ARTICLES

# The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data

Marc Parisien[1] & François Major[1]

**The classical RNA secondary structure model considers A•U and G•C Watson–Crick as well as G•U wobble base pairs. Here we substitute it for a new one, in which sets of nucleotide cyclic motifs define RNA structures. This model allows us to unify all base pairing energetic contributions in an effective scoring function to tackle the problem of RNA folding. We show how pipelining two computer algorithms based on nucleotide cyclic motifs, MC-Fold and MC-Sym, reproduces a series of experimentally determined RNA three-dimensional structures from the sequence. This demonstrates how crucial the consideration of all base-pairing interactions is in filling the gap between sequence and structure. We use the pipeline to define rules of precursor microRNA folding in double helices, despite the presence of a number of presumed mismatches and bulges, and to propose a new model of the human immunodeficiency virus-1 −1 frame-shifting element.**

The number of RNAs found to be involved in non-coding cellular roles is increasing rapidly and persistently[1,2], and many RNA transcripts of unknown function have recently been detected in eukaryotic cell maps[3]. RNAs can be grouped into families that share structural features and function. Therefore, unravelling the structure provides crucial insights into the way in which RNA works. However, producing RNA high-resolution structures by X-ray crystallography and NMR spectroscopy is slow compared to sequencing, thus creating an important gap between the number of known tertiary (three-dimensional, 3D) structures[4] and that of sequences[5].

In the search for an effective RNA structure-determination approach, we examined different theoretical schemes and studied their relative merit to attain our goal. Hope came from the fact that secondary structures would provide enough structural constraints to automate 3D building[6]. A secondary structure describes the stems of RNA—crucial building blocks that form when two complementary regions of the sequence base pair and adopt a double-helix structure. A legitimate approximation of secondary structures considers stems that consist of A•U and G•C Watson–Crick base pairs as well as G•U wobble base pairs. These base pairs are called 'canonicals'.

Secondary structures can be derived from a sequence by using a combination of free-energy minimization[7] and covariation analysis[8]. However, the presence of a few key non-canonical base pairs blurs predictions, because they contribute energies and complicate covariation interplay[9]. Even when experimental data are considered (for example, enzymatic or chemical probing), selecting the native amongst many suboptimal secondary structures remains elusive[10]. More importantly, secondary structures deprived of non-canonical base pairs are neither adequate to determine 3D structures nor sufficient to faithfully align sequences of the same family[6,11,12]. Recent attempts to replace thermodynamics by statistical scores resulted in either similar[13] or only slightly improved[14] predictive power. Furthermore, empirical scoring of 3D structures applies only to very short sequences and requires covariation data[15]. Taken together, these shortcomings and increasing needs for RNA genome-wide annotation prompted us to develop a new approach.

We extended the classical rationale underlying RNA structure prediction by incorporating all base pairs. To do so, we introduced a new first-order object to represent nucleotide relationships in structured RNAs: the nucleotide cyclic motif (NCM). The NCMs became apparent to us from an analysis of the X-ray crystallographic structure of the 23S ribosomal RNA of *Haloarcula marismortui*[16]. Adjacent NCMs share common base pairs—a property providing enough base-pairing context information to derive an effective scoring function and making possible the use of the same algorithm for predicting secondary and tertiary structures.

We propose a new RNA-structure-prediction method based on NCMs, implemented as a pipeline of two computer programs: MC-Fold and MC-Sym (Supplementary Fig. 1). We illustrate the predictive power of the pipeline by reproducing experimentally determined 3D structures from a single sequence, building 3D structures of precursor microRNA (pre-miRNA) that are compatible with Dicer docking, and proposing a new 3D structure of the human immunodeficiency virus (HIV-1) *cis*-acting −1 frame-shifting element. In practice, judicious pipeline predictions from a single sequence are expected for fragments of up to approximately 150 nucleotides.

## Folding single sequences

We evaluated the predictive power of MC-Fold by comparing the base pairs in the lowest-energy (best) predicted structure of each sequence with those found in experimental hairpin loop structures (Table 1). Compared to the thermodynamic approach, MC-Fold predicts over 6% more canonical base pairs, despite a lower positive predictive value, concurrently makes less false positives and negatives, and obtains a higher Matthews correlation coefficient ratio (MCCR) (see Supplementary Table 1). In addition, the optimal solution for each hairpin includes more than 60% of the non-canonical base pairs, and this number goes up to more than 80% if the top five solutions are considered. The low rate of false negatives is a prerequisite for building 3D structures.

