Motifs in Nucleic Acids: Molecular Mechanics Restraints for Base Pairing and Base Stacking

STEPHEN C. HARVEY,[1] CHUNLIN WANG,[1] STEPHANE TELETCHEA,[2,*] RICHARD LAVERY[2]
[1]*Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294-0005*
[2]*Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France*

Received 22 April 2002; Accepted 1 August 2002

Abstract: In building and refining nucleic acid structures, it is often desirable to enforce particular base pairing and/or base stacking interactions. Energy-based modeling programs with classical molecular mechanics force fields do not lend themselves to the easy imposition of penalty terms corresponding to such restraints, because the requirement that two bases lie in or near the same plane (pairing) or that they lie in parallel planes (stacking) cannot be easily expressed in terms of traditional interactions involving two atoms (bonds), three atoms (angles), or four atoms (torsions). Here we derive expressions that define a collection of pseudobonds and pseudoangles through which molecular mechanics restraints for base pairing and stacking can be imposed. We have implemented these restraints into the JUMNA package for modeling DNA and RNA structures. JUMNA scripts can specify base pairing with a variety of standard geometries (Watson–Crick, Hoogsteen, wobble, etc.), or with user-defined geometries; they can also specify stacking arrangements. We have also implemented "soft-core" functions to modify van der Waals and electrostatic interactions to avoid steric conflicts in particularly difficult refinements where two backbones need to pass through one another. Test cases are presented to show the utility of the method. The restraints could be adapted for implementation in other molecular mechanics packages.

 J Comput Chem 24: 1–9, 2003

Key words: nucleic acid structures; base pairing; base stacking interactions

Background

The study of structure–function relationships in RNAs and RNA/protein complexes has been receiving increased attention. A growing body of high resolution structural data from X-ray crystallography and NMR is available in the Nucleic Acids Database (NDB),[1] and it is being supplemented by lower resolution studies utilizing a variety of techniques, particularly cryo-electron microscopy (cryo-EM).[2]

Consider, as just one example, the ribosome, which is responsible for translating the sequence of nucleotides in the messenger RNA (mRNA) into the sequence of amino acids in the corresponding protein. A few structural details of the translational cycle have been revealed by the crystal structures of the 70S ribosome[3] and its components.[4–7] But these structures and those of smaller ribosomal components only provide snapshots at isolated points in the translational cycle. Lower resolution views of many more points in the cycle are provided by cryo-EM.[8–11] The resolution of the cryo-EM maps is, under favorable circumstances, sufficiently high that it should be possible to model the details of the conformational changes corresponding to different steps in the translational cycle, by fitting the crystal structures into the cryo-EM maps and making the changes required to optimize the fits, sometimes using other data as restraints. Such efforts require powerful and accurate tools for modeling RNA structures.

There are several tools available for building, examining, and refining three-dimensional RNA models.[12–14] Although the resulting models can be refined by traditional molecular mechanics methods,[15,16] existing molecular mechanics packages do not explicitly pay attention to the favored geometries of base stacking and base pairing that are responsible for the propensity of nucleic acids to form helical structures. The present work was designed to remedy this shortcoming by defining restraint terms appropriate to stacking and pairing.

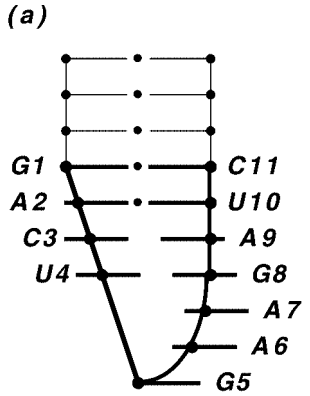
*Present address: Université René Descartes, UMR CNRS 8601, 45 rue des Saints-Pères, 75270 Paris, France

Correspondence to: S. C. Harvey; e-mail: harvey@uab.edu

Contract/grant sponsor: NIH; contract/grant number: Fogarty grant F06-TW02350 (to S.C.H.) and P41-RR12255 (C. L. Brooks, III, PI)

We have implemented restraints for base stacking and pairing within JUMNA.[17,18] JUMNA is an all-atom molecular mechanics package with a standard force field, but it utilizes a set of helicoidal parameters as independent variables, because these are the natural variables of nucleic acids. A standard convention is used for naming these variables.[19] (For completeness we note that this convention has recently been updated[20]). Because bases can generally be treated as rigid objects, the position and orientation of one base relative to its predecessor can be specified by three translational and three rotational variables. In JUMNA, these variables are defined relative to the helical axis (x-displacement, y-displacement, rise, inclination, tip, and twist). One can specify the values of any subset of these, holding them fixed, and the program will minimize the conformational energy within the limitations of those constraints by varying the remaining helicoidal and backbone variables.

The purpose of the present work is to develop force field terms to promote standard geometries of base pairing (coplanarity; proper hydrogen bonding distance) and of base stacking (parallel alignment; proper separation; alignment favoring the A-form helix). These restraints have been developed within the framework of JUMNA, but they could also be implemented in other modeling or simulation programs that use molecular mechanics force fields.



(b)

```
PR W 1 11
PR W 2 10
ST  1  2  3  4
ST  5  6  7  8  9 10 11
```

Figure 1. (a) Schematic structure of the model for the $tRNA^{Phe}$ anticodon loop and the bottom two base pairs of the anticodon stem (heavy lines). The complete stem has five base pairs, and the three that are not modeled are indicated here by light lines. The first two bases of the loop are stacked onto the 5′ side of the stem. Following the U-turn betwen U4 and G5 (which correspond to U33 and G34 in $tRNA^{Phe}$), the remaining five bases of the loop are stacked onto the 3′ side of the stem. Numbering corresponds to the model-building exercise described in Figure 5. (b) JUMNA restraints for a conformation resembling a tRNA anticodon stem/loop, contained in the file "loop2.noe." The truncated model stem has two base pairs, G1-C11 and A2-U10, with the Watson–Crick base pairing geometry indicated in the restraint file by "PR" to enforce pairing, and "W" to enforce Watson-Crick geometry. Stacking on the 5′ and 3′ sides of the loop is defined by the two "ST" commands, respectively. The break in stacking mimics that of the U-turn at the bottom of the anticodon loop in tRNA. These restraints are contained in the file "loop2.noe" referred to in Figure 5.

