T)_{10}$ (7,23) (left). Examination of the models shows that the major difference in the gross structure of these two classes of triplex is associated with a shift in strand

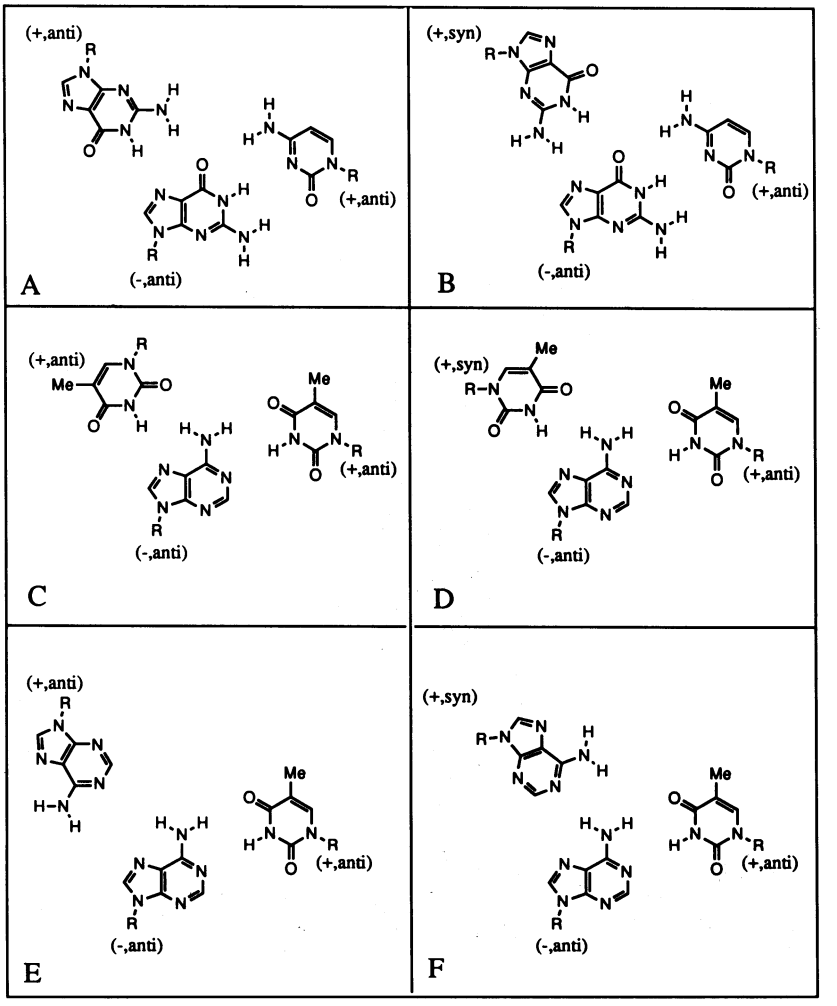

Figure 3. Hydrogen-bonding schemes in YRR triplexes. Top: CGG triples; middle: TAT triples; bottom: TAA triples. In each case the *anti* conformation is on the left, and the *syn* on the right.

Table 2. Energy breakdowns (per base triplet) for the different *in vacuo* $Y^+ \cdot R^- \cdot R^+$ triplex models.

Model	E_{bonded}	E_{nb}	E_q	E_{total}
A*a*	42.8	−95.4	−14.9	−64.9
B*a*	46.7	−96.4	−18.8	−68.5
A*s*	43.3	−100.0	−8.1	−64.9
B*s*	41.8	−97.4	−14.5	−70.0

Values are in kcal/mol. E_{bonded} = sum of bond, angle and dihedral energies, E_{nb} = non-bonded (including hydrogen-bond) energy, E_q = electrostatic energy.

Table 3. Averaged backbone torsion angles α-ζ and χ, and sugar pseudorotations, *P* (all in degrees), in final, unhydrated, $Y^+ \cdot R^- \cdot R^+$ triplex models.