We evaluated the predictive power of the MC-Fold and MC-Sym pipeline by analysing and comparing the best predictions for thirteen experimental 3D structures (Table 2). Eleven of the thirteen examples rank first (that is, match the lowest-energy structure). Eight of the thirteen examples have MCCRs of 100% (average = 98.2%). Seven of

[1]Institute for Research in Immunology and Cancer (IRIC), Department of Computer Science and Operations Research, Université de Montréal, PO Box 6128, Downtown Station, Montréal, Québec H3C 3J7, Canada.

**Table 1 | MC-Fold predictive power**

| Predicted base pairs (%) | Zipper (lower bound) | RNAsubopt (thermodynamics) | MC-Fold (NCM) |
|---|---|---|---|
| $PPV = \frac{TP}{(TP + FP)}$ | 59.6 | **91.8** | 83.4 |
| $STY = \frac{TP}{(TP + FN)}$ | 74.1 | 74.8 | **89.9** |
| $Matthews = \sqrt{\frac{TP}{(TP+FN)}\frac{TP}{(TP+FP)}}$ | 66.5 | 82.9 | **86.6** |

Best predictions over 2,093 base pairs (1,784 canonical base pairs) in 264 hairpin loops extracted from 182 different PDB structures. Columns show programs and rows show coefficients. Zipper is a program that implements a greedy algorithm that folds a sequence from bottom-up using exclusively tandems of base pairs. This gives a lower bound on the predictive power. RNAsubopt implements the current thermodynamics model and enumerates systematically all suboptimal structures. The numbers of nucleotides in these hairpin loops vary from 8 to 35 base pairs (average = 19.6). The best value for each row is shown in bold. FN, number of false negatives; FP, number of false positives; TP, number of true positives; Matthews, Matthews correlation coefficient ratio; PPV, positive predictive value; and STY, sensitivity.

**Table 2 | MC-Fold and MC-Sym pipeline predictive power**

| RNA (PDB code) | Size (nucleotides) | Rank | Matthews (%) | r.m.s.d. (Å) |
|---|---|---|---|---|
| Hairpins | | | | |
| Loop E (430D†) | 29 | 1 | 100 | 1.7 |
| IRE (1NBR) | 29 | 1 | 100 | 2.4 |
| Classical swine fever virus IRES domain III (2HUA*) | 40 | 4 | 100 | 2.6 |
| RNA thermometer (2GIO*) | 29 | 1 | 100 | 1.7 |
| Eel *UnaL2* LINE 3′ element (2FDT*) | 36 | 1 | 100 | 2.0 |
| Telomerase RNA domain IV (2FEY*) | 43 | 1 | 94 | 3.3 |
| RNase P RNA P4 (2CD1*) | 27 | 2 | 96 | 2.1 |
| GNYA tetraloop (2EVY*) | 14 | 1 | 100 | 1.8 |
| U2 snRNA (2O33*) | 20 | 1 | 100 | 2.0 |
| Group II intron branchsite (2AHT*) | 27 | 1 | 96 | 1.9 |
| Y-shape | | | | |
| Hammerhead ribozyme (1NYI†) | 36 | 1 | 100 | 2.7 |
| 5S rRNA (2HGH*) | 47 | 1 | 96 | 2.9 |
| Pseudo-knot | | | | |
| Yellow leaf virus (2AP5*) | 18 | 1 | 94 | 2.7 |

The MC-Fold and MC-Sym pipeline is applied to single sequences. Three different RNA topologies were tested: hairpin, multi-branch (Y-shape) and pseudo-knot. The best predictions (best MCCRs) are reported. The r.m.s.d. values were calculated over all heavy atoms. The average MC-Fold real time for all but one sequence is 7 s on a typical workstation processor (AMD Athlon 64, 2.2 GHz); real time for the 5S rRNA sequence is 143 s. The best 3D models are selected amongst all models generated using a probabilistic search over a period of 12 h.
* A recent NMR structure.
† X-ray crystallographic structure.