Methods

The problem is most easily illustrated by introducing an example. Suppose we wish to model an RNA stem-loop with a sequence and structure similar to that of the tRNA anticodon stem-loop (Fig. 1a). The conserved pyrimidine at the second position within the loop (U4 in the model stem-loop corresponds to U33 in $tRNA^{Phe}$) is at the start of the classic tRNA U-turn. This leads to a 725 loop geometry, using our earlier notation.[21] That is, the loop contains seven nucleotides, and two of these are stacked on the 5′ side of the loop, while the other five are stacked on the 3′ side of the loop.

Figure 1b shows the minimal JUMNA restraint file for building a truncated stem-loop with this geometry. (The restraint file is referenced in the JUMNA control file with the command "noe=*filename*"). The loop sits at the end of a two base pair stem. The two "PR W" commands enforce Watson–Crick pairing for basepairs G1—C11 and A2—U10. The two "ST" commands require that bases 1-2-3-4 be stacked, and that bases 5-6-7-8-9-10-11 be stacked; these commands are intended to guarantee the 725 geometry.

JUMNA interprets the "PR" and "ST" commands to generate a series of restraint terms that are added to the potential energy function. These will now be discussed in detail.

Base Pairing

The base pairing penalty involves distance restraints (pseudobonds) between pairs of atoms that are hydrogen bonded to one another, angle restraints to ensure colinearity of triplets of atoms in a hydrogen bonded complex (e.g., N—H—O), and an angular term for enforcing coplanarity of the bases. The coplanarity restraint calculates the dot product of unit vectors perpendicular to the two paired bases and requires that it be either +1 or −1, depending on the geometry of the base pairing interaction (Watson–Crick, Hoogsteen, etc.); this will be discussed in detail below. We have developed a convenient nomenclature for defining spatial relationships between bases in nucleic acids containing from one to four strands.[22]

Base pairing within the stem is indicated by the "PR W" commands in Figure 1b, where "PR" indicates pairing and "W" will produce Watson–Crick geometry. "W" can be applied to A–U, A–T, G–C, or I–C pairs; if applied to a G–U or G–T pair, the command is interepreted as "wobble" and will produce a standard wobble hydrogen bonding geometry.

Other pairing geometries are also available (Table 1), including Hoogsteen (H), reverse Hoogsteen (RH), reverse Watson–Crick (RW), and the sheared geometry (S) frequently found in A–A and G–A pairs at the ends of double helices.[23] If the user wishes to enforce some other pairing geometry, the program also accepts the "PR M" command, where "M" indicates "manual." Each such

Table 1. Base Pairing Restraints Available in JUMNA

Common name	Saenger ID[a]	Paired bases	Code	Bonded atom pairs
Watson-Crick	XX	AU or AT	W	N1—HN3, H1N6—O4
Watson-Crick	XIX	GC	W	H1N2—O2, O6—H1N4
Watson-Crick	XIX	IC	W	HN1—N3, O6—H1N4
Wobble	XXVIII	GU or GT	W	O6—HN3, HN1—O2
"Imino"[b]	VIII	GA	W	O6—H1N6, HN1—N1
Reverse Watson-Crick	XXII	GC	RW	H1N2—N3, HN1—O2
	XXI	AU or AT	RW	N1—HN3, H1N6—O2
	XXVII	GU or GT	RW	O6—HN3, HN1—O4
Hoogsteen	XXIII	AU or AT	H	N7—HN3, H2N6—O4
Reverse Hoogsteen	XXIV	AU or AT	RH	N7—HN3, H2N6—O2
	V	AA	RH	N7—H1N6, H2N6—N1
	VII	GG	RH	N7—HN1, O6—H1N2
Sheared[b]	XI	GA	S	N3—H2N6, H2N2—N7
Sheared[b]	[c]	AA	S	N3—H2N6, H2C2—N7
Manual	[d]	[d]	M	[d]

[a]ID from Figure 6.1 of Saenger.[24]
[b]Nomenclature of Elgavish et al.[23]
[c]The sheared AA is isosteric with the sheared GA (Saenger XI). Although not as strong as a traditional hydrogen bond, the H2C2—N7 interaction provides some stabilization.
[d]The "manual" feature can be used to pair any two bases in any desired geometry. See text for discussion.

command generates a pairing restraint between a designated pair of bases by defining two pairs of atoms to be bonded, along with an indication of the relative orientation of the base pair normals. The command is of the form

PR M atom1 atom2 atom3 atom4 unit1 unit2 sign

Atoms 1 and 3 are in the nucleotide number identified as unit 1, and atoms 2 and 4 are in unit 2; bonds are enforced between atoms 1 and 2, and between atoms 3 and 4. Sign is either +1 or −1, depending on whether the base normals are to be parallel or antiparallel. For example, if we wished to generate the asymmetric U–U pair XVI in Figure 6.1 of Saenger[24] in a hypothetical antiparallel duplex where we wanted to pair U7 with U12, we would use the command

PR M HN3 O4 O2 HN3 7 12 −1

where the final "−1" indicates that the two base normals are to be kept antiparallel. The command would generate the two hydrogen bond restraints in panel XVI of Saenger's Figure 6.1. Table 1 gives a full listing of the JUMNA base pairing codes and the restraints generated by each of them.

Base Stacking

Stacking is specified in the JUMNA restraint file with "ST" commands, as shown in Figure 1b. The command "ST 1 2 3 4" generates three independent sets of stacking restraints, one for stacking base 2 on base 1, a second for stacking base 3 on base 2, and a third for stacking base 4 on base 3. Up to 20 bases can be stacked with a single command.