Strand	α	β	γ	δ	ϵ	ζ	χ	*P*
A*a*:								
1	−74	175	60	79	−167	−66	−148	30
2	−76	174	64	80	−168	−64	−158	26
3	−73	171	68	91	−180	−76	−115	13
B*a*:								
1	−74	174	60	79	−167	−66	−149	30
2	−76	174	63	80	−169	−63	−158	30
3	−52	138	59	121	−145	−146	−103	117
A*s*:								
1	−73	173	61	79	−168	−66	−147	35
2	−77	175	64	80	−169	−66	−159	22
3	−76	172	67	76	−168	−60	31	52
B*s*:								
1	−74	173	61	79	−168	−66	−147	36
2	−77	175	63	79	−169	−65	−160	22
3	−66	174	55	106	−174	−83	55	108

3 across the groove between strands 1 and 2, which is analogous to the major groove in an A-DNA duplex. The result of this shift is to place the phosphate groups of strand 3 much closer to those of strand 2 in the $Y^+R^-R^+$ triplex compared to the $R^+Y^-Y^+$ one. The relative sizes of the two grooves either side of strand 3 are reversed, so the narrow and deep groove of highly negative electrostatic potential in the $R^+Y^-Y^+$ triplex is lost and instead we observe a deep, but less electronegative, groove between strands 1 and 3 (Figure 5).

The results of including an explicit treatment of solvent

Because of the potential importance of electrostatic and solvation effects on the relative stabilities of these highly charged structures, each was re-minimised in the presence of counterions and a 5Å shell of water molecules. The energy components are shown in Table 4, and the results of the helix analyses in Table 5. The major change in the energy components is in the electrostatic term. All models have very similar bonded and non-bonded energy components, but model B*a* is observed to be the most stable because of a particularly favourable electrostatic energy. Interestingly, a more detailed analysis of the energy terms reveals that this improved electrostatic energy term results not from superior interactions with counterions or solvent in the B*a* model, but from a superior inter-triplet interaction energy. A major contribution to this comes from inter-strand phosphate−phosphate interactions, as is confirmed by a strand-based energy analysis (Table 6). Examination of the backbone torsion angles shows that as before, strands 1 and 2 show little change in conformation. For strand 3 of model A*a* the reduction in sugar pseudorotation that accompanied the *in-vacuo* minimisations is enhanced. For

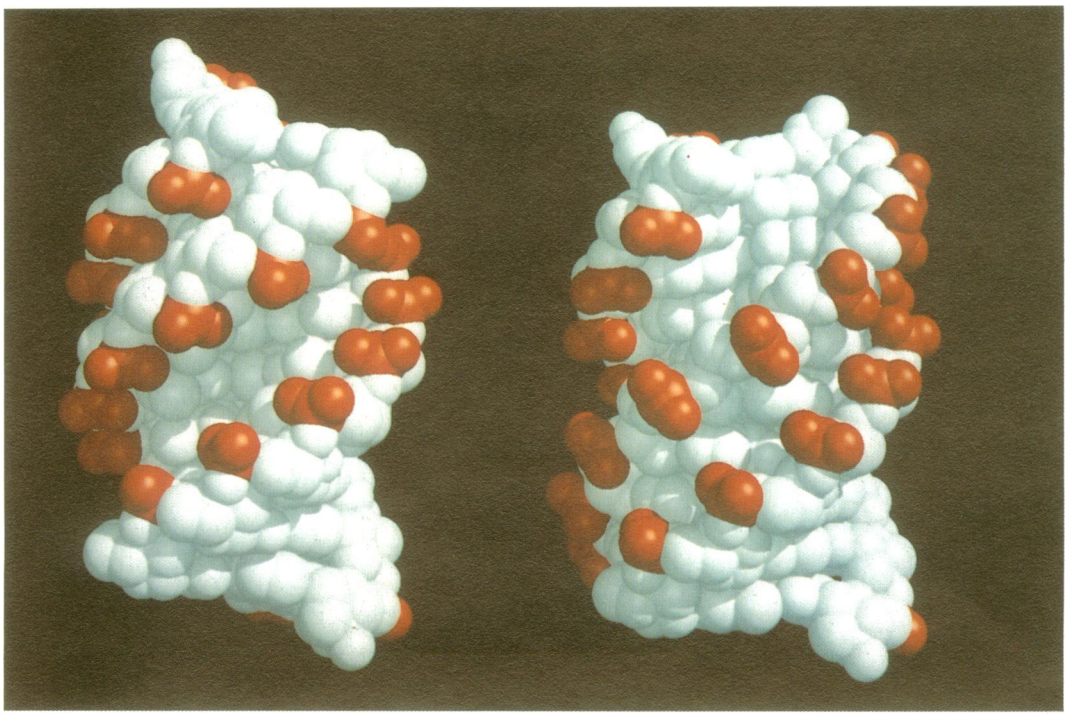

Figure 4. Space-filling representations of (left) the RYY triplex $d(A)_{10}.d(T)_{10}.d(T)_{10}$ from the fibre-diffraction model (7,22) and (right) the *in vacuo* minimised YRR triplex $d(C)_{10}.d(G)_{10}.d(G)_{10}$ in conformation B*a* from this study. The phosphate groups are coloured red.

