A benchmark of multiple sequence alignment programs upon structural RNAs

Paul P. Gardner*, Andreas Wilm[1] and Stefan Washietl[2]

Department of Evolutionary Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen Ø, Denmark, [1]Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany and [2]Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria

ABSTRACT

To date, few attempts have been made to benchmark the alignment algorithms upon nucleic acid sequences. Frequently, sophisticated PAM or BLOSUM like models are used to align proteins, yet equivalents are not considered for nucleic acids; instead, rather *ad hoc* models are generally favoured. Here, we systematically test the performance of existing alignment algorithms on structural RNAs. This work was aimed at achieving the following goals: (i) to determine conditions where it is appropriate to apply common sequence alignment methods to the structural RNA alignment problem. This indicates where and when researchers should consider augmenting the alignment process with auxiliary information, such as secondary structure and (ii) to determine which sequence alignment algorithms perform well under the broadest range of conditions. We find that sequence alignment alone, using the current algorithms, is generally inappropriate <50–60% sequence identity. Second, we note that the probabilistic method ProAlign and the aging Clustal algorithms generally outperform other sequence-based algorithms, under the broadest range of applications.

INTRODUCTION

Motivation

The use of multiple sequence alignments is an essential step for many RNA sequence analysis methods [e.g. RNA structure analysis (1–5), RNA homology search (6,7), non-coding RNA (ncRNA) detection (8,9) and RNA-based phylogenetic inference (10,11)]. Structural alignment of RNA is, however, an open problem. An algorithm for simultaneous structural RNA sequence alignment, structure prediction and phylogenetic reconstruction has been proposed (12), yet current implementations are limited in terms of functionality and sequence size (13–18). A second structural RNA alignment approach employs (predicted) structures and aligns these directly (19–21). Again these structures are generally limited in terms of sequence size, but primarily suffer from the inaccuracy of single sequence structure prediction (22–24) [although, pruning low-probability base pairs yields modest improvement (25,26)]. In addition, when structure is not conserved algorithms which attempt to include RNA structural information are likely to fail [e.g. methylation-guide snoRNAs, Air-RNA]. Many of these methods are also impractical when sequence length is large or when single sequence structure prediction (27,28) performs poorly (24).

The performance of current sequence alignment methods has been thoroughly analysed in terms of protein alignment accuracy (29–32). Benchmarking has also been performed for simulated non-coding DNA (33). However, the results of these studies do not explore methods for the specific problem of aligning structural RNAs. In this work, we extend these studies and test the performance of current alignment algorithms upon structural RNA datasets.

The aims of this work are 2-fold. First, to identify the 'twilight zone' of RNA sequence alignment—the homology range below which sequence alignment alone is unlikely to produce reliable results and researchers should seriously consider augmenting the alignment process with auxiliary information such as secondary structure. Second, to identify algorithms capable of reliably aligning structural RNA sequences under a range of sequence identities.

Alignment algorithms

The simplest form of an alignment is the pairwise sequence alignment. This can be performed by aligning sequences globally (34) or locally (35), both employ dynamic programming, thus resulting in a quadratic time complexity. Different scoring schemes may be used, which already produce varying results.

The alignment of multiple sequences is far more complex, as the mathematically optimal solution imposes exponential complexity. Therefore, heuristics are used, which do not guarantee an optimal solution, but perform multiple sequence alignment in reasonable time.

One common approach is called progressive alignment (36), which builds a multiple alignment from pairwise alignments. The idea is that an alignment of sequences, which have more recently diverged, is more likely to be reliable. Thus, high-scoring pairwise alignments are aligned first and next closely related sequences (or alignments of sequences) are added progressively. The order of this progressive alignment is, in most cases, defined by a guide tree, which is created beforehand from a distance matrix, produced by aligning all $n(n-1)/2$ possible pairs of sequences first. The basic drawback of this method is the fact that gaps introduced in an early step cannot be removed during the later addition of sequences, e.g. errors made in an early step propagate during the alignment process ('Once a gap always a gap'). The so-called iterative methods prevent this by realigning sequences or sequence groups in the multiple alignment, thus, theoretically optimizing the alignment until successive iterations fail to improve the alignment or reach a predefined limit (convergence).

Another class of algorithms is called consistency-based. Here, a multiple alignment is constructed by extracting (maximum-scoring) pairwise alignments from a library such that these combined pairwise solutions are not contradictory or mutually exclusive.