the thirteen examples combine first rank and 100% MCCRs. The average root mean squared deviations[17] (r.m.s.d.) of the thirteen examples when optimally superimposed on their corresponding experimental structures are near 2 ångströms (Å) (Fig. 1). The nucleotides that increase the r.m.s.d. are those with more degrees of freedom, that is, those not involved in base-pairing interactions (see, for instance, nucleotides A14 and U16 in the iron-responsive element (IRE) hairpin loop in Fig. 1a). Another source of high r.m.s.d. is the presence of false positives and negatives. For instance, the telomerase RNA domain IV has an MCCR of 94% and a r.m.s.d. of 3.3 Å (Fig. 1b). Interestingly, the false positive and the false negative are made in the hairpin loop. The NMR structure has a single-nucleotide bulge, A22, which stacks inside the helix on the 5′ side of a CUAU tetra-loop. The C23•U26 base pair that closes the tetra-loop is stabilized by a single hydrogen bond, and the two bases are perpendicular to each other. Although relatively stable, these features are

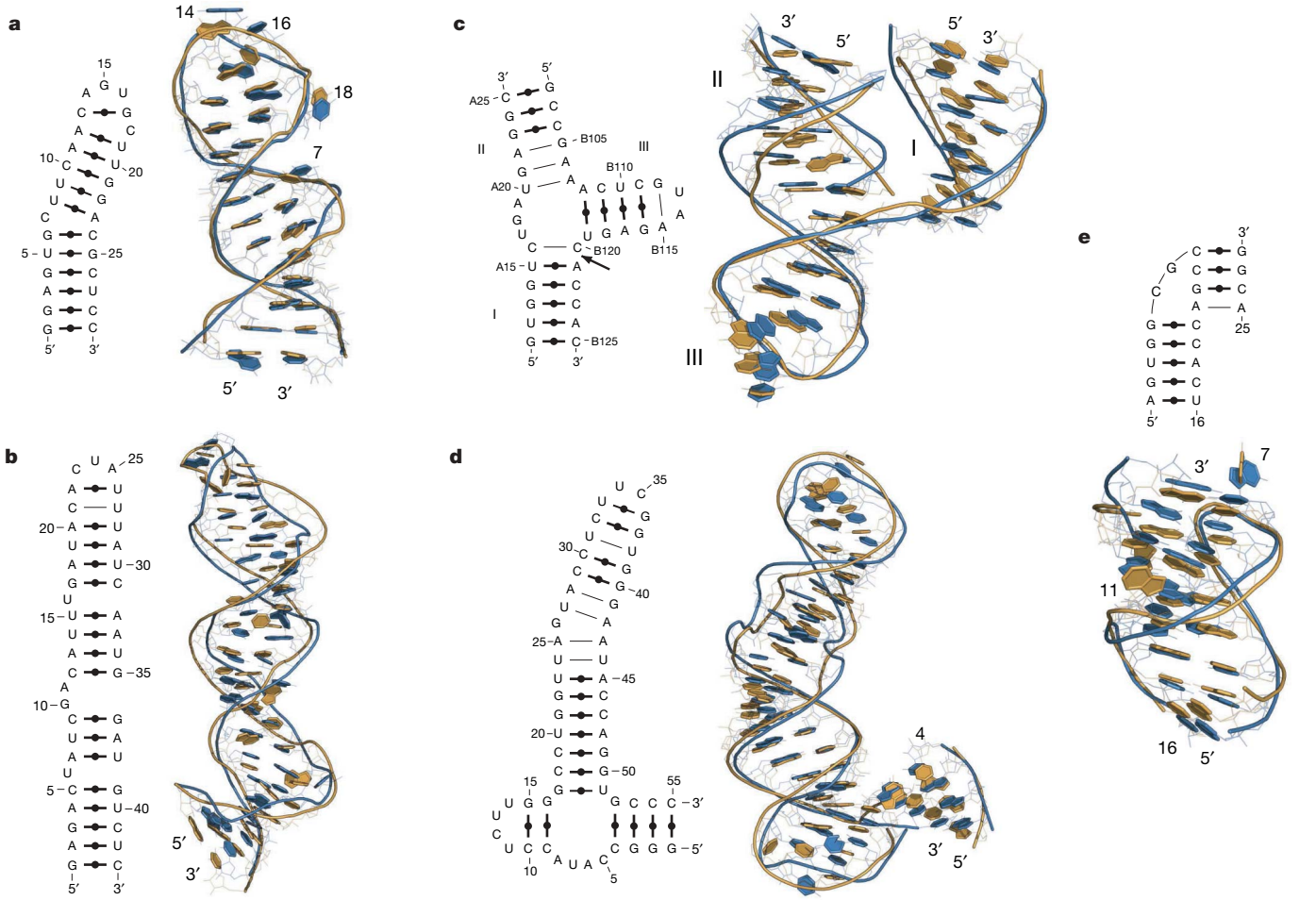


**Figure 1 | A selection of 3D structures predicted from sequence.** The canonical (bold lines, black dots) and non-canonical (non-bold lines) base pairs predicted by MC-Fold are shown on the left of the 3D structures. The closest structure (minimum r.m.s.d.) over all heavy atoms is shown (blue) is superimposed on its respective experimental structure (gold). **a**, IRE. **b**, Telomerase RNA domain IV. **c**, Pre-catalytic conformation of a hammerhead ribozyme. The arrow points the cleavage site. **d**, Subdomain of the *X. laevis* 5S rRNA. **e**, Yellow leaf virus pseudo-knotted element.

rather rare and might be induced by particular experimental conditions or structure resolution methods. The NMR hairpin loop is less stable than the penta-loop proposed by the MC-Fold and MC-Sym pipeline.

When the experimental structure does not correspond to the lowest-energy structure, it is generally due to the formation of extra base pairs in the latter. The lowest-energy structure is often referred to the 'ground state'. The base pair formation/disruption phenomenon is known to be dependent on conformational changes induced by cofactors[18], which are difficult to represent in any scoring scheme. Consequently, polymorphic structures are found in MC-Fold's suboptimal solutions.

The conserved sequence of the *Deinococcus radiodurans* and *Escherichia coli* 23S rRNA helix 40 (ref. 19) contains an interior loop, $\begin{matrix}5'-\text{CU}\mathbf{AA}\text{G}-3'\\3'-\text{GA}\mathbf{\mathit{AG}}\text{C}-5'\end{matrix}$, the structure of which differs whether it is solved by NMR or by X-ray crystallography. The NMR conditions favour the formation of a non-canonical A•A/A•G base-pair tandem, which MC-Fold ranks first (shown in bold above). The X-ray crystallographic structure is bound to a protein that possibly induces the disruption of the A•A non-canonical base pair and the apparition of a single bulged-out A (shown in bold-italic above), which MC-Fold ranks fifth.