model B*a* the sugar pseudorotation, rather than increasing, shows a large reduction, though the trends in the changes in α and γ are maintained. Whereas model A*s* showed little conformational change in the previous minimisations, it shows sizeable changes in both χ and *P* when minimised in the presence of solvent, while model B*s* shows little further change relative to its *in-vacuo* minimised structure. Despite these differences in detail, and the very different energies obtained, the overall structures of the triplexes minimised with explicit solvent do not differ greatly from their *in vacuo* counterparts (rms deviations A*a*: 0.34 Å, B*a*: 0.41 Å, A*s*: 0.51 Å, B*s*: 0.24 Å). In view of these small rms deviations, we do not present space-filling or electrostatic potential diagrams for the reminimised structures.

Effects on stability resulting from the introduction of TAA and TAT triples

In order to see how the different triplex models responded to the introduction of alternative base triples, the sixth triple in each minimised model was replaced by TAA and TAT and then re-minimised (without torsion restraints). The energy breakdowns for the resulting structures are shown in Tables 7 and 8. The structures are labelled according to their 'parent' conformations with the addition of a postfix identifying the modified third strand base, so structure B*a* gives rise to B*a*A and B*a*T, etc. The values in the tables refer to the modified triplet only, the average energy components for the remaining CGG triplets were almost identical to the values for the corresponding 'parent' structures. For the TAA triplet substitution, we see that the order of stability is A*a*A < A*s*A < B*a*A < B*s*A; while for the TAT substitution the order is A*a*T < A*s*T < B*s*T < B*a*T, though the energy differences are very small (total range 23 kcal/mol).

Table 4. Energy breakdowns (per base triplet) for the different hydrated $Y^+ \cdot R^- \cdot R^+$ triplex models

Model	E_{bonded}	E_{nb}	E_q	E_{total}
A*a*	51.4	−110.5	−223.5	−282.6
B*a*	59.9	−110.4	−282.5	−332.9
A*s*	54.3	−108.1	−224.7	−278.5
B*s*	52.2	−108.1	−214.6	−270.5

Values are in kcal/mol. For definitions of column titles, see Table 2.

DISCUSSION

The *ab initio* prediction of the molecular structure of a DNA triple helix is clearly a major undertaking. However it is greatly facilitated by the assumption, based on experimental evidence, that the structure is unlikely to be very different from that of the $R^+Y^-Y^+$ triplex. Thus we have started from the assumption that strands 1 and 2 of the $Y^+R^-R^+$ triplex adopt the same right-handed, A-DNA-like conformation, and have only considered the case of the third strand running antiparallel to strand 2 of the duplex. On this basis, two possible conformations and hydrogen-bonding schemes for the third strand bases are apparent, Hoogsteen hydrogen-bonding in the *anti* conformation or reverse-Hoogsteen hydrogen-bonding in the *syn* conformation. From a preliminary examination of the 2-dimensional structures of these possibilities (Figure 3), it seemed most likely that the Hoogsteen, *anti* mode (Figure 3A) would be preferable as this