The stacking penalty involves two distance restraints and an angle term. One of the distance restraints assures that reference atoms in two successive bases are spaced at an appropriate stacking distance, 3.3–3.6 Å, in the "vertical" direction (along a vector perpendicular to the mean plane of the two base pairs). The second distance term restrains the sliding of those same reference atoms in the "horizontal" direction (perpendicular to the aforementioned vector), so that the reference atoms are reasonably well aligned when viewed along the vertical axis. The exact definitions of "vertical" and "horizontal" are given in the next section.

Reference atoms for defining the stacking restraints were chosen by a survey of A-form RNA double helices from the Nucleic Acid Database (NDB). Four pairs of reference atoms are used, one pair each for RR, RY, YY, and YR stacking (R = purine; Y = pyrimidine). The stacking angle restraint resembles that for the base pairing case, guaranteeing that the two base planes are essentially parallel. Although currently restricted to A-form helices, this approach could easily be adapted to B-DNA conformations.

Defining the Orientation of the Base Plane

Figure 2 shows the definitions of various vectors required for the pairing and stacking restraints. Each base plane is characterized by a vector normal to it. We define a vector $\boldsymbol{d}$ along the bond from N1 to C1′ (pyrimidines), or along N9 to C1′ (purines), and a second vector $\boldsymbol{e}$ along the bond from N1 to C6 (pyrimidines), or along N9 to C8 (purines). The cross product of these vectors

$$\mathbf{u} = \frac{\boldsymbol{e} \times \boldsymbol{d}}{|\boldsymbol{e}||\boldsymbol{d}|}$$

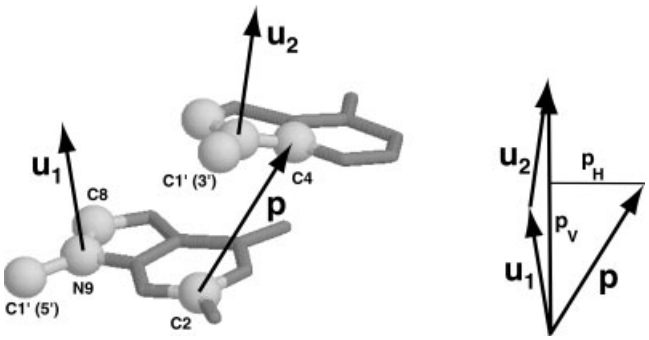


Figure 2. Vectors required for the pairing and stacking restraints. Two vectors are used to define the base plane: one, $\boldsymbol{d}$, along the glycosidic bond connecting N9 (purines) or N1 (pyrimidines) with the ribose C1′ atom, and the other, $\boldsymbol{e}$, along the bond connecting N9 with C8 (purines), or along the bond connecting N1 with C6 (pyrimidines). The crossproduct of $\boldsymbol{e}$ with $\boldsymbol{d}$ defines a vector $\boldsymbol{u}$ that is perpendicular to the base plane. Two of the $\boldsymbol{u}$ vectors are shown here, $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, although the corresponding $\boldsymbol{d}$ and $\boldsymbol{e}$ vectors are not shown. The subscripts identify bases 1 and 2, which are numbered in the 5′ to 3′ direction; the backbone connecting those bases is not shown. The stacking restraint requires an additional vector, $\boldsymbol{p}$, connecting reference atoms that are vertically well aligned in successive stacked bases in A-form RNA structures. For successive purines, $\boldsymbol{p}$ connects C2 of the 5′ base with C4 of the 3′ base. The "vertical" direction is defined by a vector $\boldsymbol{w} = \boldsymbol{u_1} + \boldsymbol{u_2}$ that is essentially perpendicular to the mean plane of the two bases, and $\boldsymbol{p}$ is resolved into vertical and horizontal components $\boldsymbol{p_V}$ and $\boldsymbol{p_H}$, respectively. See text for formal definitions and further explanation.

is then perpendicular to the base plane. Figure 2 shows two such vectors, $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, for two purines. (It should be noted that JUMNA uses idealized, planar bases. If one is manipulating real structures, a CURVES analysis[25,26] can be carried out to find the best fit planes for the bases.)

Base Pairing Restraints

For Watson–Crick base pairing, hydrogen bonding distances are measured between two donor/acceptor pairs (N1–HN3 and HN6–O4 for A–U; O6–HN4 and HN2–O2 for G–C). No energy penalty is charged if the H-acceptor distance is less than 1.75 Å or greater than 2.00 Å. Outside this range, harmonic energy penalties are applied with a user-specified force constant, with a default value of 12 kcal/(mol · Å^2). Restraint forces are calculated from the negative gradient of the energy. Standard base pairing geometries that are recognized by the program are indicated in Table 1. Any other base pairing geometry can be enforced by entering user-specified donor/acceptor pairs into the NOE restraint file with the manual pairing restraint "PR M" discussed above.

Approximate linearity of each donor-hydrogen-acceptor angle is enforced by an additional angle restraint term in the potential function, for example, N—H–O. No penalty is applied if the angle is within 15° of 180°, but a harmonic restraint is added outside of this range. It includes a user-defined force constant whose default is 12 kcal/(mol · rad^2). The negative gradient of this energy term gives a restraining torque that is converted to appropriate forces on the atoms.