The 'on' and 'off' conformational states of the cytoplasmic eukaryotic rRNA A site[20] contains an interior loop, $\begin{matrix}5'-\text{CGC}-\text{U}-3'\\3'-\text{A}\mathit{\mathit{A}}\mathbf{AG}-5'\end{matrix}$, the structure of which differs whether the ribosome is active in protein translation (on) or not (off). X-ray crystallographic data of the *Homo sapiens* A site reveal these two distinct structures[20]. The on state has two unpaired A nucleotides that bulge out of the main helix (shown in bold above), whereas only one A, 3′ of the two bulges in the on state, is unpaired in the off state (shown in italic above). MC-Fold ranks the on state as sixth and the off state as fourth.

Multi-branched RNAs are made of more than two helical stems that are joined by a multi-branch loop. We used the pipeline to reproduce the 3D structure of a pre-catalytic conformation of the hammerhead ribozyme[21] (Fig. 1c), as well as that of the recent NMR structure of the *Xenopus laevis* 5S rRNA bound to zinc fingers[22] (Fig. 1d). When more than two stems are selected by MC-Fold, the coaxial energies are computed and accounted for in the final score (Supplementary Methods). The key base pairs to project properly the hammerhead in 3D space are located near the multi-branch: the three base pairs at the bottom of stem II and the C•C base pair in stem I. The C•C base pair is particularly important to avoid coaxial stacking between stem I and stem III.

Finally, inserting a stem that creates a nested structure generates a pseudo-knot, as shown in the structure of the yellow leaf virus[23] (Fig. 1e). In this model, a false positive non-canonical A•A base pair is made at the bottom of the upper stem. Nevertheless, the closest generated model shares 2.7 Å of r.m.s.d. when optimally superimposed on the NMR structure.

## Folding human precursor microRNAs

When we submitted the pre-miRNA sequences of let-7c, mir-19a and mir-29a, our predictions were almost identical, and were similar to the A-RNA double helix (Fig. 2). In fact, we did not find any pre-miRNA sequence in mirBase[24] that could not be folded in the double helix (data not shown), despite an overrepresentation of U•U and U•C mismatches. The double helix offers a fixed and stable reference to the scissile phosphates that are cleaved by the Drosha complex upstream of the pre-miRNA[25], as well as by Dicer near the terminal loop[26].