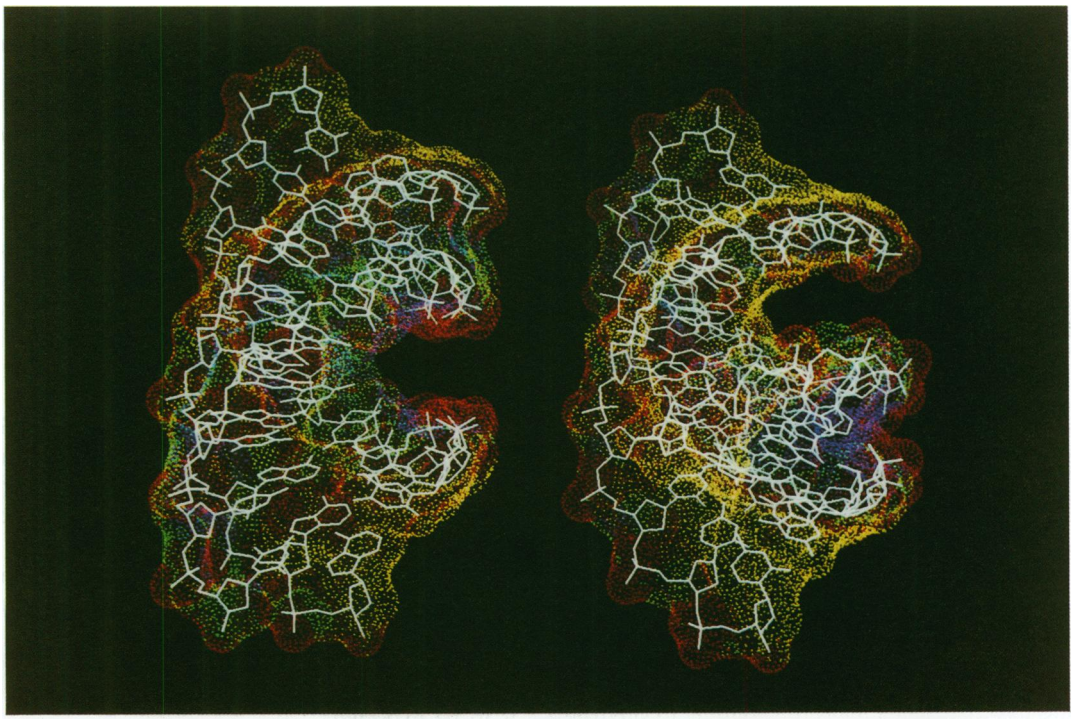

Figure 5. Dot-surfaces for the same triplex models shown in Figure 4, colour-coded by electrostatic potential (blue = most negative potential; yellow = most positive). The models have been reorientated to highlight the grooves in the structures.

Table 5. Averaged backbone torsion angles α-ζ and χ, and sugar pseudorotations, P (all in degrees), in final, hydrated, $Y^+ \cdot R^- \cdot R^+$ triplex models

Strand	α	β	γ	δ	ϵ	ζ	χ	P
A*a*:								
1	−68	169	58	77	−166	−67	−153	26
2	−78	172	65	82	−172	−61	−157	32
3	−71	168	67	92	−179	−76	−115	0
B*a*:								
1	−70	171	59	79	−167	−64	−152	26
2	−77	170	67	81	−168	−62	−157	31
3	−45	127	64	103	−147	−133	−114	82
A*s*:								
1	−66	165	60	78	−166	−67	−150	40
2	−79	171	66	85	−171	−68	−155	47
3	−69	167	60	89	−170	−70	43	83
B*s*:								
1	−68	167	61	78	−167	−66	−151	40
2	−80	173	67	80	−169	−63	−158	23
3	−66	171	56	102	−174	−79	53	106

Table 6. Inter-strand energy breakdowns for the $Y^+ \cdot R^- \cdot R^+$ triplex minimised in the presence of explicit solvent

	E_{nb}			E_q		
Model	1−2	2−3	1−3	1−2	2−3	1−3
A*a*	−71	−93	−7	210	1212	405
B*a*	−72	−95	−7	208	1117	424
A*s*	−69	−70	−20	193	1319	490
B*s*	−69	−69	−20	197	1325	487

Values are in kcal/mol.

Table 7. Energy breakdowns for the TAA triplet in the modified $Y^+R^- \cdot R^+$ triplex models (minimised in the presence of explicit solvent)

Model	E_{bonded}	E_{nb}	E_q	E_{total}
A*a*A	53.5	−102.7	−110.2	−159.4
B*a*A	60.1	−98.6	−153.5	−192.0
A*s*A	55.9	−100.7	−121.7	−166.5
B*s*A	51.2	−97.3	−166.8	−213.0

Values are in kcal/mol. For definitions of column titles, see Table 2.

Table 8. Energy breakdowns for the TAT triplet in the modified $Y^+ \cdot R^- \cdot R^+$ triplex models (minimised with explicit solvent)

Model	E_{bonded}	E_{nb}	E_q	E_{total}
A*a*T	58.9	−100.8	−109.2	−151.0
B*a*T	61.3	−97.9	−137.2	−173.7
A*s*T	55.4	−96.9	−112.9	−154.3
B*s*T	58.5	−86.8	−140.5	−168.9

Values are in kcal/mol. For definitions of column titles, see Table 2.

would place the sugar-phosphate backbone of the third strand more centrally in the groove between strands 1 and 2. We were therefore rather surprised when our initial calculations, performed *in vacuo* with a distance-dependent dielectric constant to simulate the shielding effect of solvent, indicated that *syn* and *anti* models had very similar energies (Table 2). However, a similar result has been observed previously (16,19).