Coplanarity requires that the u-vectors for Watson–Crick base pairs be antiparallel. If the normals to the two paired bases are $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, respectively, we calculate the angle and associate with it an energy penalty $E_c = k(\theta - \theta_{\text{tol}})$,^{2} where $\theta = \cos^{-1}(\boldsymbol{u_1} \cdot \boldsymbol{u_2}) - 180°$ is a limiting value up to which no energy penalty is charged (10° in the current implementation) and k is a user-specified force constant, with a default value of 12 kcal/(mol · radian^2). Base pairing patterns other than Watson–Crick are treated identically, except that the factor of 180° is changed to 0° for those cases where the two $\boldsymbol{u}$ vectors are required to be parallel, for example, Hoogsteen pairs. Bearing in mind that the protocols described here are intended to build idealized models, users who are concerned that the coplanarity restraint will unduly restrict propeller twist or buckle should do a final JUMNA optimization without pairing restraints at the end of the model building.

The negative gradient of the coplanarity energy term provides a restraining torque, and forces on the three atoms defining each base plane are calculated from this torque by repeated application of the chain rule.

Stacking Restraints

For bases to be properly stacked, three conditions must be met. First, the bases must be essentially parallel. Second, the two bases must be separated by an appropriate contact distance, about 3.3–3.6 Å. Third, the sliding of one base relative to the other must be small, in the sense that one base cannot be displaced too far away from "vertical" alignment with respect to the second base. "Vertical" is defined by a vector that is essentially perpendicular to the mean plane of the two bases (Fig. 2). To measure the "horizontal" sliding of one base relative to the other, we require a pair of reference atoms, one in each base, chosen such that the vector connecting them, $\boldsymbol{p}$ in Figure 2, is essentially parallel to $\boldsymbol{w}$ in a standard A-form RNA double helix. For any real structure, $\boldsymbol{p}$ is decomposed into vertical and horizontal components $\boldsymbol{p_V}$ and $\boldsymbol{p_H}$. $\boldsymbol{p_V}$ and $\boldsymbol{p_H}$ are scalars, and the vertical stacking restraint corresponds to holding $\boldsymbol{p_V}$ close to the van der Waals contact distance (3.3–3.6 Å), while the horizontal sliding restraint corresponds to holding $\boldsymbol{p_H}$ close to zero. The formal definitions of the various vectors in Figure 2, and how they are used to measure deviations from ideal stacking, will now be developed.

The deviation of any pair of bases from parallel stacking is measured through the use of an angular term similar to that used for coplanarity in base pairing: we calculate the angle and again associate with it an energy penalty. The force constant and limit angle are the same as those used to ensure coplanarity in the base pairing restraint. As in that case, the gradient of this function provides torques that are converted to forces on the atoms used to define the $\boldsymbol{u}$ vectors through repeated application of the chain rule.

We need a vector $\boldsymbol{p}$ connecting a pair of reference atoms to measure the vertical separation of the bases, and to quantify their horizontal sliding (Fig. 2). To find appropriate reference atoms, we examined all the double-helical regions in the crystal structures of the 50S and 30S ribosomal subunits[4,6] and generated histograms for the probability distributions of $\boldsymbol{p_H}$ for all $\boldsymbol{i,j}$ pairs, where $\boldsymbol{i}$ is an atom in one base and $\boldsymbol{j}$ is an atom in an adjoining stacked base. The

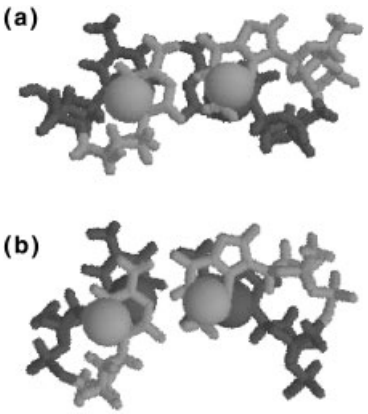


Figure 3. Stacking of control atoms. In both panels a T-A base pair (dark gray) lies below a C–G base pair (light gray), with the 5′-to-3′ direction pointing upward for the strand on the left (5′-TC-3′) and downward for the strand on the right (5′-GA-3′). (a) The standard A-form RNA double helix places the C2 atom of the thymine base directly below the C6 atom of the cytidine base above it (light gray sphere on the left), and it places the C4 atom of the adenine base directly below the C2 atom of the guanine base above it (large gray sphere on the right). Therefore, although the control atoms in the T–A base pair are represented as dark gray spheres, they are almost completely obscured by the control atoms in the C–G base pair above them. (b) In a B-form double helix, the lower control atoms are visible because they are not stacked directly below the control atoms in the upper base pair. Note that in both panels the view is perpendicular to the mean plane of the base pairs. The view is directly along the helix axis in the B-form helix, because base pairs are perpendicular to the axis, but the view is not along the axis in the A-form helix, because base pairs have large nonzero inclination and tip angles.

best pairs are those that consistently have the smallest value of $\boldsymbol{p}_H$. For each of the following pairs of reference atoms, over 80% of the horizontal sliding distances were less than 1 Å in our survey of all stacked bases in the ribosomal subunit crystal structures (R = purine; Y = pyrimidine):

R C2−R* C4 for a 5′-RpR*-3′ stack

R N1−Y* N3 for a 5′-RpY*-3′ stack

Y C2−Y* C6 for a 5′-YpY*-3′ stack

Y N3−R* N7 for a 5′-YpR*-3′ stack

Figure 3 shows the stacking of two successive RY base pairs from idealized A-form and B-form double helices and illustrates how the control atoms are very well aligned for the former, and not so well aligned for the latter.