Probabilistic methods are an increasingly popular way of generating solutions to biological problems. The basic premise is to produce a model that one believes best describes the system behaviour. Model parameters are subsequently estimated from reliable data. In terms of sequence alignment, a pairwise hidden Markov model-based approach has been proposed (37), and now implemented and extended to multiple sequence alignment (38,39).

The structural RNA alignment approach of Sankoff (12) merges the recursions of Smith–Waterman (35) type sequence alignment and Nussinov *et al.* (maximal base-pairing) (40) or Zuker and Stiegler (energy-based) (41) RNA structure prediction (15). The basic idea is to implicitly include base-pairing interactions into the alignment procedure such that homologous base pairs are aligned correctly. Unfortunately, the algorithm is computationally expensive [$O(n^{3m})$ in time, and $O(n^{2m})$ in space, where n is the sequence length and m is the number of sequences]. Current implementations, Dynalign (13,14), Foldalign (16), PMcomp (15) and Stemloc (17,18), are restricted implementations of the Sankoff algorithm, which impose practical limits on the size or the shape of substructures. In addition, sensible score routines such as thermodynamics (13,14), a combination of sequence and thermodynamic scores (16) and partition-function derived probability matrices (15), can be used to score alignments.

Figure 1 shows a classification of all the programs used in this study into the above categories.

MATERIALS AND METHODS

Accuracy measures

In order to evaluate alignment methods on structural RNAs, we use two independent measures. First is the traditional
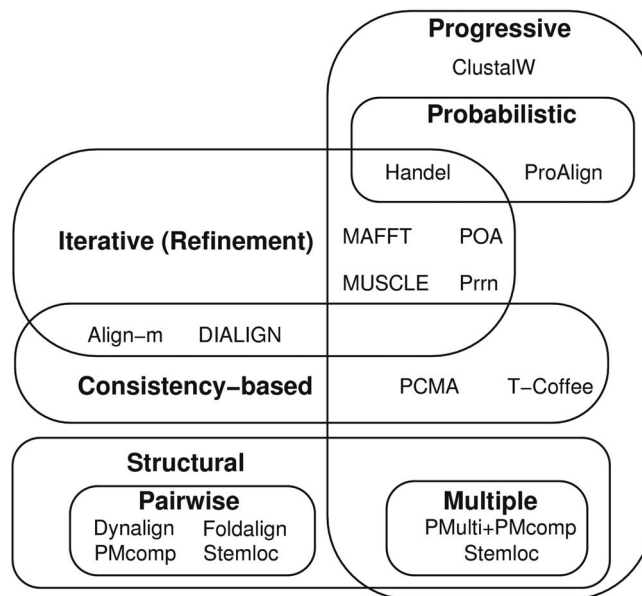

Figure 1. An overview of alignment programs used in this work. Programs were classified into the categories described in detail in Alignment algorithms section.

sum-of-pairs score (SPS) employed in many previous alignment benchmarks (29,32,33). SPS is defined as the fraction out of all possible character pairs that are aligned in both the predicted and the reference alignments. Perfectly predicted (concordant) alignments receive an SPS of one, absolutely imperfectly predicted (discordant) alignments receive an SPS of zero. Although, for nucleotide alignments an SPS of zero is rarely observed. Essentially, SPS provides a measure of the sensitivity of the prediction. The second measure, dubbed the structure conservation index (SCI), provides a measure of the conserved secondary structure information contained within the alignment (9). It is a derivative of the score calculated by the RNAalifold consensus folding algorithm (5,42) which is based upon the sum of a thermodynamic and a covariance term and, in contrast to the SPS, is independent from a reference alignment. The SCI is close to zero if RNAalifold identifies no common RNA structure in the alignment, whereas a set of perfectly conserved structures has an SCI ≈ 1. An SCI >1 indicates that there is a conserved RNA secondary structure which is, in addition, supported by compensatory and/or consistent mutations preserving the common structure. We note that the SCI scores alignment accuracy only in terms of the secondary structure information. For example, if the helices of a secondary structure are accurately aligned, it does not affect the SCI whether the loop regions are well aligned in terms of sequence similarity. The SCI specifically points out the structural aspect of alignment accuracy and, therefore, appears to be a useful measure in addition to the SPS.