Examination of the models revealed that in both cases, and for both A- and B-type backbone conformations, the structures had the phosphate groups of strands 2 and 3 in very close proximity. In view of this we suspected that electrostatic interactions might play a crucial role in the relative stabilities of these models and that a more rigorous treatment of this term might be necessary. The energetic analysis of the structures resulting from the minimisations performed with the explicit treatment of counterions and water, and a distance-independent dielectric constant of one, confirmed these suspicions. Model B*a*, with a B-type sugar-phosphate backbone and third strand bases in an *anti* conformation, was found to be the most stable by a significant amount (Table 4). This was not the result of better interactions with the counterions or solvent, but because of superior inter-base triple interactions. An alternative analysis, examining the inter-strand interactions, confirmed that the major source of the relative stability of the B*a* model was the reduced electrostatic repulsion between strands 2 and 3 (Table 6).

The question then arose as to whether this preference for the B*a* model would be maintained in the presence of alternative base triples. Two other base-triples have been found to be compatible with the CGG, $Y^+R^-R^+$, triplex modelled here: the TAA triplet and the TAT triplet (Figure 3). We examined the effect on the triplex structure and stability of the replacement of the sixth CGG triple by either of these two alternatives. As with the CGG triple, both *syn* and *anti* conformations were investigated. It was immediately observed that in both cases (TAA and TAT) mixed conformations were not possible; that is, if the CGG bases were in the *anti* conformation then the third-strand A or T could not be *syn* , and vice-versa. In view of our observations above, the solvated model protocol was used to obtain energetic parameters for the different models. In all cases it was found that the introduction of the substituted bases did not destabilise the remaining CGG triples to any meaningful extent, so for these, the B*a* conformation remained the most favourable.

For the TAA triple itself, the B*s* conformation was found to be favoured over B*a* (Table 7). This was observed to stem largely from a good electrostatic term, resulting from improved base-counterion interactions. Because of counterion mobility, the reliability of this result must be in some doubt, but a bias in favour of the *syn* conformation is also evident in the bonded term. This is presumably the result of the fact that although the *anti* conformation permits two hydrogen-bonds between the two A bases (Figure 3E), in this relative orientation they are not isomorphous with a CGG triple. In contrast, in the *syn* conformation (Figure 3F) they are structurally analogous to a *syn* CGG triple (Figure 3B) and so introduce less distortion, and therefore strain, into the backbone.

For the TAT triple, the B*a* conformation was found to remain the most favourable (Table 8), but only by a small margin and in general the differences in stability between the different conformations was much reduced over that observed for CGG triples. This is not so surprising; in all cases the substitution of a pyrimidine for a purine in the third strand would be expected to produce major alterations to the local conformation of the helix in order to achieve the optimal geometry for the modified base triple or, alternatively, less conformational change and a suboptimal local geometry. In either case, an energetic penalty is to be expected.

In conclusion, molecular models for the $R^+Y^-Y^+$, CGG, triplex have been generated and tested for physical probability on the basis of their molecular mechanical energies. In the favoured model the two Watson−Crick hydrogen-bonding strands adopt a similar conformation to that observed in the

$R^+Y^-Y^+$ triplex, and the third strand binds with *anti* glycosidic torsion angles and a sugar-phosphate backbone conformation similar to that of B-DNA. This is in agreement with the nmr evidence of Radhakrishnan *et al* (25).

The steric and electrostatic nature of the helix is found to differ considerably from that of the $R^+Y^-Y^+$ triplex, particularly with respect to the grooves. The $Y^+R^-R^+$ triplex structure involves closer contacts between phosphate groups on neighbouring strands than does the $R^+Y^-Y^+$ triplex, and so its stability may be expected to be more sensitive to factors that could modify electrostatic interactions. Thus it is interesting to note that a supercoiled plasmid DNA sequence containing homopurine-homopyrimidine sequences has been observed to form a $R^+Y^-Y^+$ triplex in the absence of Mg^{2+} , but a $Y^+R^-R^+$ triplex when Mg^{2+} is present (26).