$\boldsymbol{p}$ is decomposed into vertical and horizontal components by projecting it onto the vertical axis. If we define a unit vertical vector

$$\boldsymbol{v} = \boldsymbol{w}/|\boldsymbol{w}|$$

then the vertical and horizontal components of $\boldsymbol{p}$ are given by the scalar quantities

$$\boldsymbol{p}_V = \boldsymbol{p} \cdot \boldsymbol{v}$$

and

$$\boldsymbol{p}_H = (\boldsymbol{p}^2 - \boldsymbol{p}_v^2)^{1/2}$$

Probability distributions for $\boldsymbol{p}_H$ for the pairs of reference atoms listed above are quite narrow, with over 80% of the stacked pairs having $\boldsymbol{p}_H$ less than 1 Å. Figure 3 shows a typical case. We define horizontal restraints using

$$\boldsymbol{E}_H = \boldsymbol{k}(\boldsymbol{p}_H - \boldsymbol{p}_H^*)^2 \text{ if } \boldsymbol{p}_H \geq \boldsymbol{p}_H^*;$$

otherwise, $E_H = 0$. In the current implementation $\boldsymbol{p}_{\mathbf{H}}^* = 1$ Å. The force constant is again user-specified, with a default value of 12 kcal/(mol · Å^2). This force constant is sufficiently strong to substantially penalize the sliding characteristic of B-form stacking relative to A-form stacking (Fig. 3), so optimization with these stacking restraints favors A-form helices. The vertical restraint energy has a similar form,

$$\boldsymbol{E}_V = \boldsymbol{k}(\boldsymbol{p}_V - \boldsymbol{p}_{hi})^2 \text{ if } \boldsymbol{p}_V \geq \boldsymbol{p}_{hi};$$

$$\boldsymbol{E}_V = \boldsymbol{k}(\boldsymbol{p}_V - \boldsymbol{p}_{lo})^2 \text{ if } \boldsymbol{p}_V \leq \boldsymbol{p}_{lo};$$

$$\boldsymbol{E}_V = 0 \text{ otherwise.}$$

We have used $\boldsymbol{p}_{\mathrm{hi}} = 3.6$ Å and $\boldsymbol{p}_{\mathrm{lo}} = 3.3$ Å. Structure refinement is not particularly sensitive to the lower boundary, because of interbase van der Waals repulsions. The upper limit was chosen because our survey of the ribosomal crystal structures revealed that 80% of the stacked pairs have $\boldsymbol{p}_V$ less than 3.6 Å.

Finally, we note that the gradients of E_H and E_V are expressed in terms of the forces on all eight labeled atoms in Figure 2.

Soft-Core Functions

In the case of large RNA fragments with complex folds, the folding pathway induced by the stacking and pairing restraints can become blocked due to collisions between different strand segments. This problem can be overcome by using soft-core functions for the nonbonded interactions, allowing the strands to pass through one another.

We have developed specific soft-core functions for both electrostatic and Lennard–Jones terms to replace the standard force field terms when the interatomic distance is less than a defined interaction limit distance (Fig. 4). For the electrostatic terms (whether or not distance-dependent dielectric damping is used), we employ a quadratic function of the form $ar^2 + b$, where the parameters a and b are defined by requiring that the soft-core function has the same value and the same derivative as the pairwise electrostatic interaction at the limit distance r_O (currently set to 2.5 Å); note that the values of a and b are therefore different for different pairs of charged atoms. For the Lennard–Jones terms, we use a cubic function of the form $ar^3 + br^2 + c$, where the

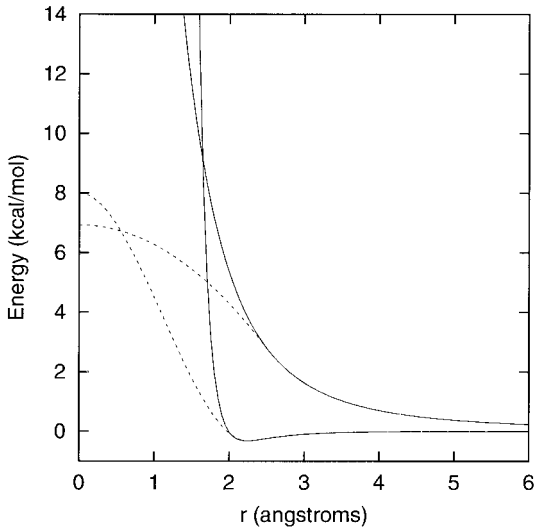


Figure 4. The standard force-field terms (solid lines) and the soft-core functions (dotted lines) are shown as a function of the interatomic distance r for typical Lennard–Jones (lower curve) and electrostatic interactions (upper curve).

parameters a and b are again fixed by requiring a continuous fit to the pairwise Lennard–Jones interaction at its energy minimum, and the parameter c can be adjusted to set the maximal repulsive interaction at $r = 0$; again the values of a and b are unique to each Lennard–Jones pair. Note that the absence of linear terms with respect to r in both these functions ensures that their derivatives become zero at $r = 0$, ensuring smooth passage of the atoms through one another. Note also that attractive electrostatic terms will remain attractive at $r = 0$ using the present soft-core function. The freedom in the cubic Lennard–Jones function should thus be used to adjust the parameter c to ensure that repulsive interactions dominate when two segments of the molecule become superimposed. These soft-core functions are illustrated graphically in Figure 4 for two typical nonbonded interactions; $c = 8$ kcal/mol in this example.

Results

Modeling a Stem-Loop Structure

As mentioned above, the anticodon loop of $tRNA^{Phe}$ has two bases stacked on the 5′ side of the loop and then, following the U turn at residue 33, there are five bases stacked on the 3′ side of the loop. If this loop is placed at the end of a short two base pair stem, it satisfies the pairing and stacking restraints given in Figure 1b.