The pre-miRNA double helix of let-7c (Fig. 2a) is bulge-free and presents to Dicer the expected docking surface[26], despite the non-canonical base pairs. In the 3D structure of mir-29a (Fig. 2b), the unpaired C23 nucleotide stacks inside the helix, acting as a lever to push the scissile phosphate of A26 into its proper position. Finally, in the 3D structure of mir-19a (Fig. 2c), the two unpaired nucleotides, A56 and U57, form a bulge behind the docking surface, and hence do not interfere with Dicer binding. These strict 3D structural restraints should further help in distinguishing between RNA stem-loop structures that can be processed by Dicer.

The presumed microRNA mismatches, in fact, adopt a geometry isosteric to Watson–Crick base pairs[27]. Their energies are less than that of canonical ones, which may facilitate the unwinding of the double helix and loading of the mature miRNA into the RNA-induced silencing complex (RISC). Interestingly, we find very few G•A mismatches in the miRNA region interfacing Dicer because their propensity for the sheared conformation is not isosteric to Watson–Crick base pairs. The sheared geometry distorts the backbone path of the double helix and, thus, might interfere with Dicer binding. The natural selection for non-canonical base pairs increases the diversity of possible pre-miRNA sequences, while increasing target specificity and, simultaneously, decreasing off targeting.

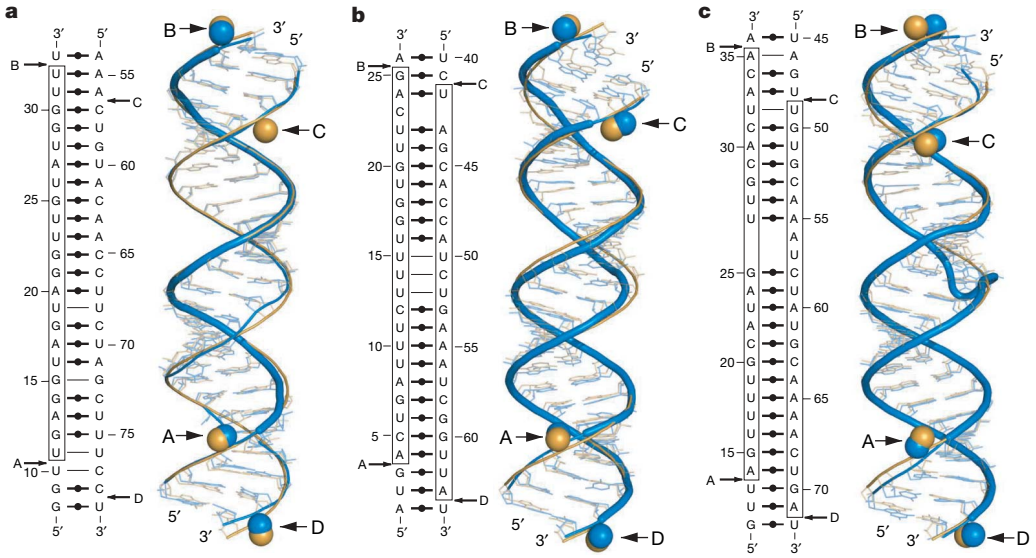


**Figure 2 | A selection of pre-miRNA 3D structures.** The predicted structures (blue) are optimally superimposed on a theoretically generated A-RNA double helix (gold). For each pre-miRNA, the nucleotides that form the mature miRNA are shown inside boxes. The spheres represent the scissile phosphate atoms: Drosha complex cleavage (A and D) and Dicer (B and C). **a**, Human let-7c. **b**, Human miR-29a. **c**, Human miR-19a.

### Folding using probing data

As shown above, MC-Fold does not always rank experimental and activated structures first. Reaching these structures is nevertheless of principal importance. Here we show how experimental data can be incorporated to restrain the conformational space of MC-Fold to identify such induced structures.

For example, a recent study investigated the yeast transfer RNA$^{Asp}$ structure by selective 2′-hydroxyl acylation and primer extension (SHAPE)[28]. SHAPE data reveal the flexible and constrained nucleotides, subject to experimental conditions. The tRNA sequence tested is deprived of the modified nucleotides, and has been shown to adopt the cloverleaf structure[29]. The top MC-Fold prediction of this tRNA$^{Asp}$ sequence is a hairpin, not a cloverleaf.