The replacement of CGG triples with TAA triples is found to increase the relative stability of the model in which the third strand bases have the *syn* glycosidic torsion angle. However, if the various penalties associated with the substitutions are additive and sequence-independent, which they almost certainly are not, it would require a triplex composed of 75% TAA triples before the *syn* conformation became the most favoured. The *anti* conformation is thus expected to be generally observed, and there is some experimental evidence to support this (17). The replacement of CGG triples by TAT triples does not have the same result; the *anti* conformation is favoured at all base compositions.

Having a plausible model for the $Y^+R^-R^+$ triplex, we are now in a position to model alternatives to A or T for the recognition of TA base-pairs which may be predicted to have improved stabilities. In addition, the structural consequences of combining the $Y^+R^-R^+$ and $R^+Y^-Y^+$ motifs within a single triplex can now be examined in detail.

ACKNOWLEDGMENTS

This work was supported by the Cancer Research Campaign of Great Britain. Supercomputing facilities at the University of London Computer Centre were obtained courtesy of SERC grant GR/H03230. We are grateful to Prof. Dinshaw Patel for stimulating discussions.

REFERENCES

1. Beal, P.A. and Dervan, P.B. (1992) *J. Amer. Chem. Soc.*, **114**, 4976–4982.
2. Grigoriev, M., Praseuth, D., Robin, P., Hemar, A., Saison-Behmoaras, T., Dautry-Varsat, A., Thuong, N.T., Helenr, C. and Harel-Bellan, A. (1992) *J. Biol. Chem.*, **267**, 3389–3395.
3. Felsenfeld, G., Davies, D.R. and Rich, A. (1957) *J. Am. Chem. Soc.*, **79**, 2023–2024.
4. Moser, H.E. and Dervan, P.B. (1987) *Science* **238**, 645–650.
5. Broitman, S.L., Im, D.D. and Fresco, J,R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5120–5124.
6. Beal, P.A. and Dervan, P.B., (1991). *Science*, **251**, 1360–1363.
7. Arnott, S., Bond, P.J., Selsing, E. and Smith, P.J.C. (1976) *Nucleic Acids Res.*, **11**, 4141–4155.
8. Rajagopal, P. and Feigon, J. (1989). *Biochemistry*, **28**, 7859–7870.
9. de los Santos, C., Rosen, M. and Patel, D. (1989) *Biochemistry*, **28**, 7282–7289.
10. Pilch, D.S., Levenson, C. and Shafer, R.H. (1990). *Proc. Natl. Acad. Sci. USA*, **87**, 1942–1946.
11. Umemoto, K., Sarma, M.H., Gupta, G., Luo, J. and Sarma, R.H. (1990). *J. Am. Chem. Soc*, **112**, 4539–4545.
12. Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. and Patel, D.J. (1991) *Biochemistry*, **30**, 9022–9030.
13. Macaya, R.F., Schultze, P. and Feigon, J. (1992) *J. Amer. Chem. Soc.*, **114**, 781–783.
14. Johnson, K.H., Durland, R.H. and Hogan, M.E. (1992) *Nucleic Acids Res.*, **20**, 3859–3864.
15. Durland, R.H., Kessler, D.J., Gunnell, S., Duvic, M., Pettitt, B.M. and Hogan, M.E. (1991) *Biochemistry*, **30**, 9246–9255.
16. Cheng, Y.-K. and Pettitt, B.M. (1992) *J. Amer. Chem. Soc.*, **114**, 4465–4474.
17. Pilch, D.S., Levenson, C. and Shafer, R.H. (1991) *Biochemistry*, **30**, 6081–6087.
18. Jayasena, S.D. and Johnston, B.H. (1992) *Biochemistry*, **31**, 320–327.
19. van Vlijmen, H.W., Rame, G.L. and Pettitt, B.M. (1990) *Biopolymers*, **30**, 517–532.
20. Pearlman, P.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman, P.A. (1991) AMBER 4.0, University of California, San Francicso.
21. Beveridge, A.J. (1991) GEMINI 1.03, available from the Institute of Cancer Research, Sutton, Surrey, U.K.
22. Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R. (1988). *J. Mol. Graphics* **6**, 13–27.
23. Laughton, C.A. and Neidle, S. (1992). *J. Mol. Biol.*, 223, 519–529.
24. Orozco, M., Laughton, C.A., Herzyk, P. and Neidle, S. (1990) *J. Biomol. Structure and Dynamics*, **8**, 359–374.
25. Radhakrishnan, I., de los Santos, C. and Patel, D.J. (1991) *J. Mol. Biol.*, **222**, 1403–1418.
26. Kohwi, Y. and Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3781–3785.