JUMNA can use the restraints of Figure 1b to quickly construct a model stem-loop, starting with a very crude initial model derived by cutting out the eleven nucleotides from a duplex with ideal A-form RNA geometry. The procedure requires three steps, and the command files for each of these are given in Figure 5. The duplex contains the sequence

5′-GACUuuc-3′

3′-CUAGAAG-5′

where the capitals indicate the nucleotides to be cut out to give the tRNA anticodon stem-loop sequence

(a)

```
/apps/a/bin/Cur5_s <<!
 &input file='ARNA.pdb',
  lis='loopstart',pdb='loopstart', axout='loopstart',
  line=.t., &end
 1 11  0  0  0
 1  2  3  4 14 13 12 11 10 9  8
 0. 0. 0. 0.
!
--------------------------------
```

(b)

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=1000,
 axe=loopstart, noe=loop1,
 out=loopclose, pdb=loopclose, &end
0 0 0 0 0
1 11 0 0 0
GACUGAAGAUC
XYIT----
H-------
-------- *9
!
--------------------------------
```

(c)

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=10000,
 axe=loopclose, noe=loop2,
 out=looprefine, pdb=looprefine, &end
0  0 0 0 0
1 11 0 0 0
GACUGAAGAUC
XYIT----
H-------
-------- *9
!
```

Figure 5. Command files for generating the stem-loop structure described by the restraints in the file "loop2.noe" of Figure 1. (a) First, a CURVES analysis of a standard RNA double helix containing seven base pairs (the file ARNA.pdb) produces the helicoidal description of the starting stem–loop structure. See text for a description of the nucleotide renumbering produced by this analysis. The command "line = .t." carries out the analysis with respect to a best fit linear axis. The important output file is "loopstart.axe," which is used as input for JUMNA minimization. The output file "loopstart.pdb" is a JUMNA graphical output file with a simplified representation of the backbone, the bases, and a vector for the best fit axis. (b) Second, a JUMNA minimization is carried out using the same base pairing restraints as given in Figure 1, but with only two stacking restraints, "ST 1 2" and "ST 10 11"; these are given in the file "loop1.noe." (c) Final minimization, carried out with the full set of base pairing and stacking restraints in "loop2.noe," which is given in Figure 1.

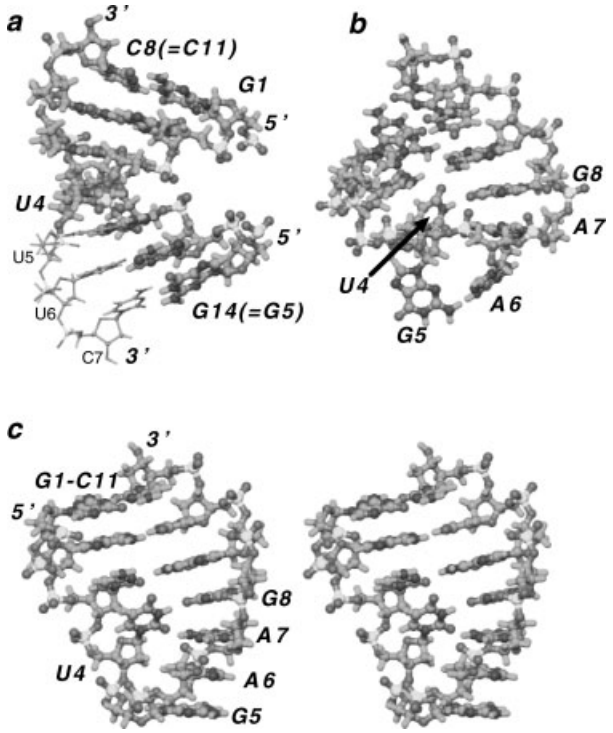


Figure 6. Structures produced by the three command files in Figure 5. (a) The starting structure is a standard A-RNA seven base pair duplex, from which three nucleotides are removed (light lines). (b) Initial refinement closes the loop while preserving base pairing and stacking in the two base pair stem, but the U-turn geometry is not particularly good, and the anticodon residues (G5, A6, A7) are not well stacked. (c) Final refinement with all pairing and stacking restraints produces the desired result, with proper stacking of U4 on the 5′ side of the loop and good stacking of the anticodon on the 3′ side (stereo view).

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356,
 opt=-1, maxn=1000,
 out=work1, pdb=work1, &end
0  0  0  0  0
2  7 -7  0  0
GACUUUC
CUAGAAG
0#------ -5.28 0. 0.0 0. 0.   0. 21.1 -24.9 38.9 -155. -67. -60.
0#------ -5.28 0. 3.4 0. 0.  33. 21.1 -24.9 38.9 -155. -67. -60. *2
0#------ -5.28 0. 1.7 0. 0. -33. 21.1 -24.9 38.9 -155. -67. -60.
0#------ -5.28 0. 1.7 0. 0.  66. 21.1 -24.9 38.9 -155. -67. -60.
0#------ -5.28 0. 3.4 0. 0.  33. 21.1 -24.9 38.9 -155. -67. -60. *2
0#------ -5.28 0. 0.0 0. 0.   0. 21.1 -24.9 38.9 -155. -67. -60.
0#------ -5.28 0. 3.4 0. 0.  33. 21.1 -24.9 38.9 -155. -67. -60. *6
!
```

Figure 7. Command file for generating the initial duplex structure with an unpaired base (U4) for building models with bulges. See text for discussion.

5′-GACUGAAGAUC-3′

In the first step, analysis of the duplex using CURVES[25,26] provides the helicoidal parameters of the initial single-stranded structure. The input duplex, ARNA.pdb was built with JUMNA, with nucleotides 1–7 comprising the first strand in a 5′-to-3′ direction, and nucleotides 8–14 comprising the other strand in a 3′-to-5′ direction. Residues 5–7 are discarded, and the next residue in the single-stranded structure is G14, followed by A13, etc. (The nucleotide numbering in the initial duplex is not carried forward into the output .axe file, where the residues are simply renumbered serially from 1 to 11.) The structure, "loopstart," is shown in Figure 6a.

This initial loop structure has two base pairs at the top of the stem (G1–C11 and A2–U10); two bases (C3 and U4) are perfectly stacked on the 5′ side of the loop, and five bases (G5, A6, A7, G8, and A9) are perfectly stacked on the 3′ side. But the backbone is broken between U4 and G5, with a gap of 18 Å. The second step of the procedure is to use JUMNA to repair this break. If the complete set of pairing and stacking restraints is imposed at this point, the program is unable to properly realign the bases during closure of the backbone, so a reduced set of stacking restraints is used. Only those restraints needed to prevent the breakup of the two base pair stem during JUMNA energy minimization are used. Thus, the file loop1.noe contains two pairing restraints (PR W 1 11 and PR W 2 10) and two stacking restraints (ST 1 2 and ST 10 11).

The resulting intermediate structure, contained in the file "loopclose.pdb" and shown in Figure 6b, has a good stem structure, and the backbone is properly closed. The loop geometry is not unreasonable, but the first base of the anticodon (G5) is out of position, and neither the 5′ nor the 3′ side of the loop is optimally stacked. These deficiencies are quickly remedied in the third step of the procedure, in which JUMNA minimizes the "loopclose" structure with the full set of restraints in "loop2.noe" (Fig. 1b), producing the final structure, "looprefine." It strongly resembles a tRNA stem-loop (Fig. 6c).

Modeling Bulged Nucleotides

Consider the following duplex, in which an unpaired base (U4) interrupts an otherwise regular double helix:

5′-GUAuUGC--3′

3′-CAU-ACGG-5′

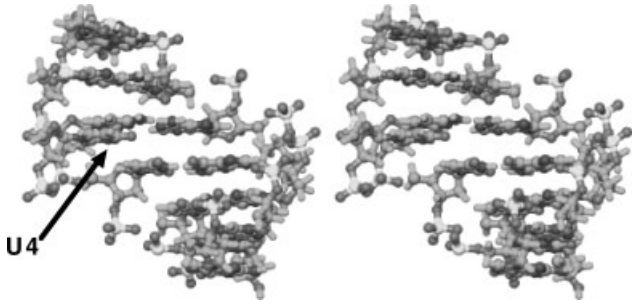


Figure 8. The structure of "work1," the starting model produced by the command file in Figure 7. Refinement of this initial model to produce two different bulge geometries is described in Figures 9–12.

(a)