If we introduce high- and medium-flexibility SHAPE constraints (Supplementary Fig. 2), MC-Fold generates cloverleaf structures and ranks the native one sixth. The D-stem–loop sequence of this tRNA has a positive folding free energy under the thermodynamic model. Amongst the solutions, one includes a correct D-stem base-pairing registry (MCCR of 100%) and the A14•A21 base pair, whereas all other solutions base pair U13 with A21. The A14 inflexibility demonstrated by SHAPE is thus sufficient to discriminate the native amongst all solutions.

Similarly, by introducing dimethyl sulphate (DMS) data[10], all known canonical (with the exception of G56•C28) and all non-canonical base pairs of the *E. coli* 5S rRNA are captured in the correct *in vivo* Y-shaped topology (Supplementary Fig. 3a). Interestingly, the MC-Fold optimal solution (Supplementary Fig. 3b) of sub-sequence 16 to 69 has a marked resemblance to the *in vitro* structure that was probed by chemical modifications[30] and by NMR[31]. The latter suggests further a bias towards structures in the ground state.

### Consensus structural assignments

Sequences that are functionally related are another source of structural data. Consider the IRE, a hairpin loop found in the 3′ untranslated region of the ferritin and transferrin receptor mRNAs. IREs are involved in maintaining iron homeostasis in vertebrate cells by acting as post-transcriptional factors[32]. MC-Fold and MC-Sym best predictions of several IRE sequences reveal the base pairs and nucleotides found to be involved in receptor binding (Fig. 1a): an upper stem of six base pairs[33], a single unpaired nucleotide 3′ of the hairpin-loop-flanking base pair[34] and a single (V) or double (W) bulge in the 5′ strand of the stem.

MC-Cons computes a structural assignment, that is, it assigns to each sequence the structure that maximizes the overall sum of pairwise structural similarities. The structural assignment returned when 30 IRE sequences available at Rfam (RNA families database)[35] and their top ten MC-Fold predictions are input to MC-Cons (see Supplementary Methods) reveals two IRE subclasses (Supplementary Fig. 4), corresponding to both helix-bending motifs, V and W. The two subclasses have been shown to be important in selective repressor binding, in particular to the human iron responsive protein 2 (ref. 36). Similarly, using ten yeast tRNA sequences, MC-Cons identified the cloverleaf structure for each sequence (see Supplementary Fig. 5).

Multiple-sequence and low-resolution data can be used in combination. Using fourteen 5S rRNA *E. coli* sequences and DMS data, the *in vivo* 5S rRNA structure is captured (Supplementary Fig. 6). In this case, a high rate of non-canonical false positives is made in the large hairpin (nucleotides 35–47). This is probably due to the fact that the RNA in the crystal structure is bound to the ribosomal complex, in which the large hairpin makes several contacts with the ribosomal protein L5. However, the consensus structural assignment of the *E. coli* 5S rRNA sequences without DMS probing data predicts the *in vitro* structure (Supplementary Fig. 7). Similarly, the Selenocysteine Insertion Sequence (SECIS) structure is also predicted using seven sequences and various RNase probing data[37] to block the base pairing of 13 out of 151 nucleotides (Supplementary Fig. 8).

### Modelling HIV-1 frame-shifting element

HIV-1 is known for encoding two proteins, pol and gag-pol[38], using the same mRNA and a −1 *cis*-acting frame-shifting mechanism owing to the formation of a structure downstream of the slippery sequence[39,40]. Recent NMR data suggest that this structure could be a hairpin loop with an asymmetric bulge of three nucleotides, 5′-GGA-3′. In a first study, clear NMR signals were obtained by modifying the sequence to include GC base pairs in the lower stem, therefore introducing a coerced registry. In a second study, the native sequence was used. However, it was extracted from the mRNA so that the lower stem was also constrained. Besides, when MC-Fold is run with both NMR sequences, the best solutions match the structures obtained by NMR.

However, using 50 randomly selected sequences out of the 753 reported in Rfam, a single and different structure makes the consensus assignment amongst these sequences. The principal difference between the new structure and those obtained by NMR is in the bulge: MC-Fold predicts a double-A bulge, 5′-AA-3′, instead of a 5′-GGA-3′ bulge (Fig. 3). This is a minor difference, but several arguments support it (see Supplementary Discussion).