```
PR W 1 8
PR W 2 9
PR W 3 10
PR W 5 11
PR W 6 12
PR W 7 13
ST  1  2  3  5  6  7
ST 14 13 12 11 10  9  8
--------------------------------
```

(b)

```
/apps/a/bin/Jum11 p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=10000,
 axe=work1, noe=bul1,
 out=bulge1, pdb=bulge1, &end
0  0  0  0  0
2  7 -7  0  0
GUAUUGC
CAUACGG
XYIT----
-------- *13
!
--------------------------------
```

(c)

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=10000, fnoes=100,
 axe=bulge1, noe=bul1,
 out=bulge1ref, pdb=bulge1ref, &end
0  0  0  0  0
2  7 -7  0  0
GUAUUGC
CAUACGG
XYIT----
-------- *13
!
```

Figure 9. Command files for refinement of the starting model of Figures 7–8 to produce a bulged base (U4) that is not stacked into the double helix. See text for discussion.

(a)

```
PR W 1 8
PR W 2 9
PR W 3 10
PR W 5 11
PR W 6 12
PR W 7 13
ST  1  2  3  4  5  6  7
ST 14 13 12 11 10  9  8
--------------------------------
```

(b)

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=10000,
 axe=work1, noe=bul2,
 out=bulge2, pdb=bulge2, &end
0  0  0  0  0
2  7 -7  0  0
GUAUUGC
CAUACGG
XYIT----
-------- *13
!
--------------------------------
```

(c)