### Discussion

Our results highlight the fact that for effective RNA structure predictions, dealing with all base-pairing types in both secondary and tertiary structures is of the utmost importance. A difference between our study and other recent attempts is the use of a first-order object based on NCMs, which incorporates more base-pairing context-dependent information; this suggests that it is key in scoring secondary structures.

The lowest free-energy states determined by MC-Fold often differ from active and experimental states. Furthermore, solving the consensus structural assignment using MC-Cons occasionally predicts such ground rather than active structures (for example, *in vitro E. coli* 5S rRNA). However, we showed that few low-resolution experimental data could be introduced to bias the search towards experimental and *in vivo* structures (for example, tRNA, *in vivo* 5S rRNA and a SECIS element). Predicting both ground state and induced fit structures for the same sequence is a strong indication that MC-Fold

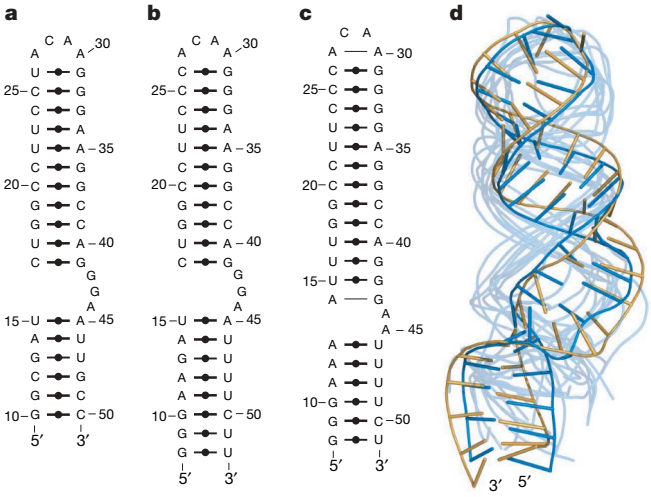


**Figure 3 | HIV-1 −1 frame-shifting-element models. a**, Secondary structure of the first NMR study (Protein Data Bank, PDB, code 1ZC5). **b**, Second NMR study (PDB code 1Z2J). **c**, MC-Cons best secondary structure prediction. The sequence used is EMBL AJ535040.1 from patient PT747 that expresses the pol and gag proteins. **d**, MC-Sym tertiary structures. The closest model (minimum r.m.s.d. of 3.4 Å over all heavy atoms) is shown in blue, optimally superimposed on the NMR structure resolved in the second study (gold), as well as ten representative structures of the conformational space of MC-Sym (light blue).

predicts correct structures, as well as structures that are accessible to any given sequence. This is reflected in the high rate of false positives when compared to experimental structures.

Thanks to the pipeline, RNA modelling is now more accurate and simpler than ever. The secondary structures generated by MC-Fold are more informative than those deprived of non-canonical base pairs and include very few false negatives. Producing 3D models consistent with these secondary structures is now a straightforward and accessible-to-all online activity. This should translate into keener RNA function hypotheses and less experimental work to verify them.

## METHODS SUMMARY

The three algorithms and the scoring function are fully described in the Supplementary Information. A web service of the three algorithms has been made publicly available on the Internet at http://www.major.iric.ca. The protocols to produce secondary and tertiary structures using the website are described elsewhere (submitted).

**Received 7 January 2007; accepted 11 January 2008.**


1. *The RNA World* 3rd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) (CSHL, Cold Spring Harbor, 2006).
2. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33,** D121–D124 (2005).
3. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316,** 1484–1488 (2007).
4. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
5. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **35,** D21–D25 (2007).
6. Shapiro, B. A. *et al.* Bridging the gap in RNA structure prediction. *Curr. Opin. Struct. Biol.* **17,** 157–165 (2007).
7. Mathews, D. H. & Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **16,** 270–278 (2006).
8. Gutell, R. R., Lee, J. C. & Cannone, J. J. The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.* **12,** 301–310 (2002).
9. Mathews, D. H. Revolutions in RNA secondary structure prediction. *J. Mol. Biol.* **359,** 526–532 (2006).
10. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA* **101,** 7287–7292 (2004).
11. Major, F. *et al.* The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science* **253,** 1255–1260 (1991).
12. Lescoute, A. *et al.* Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **33,** 2395–2409 (2005).
13. Dima, R. I., Hyeon, C. & Thirumalai, D. Extracting stacking interaction parameters for RNA from the data set of native structures. *J. Mol. Biol.* **347,** 53–69 (2005).
14. Do, C. B., Woods, D. A. & Batzoglou, S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22,** e90–e98 (2006).
15. Das, R. & Baker, D. Automated *de novo* prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA* (2007).
16. Lemieux, S. & Major, F. Automated extraction and classification of RNA tertiary structure cyclic motifs. *Nucleic Acids Res.* **34,** 2340–2346 (2006).
17. Kabsch, H. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **34,** 827–828 (1978).
18. Williamson, J. R. Induced fit in RNA-protein recognition. *Nature Struct. Biol.* **7,** 834–837 (2000).
19. Shankar, N. *et al.* The NMR structure of an internal loop from 23S ribosomal RNA differs from its structure in crystals of 50s ribosomal subunits. *Biochemistry* **45,** 11776–11789 (2006).
20. Kondo, J., Urzhumtsev, A. & Westhof, E. Two conformational states in the crystal structure of the *Homo sapiens* cytoplasmic ribosomal decoding A site. *Nucleic Acids Res.* **34,** 676–685 (2006).
21. Pley, H. W., Flaherty, K. M. & Mckay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372,** 68–74 (1994).
22. Lee, B. M. *et al.* Induced fit and ''lock and key'' recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J. Mol. Biol.* **357,** 275–291 (2006).
23. Giedroc, D. P., Theimer, C. A. & Nixon, P. L. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.* **298,** 167–185 (2000).
24. Griffiths-Jones, S. *et al.* miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34,** D140–D144 (2006).
25. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha–DGCR8 complex. *Cell* **125,** 887–901 (2006).
26. Macrae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311,** 195–198 (2006).
27. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30,** 3497–3531 (2002).
28. Merino, E. J. *et al.* RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127,** 4223–4231 (2005).
29. Perret, V. *et al.* Conformation in solution of yeast $tRNA^{Asp}$ transcripts deprived of modified nucleotides. *Biochimie* **72,** 735–743 (1990).
30. Brunel, C. *et al.* Three-dimensional model of *Escherichia coli* ribosomal 5S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.* **221,** 293–308 (1991).
31. Leontis, N. B. & Moore, P. B. NMR evidence for dynamic secondary structure in helices II and III of the RNA of *Escherichia coli*. *Biochemistry* **25,** 3916–3925 (1986).
32. Hentze, M. W. & Kuhn, L. C. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl Acad. Sci. USA* **93,** 8175–8182 (1996).
33. Jaffrey, S. R. *et al.* The interaction between the iron-responsive element binding protein and its cognate RNA is highly dependent upon both RNA sequence and structure. *Nucleic Acids Res.* **21,** 4627–4631 (1993).
34. Sierzputowska-Gracz, H., Mckenzie, R. A. & Theil, E. C. The importance of a single G in the hairpin loop of the iron responsive element (IRE) in ferritin mRNA for structure: an NMR spectroscopy study. *Nucleic Acids Res.* **23,** 146–153 (1995).
35. Griffiths-Jones, S. *et al.* Rfam: an RNA family database. *Nucleic Acids Res.* **31,** 439–441 (2003).
36. Leipuviene, R. & Theil, E. C. The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. *Cell. Mol. Life Sci.* (in the press).
37. Clery, A. *et al.* An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res.* **35,** 1868–1884 (2007).
38. Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* **331,** 280–283 (1988).
39. Gaudin, C. *et al.* Structure of the RNA signal essential for translational frameshifting in HIV-1. *J. Mol. Biol.* **349,** 1024–1035 (2005).
40. Staple, D. W. & Butcher, S. E. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J. Mol. Biol.* **349,** 1011–1023 (2005).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank P. Thibault for updating MC-Sym and P. Gendron for helping us with the Condor and web services. We thank D. D'Amours, M.-F. Gaumont-Leclerc and V. Lisi for making suggestions to improve the manuscript. We thank D. H. Mathews and E. Westhof for discussions about MC-Fold. This project was supported by grants from the Canadian Institutes of Health Research (CIHR) and from the Natural Sciences and Engineering Research Council (NSERC) of Canada. M.P. holds Ph.D. scholarships from the NSERC and the Fonds Québécois de la Recherche sur la Nature et les Technologies. F.M. is a member of the Centre Robert-Cedergren of the Université de Montréal.

**Author Contributions** Both authors were involved in every aspect of the research. M.P. programmed MC-Fold and MC-Cons.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.M. (francois.major@umontreal.ca).