```
/apps/a/bin/Jum11_p <<!
 &input  lib=/home/seurat/112,
 slope=.356, opt=1, maxn=10000, fnoes=100,
 axe=bulge2, noe=bul2,
 out=bulge2ref, pdb=bulge2ref, &end
0  0  0  0  0
2  7 -7  0  0
GUAUUGC
CAUACGG
XYIT----
-------- *13
!
```

Figure 11. Command files for refinement of the starting model of Figures 7–8 to produce a bulged base that is stacked into the double helix. The only difference between these files and those of Figure 9 is the inclusion of U4 in the stacking restraint list in (a). See text for discussion.

where uppercase letters indicate paired bases, and the lowercase u identifies an unpaired uracil. This bulged base can adopt either of two principal conformations, depending on its identity, the neighboring sequences, and the conditions of the experiment. It may be stacked between the bases on either side of it, producing a bend in the helix axis because one strand has more stacked bases than the other. Alternatively, the unpaired base may swing out of the double helix and dock against it in either the major groove or in the minor groove.

It is easy to build both of these models using the new pairing and stacking restraints. To begin with, we need a duplex with the bulged U positioned between paired nucleotides on either side of it. A simple JUMNA script for generating such a starting structure is given in Figure 7, and the initial structure is shown in Figure 8. As will be seen presently, we need to be able to easily manipulate the initial rise and twist angles of the bulged U4, and that is most easily done by setting the initial tip and inclination of all bases to zero (Fig. 7). In this sense, the starting structure is like a B-form helix, but we have preserved the large negative x-shift characteristic of A-form helices. These differences will be blended in the JUMNA refinement of this hybrid helix with the interrupting bulge. An examination of the helicoidal parameters (Fig. 7) or a visual inspection of the structure (Fig. 8) reveals that the bulged base is positioned midway between two bases separated by a total rise of only 3.4 Å. This is to permit easy formation of the base pairs flanking the bulged U. In subsequent refinements, either the helix will be forced to bend (if the bulged U4 is restrained to be stacked between A3 and U5), or the bulged U will be squeezed out of the helix (if bases 3 and 5 are required to be stacked). The total twist going from base A3 to U5 is 33° (−33° + 66° in Fig. 7), because we found in a series of trial runs that it is much easier to squeeze the bulged U out into the minor groove than the major groove.

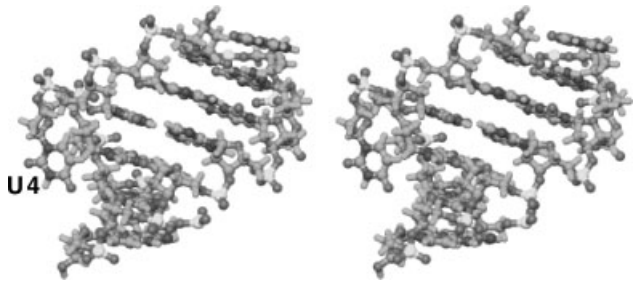


Figure 10. The final model produced by the command files of Figure 9.

Figure 9 shows the NOE restraint file and the two script files needed for JUMNA refinement when the bulged U4 is to be squeezed out of the helix, and Figure 10 shows the resulting

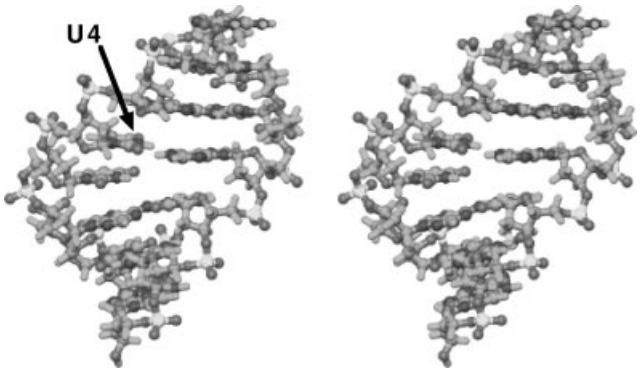


Figure 12. The final model produced by the command files of Figure 11.

structure. The uracil base is positioned against the floor of the minor groove, just as pyrimidines are frequently observed to do.

The corresponding NOE restraints and scripts for the case where the bulged U4 is required to remain stacked in the helix are given in Figure 11. The only difference between these files and those of Figure 9 are in the set of "ST" restraints along the first strand, but the resulting structure (Fig. 12) is quite different from that of Figure 10. A CURVES analysis of the structure in Figure 12 reveals that the stacked U4 induces a bend of 22° in the axis of the double helix.

Discussion

The automated procedures described here are rapid, simple, and accurate. With a modest workstation the stem-loop model requires about 10 min of total CPU time for building the initial duplex, carrying out the CURVES analysis to generate the initial single-stranded stem-loop, and doing the two steps of JUMNA optimization. About the same amount of CPU time was required for each of the two bulge models. The construction, manipulation, and running of the control files for each of these cases required about an hour, even with the examination of intermediate and final results and repeats of some steps.

Most important, the structures generated by these procedures are sterically plausible, and are of considerably higher quality than can be built by the usual procedures of manual manipulation with a traditional molecular graphics program, followed by energy minimization. Several non-Watson–Crick base pairing geometries are already available in the restraint library, and the user can easily define other geometries.

The addition of pairing and stacking restraints to JUMNA makes it possible to easily build, manipulate, and refine a wide range of model RNA motifs. Similar restraints could be incorporated into other molecular mechanics programs with the information provided here.

Acknowledgments

We thank Krystyna Zakrzewska for stimulating discussions and for assistance and stimulating discussions.

References

1. Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. Biophys J 1992, 63, 751.
2. Frank, J. Three-Dimensional Electron Microscopy of Macromolecular Assemblies; Academic Press: San Diego, 1996.
3. Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F. Science 2001, 292, 883.
4. Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. Science 2000, 289, 905.
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. Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan–Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. Nature 2000, 407, 327.
7. Ogle, J. M.; Brodersen, D. E.; Clemons, W. M., Jr.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. Science 2001, 292, 897.
8. Frank, J.; Penczek, P.; Agrawal, R. K.; Grassucci, R. A.; Heagle, A. B. Methods Enzymol 2000, 317, 276.
9. Frank, J. Bioessays 2001, 23, 725.
10. van Heel, M.; Gowen, B.; Matadeen, R.; Orlova, E. V.; Finn, R.; Pape, T.; Cohen, D.; Stark, H.; Schmidt, R.; Schatz, M.; Patwardhan, A. Q Rev Biophys 2000, 33, 307.
11. van Heel, M. Curr Opin Struct Biol 2000, 10, 259.
12. Gautheret, D.; Major, F.; Cedergren, R. J Mol Biol 1993, 229, 1049.
13. Macke, T.; Case, D. A. In Molecular Modeling of Nucleic Acids; Leontis, N. B.; J. Santa Lucia, Eds.; American Chemical Society: Washington, DC, 1998.
14. Massire, C.; Westhof, E. J Mol Graph Model 1998, 16, 197.
15. McCammon, J. A.; Harvey, S. C. Dynamics of Proteins and Nucleic Acids; Cambridge Univ. Press: London, 1987.
16. Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins. A Theoretical Perspective of Dynamics, Structure, and Thermodynamics; John Wiley: New York, 1988.
17. Lavery, R. In Unusual DNA Structures; Wells, R. D.; S. C. Harvey, Eds.; Springer-Verlag: New York, 1987.
18. Lavery, R.; Zakrzewska, K.; Sklenar, H. Comp Phys Commun 1995, 91, 135.
19. Dickerson, R. E. Nucleic Acids Res 1989, 17, 1797.
20. Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. J Mol Biol 2001, 313, 229.
21. Harvey, S. C.; Luo, J.; Lavery, R. Nucleic Acids Res 1988, 16, 11795.
22. Lavery, R.; Zakrzewska, K.; Sun, J. S.; Harvey, S. C. Nucleic Acids Res 1992, 20, 5011.
23. Elgavish, T. E.; Cannone, J. J.; Lee, J. C.; Harvey, S. C.; Gutell, R. R. J Mol Biol 2001, 310, 735.
24. Saenger, W. Principles of Nucleic Acid Structure; Springer-Verlag: New York, 1984.
25. Lavery, R. J Biomol Struct Dyns 1988, 6, 63.
26. Lavery, R. J Biomol Struct Dyns 1989, 6, 655.