Alignment programs

We tested 11 sequence alignment programs and 4 structural alignment programs (see Supplementary Material Table 1). Where previous alignment benchmarks have only considered the default or 'out of the box' behaviour we try testing a range of parameter combinations for each alignment method.

Algorithm options are summarized in the Supplementary Material Tables 2 and 3.

Test datasets

We generated four diverse structural RNA datasets of Group II introns, 5S rRNA, tRNA and U5 spliceosomal RNA. The sequences and the reference alignments for calculating the SPS were obtained from the Rfam v5.0 database. SRP from the SRPDB database was included in the original dataset but was later discarded due to poor comparability between predicted and structural alignments contained in this dataset (the results suggest that a fraction of the SRP reference sequences have been misaligned). Using the same procedure as described previously (42), we generated ∼100 sub-alignments for each family. The alignments contained five sequences each and encompassed a range of sequence identities. This large dataset (hereafter referred to as dataset 1) was divided into high (≥75% sequence identity, 73 alignments), medium (<75 and ≥55% sequence identity, 73 alignments) and low (<55% sequence identity, 242 alignments) sequence

homology groups. An additional tRNA dataset was generated with just two sequences to each alignment (hereafter referred to as dataset 2). This was used to contrast pairwise structural alignment methods and sequence alignment methods.

Caveat: tools improve

We note here that our data reflect the state of the art in early 2005. Most of the tools tested are relatively recent, and many are still under development. Hence, not all the observations below will remain reproducible. In fact, we hope this study helps to obtain better results in the future.

RESULTS

Dataset 1: applicability of pure sequence alignment to ncRNA sequences

All the 11 sequence alignment algorithms were tested upon dataset 1. The results and the relative algorithm ranks for each homology group are summarized in Table 1. We experimented

Table 1. The mean SCI score and mean SPSs computed for dataset 1 (see text for details)

Algorithm	High homology (75% ≤ seq. id)			Medium homology (75% < seq. id ≤ 55%)			Low homology (seq. id < 55%)		
	SCI	SPS	Rank	SCI	SPS	Rank	SCI	SPS	Rank
Structural	0.9789	1.0000	0	0.9297	1.0000	0	0.7846	1.0000	0
Align-m (1)	0.9827	0.9600	**5**	0.8453	0.8825	22	0.4957	0.6748	25
Align-m (2)	0.9827	0.9600	**6**	0.8453	0.8825	21	0.4957	0.6748	24
Align-m (3)	0.9778	0.9593	**7**	0.8040	0.8742	26	0.4691	0.6635	29
Align-m (4)	0.9778	0.9593	**8**	0.8040	0.8742	25	0.4691	0.6635	28
Align-m (5)	0.8995	0.9419	30	0.7597	0.8583	30	0.4777	0.6460	30
Clustal	0.9438	0.9741	**9**	0.9100	0.9194	**2**	0.6064	0.7423	**8**
Clustal (qt)	0.9315	0.9743	12	0.8996	0.9123	**9**	0.6076	0.7345	**9**
DIALIGN	0.9071	0.9577	27	0.8018	0.8712	27	0.4979	0.6659	26
DIALIGN (o)	0.9077	0.9601	26	0.8568	0.8860	20	0.5202	0.6721	23
DIALIGN (it)	0.8590	0.9491	34	0.7519	0.8556	31	0.4669	0.6546	32
DIALIGN (it,o)	0.8492	0.9486	35	0.7092	0.8456	33	0.4467	0.6467	33
Handel	0.9604	0.9560	11	0.8570	0.8954	19	0.5360	0.7283	19
MAFFT (fftns)	0.8321	0.9145	36	0.5864	0.8030	37	0.3538	0.6448	37
MAFFT (fftnsi)	0.8840	0.9427	32	0.6655	0.8378	36	0.3845	0.6634	36
MAFFT (nwns)	0.9297	0.9502	25	0.6712	0.8349	35	0.3941	0.6724	35
MAFFT (nwnsi)	0.9330	0.9526	22	0.7004	0.8451	34	0.4071	0.6812	34
MUSCLE	0.9222	0.9684	19	0.8988	0.9181	**5**	0.6065	0.7668	**2**
MUSCLE (nj)	0.9268	0.9707	16	0.8841	0.9110	17	0.5902	0.7503	12
MUSCLE (mi32)	0.9222	0.9683	21	0.8959	0.9167	**8**	0.6069	0.7666	**1**
MUSCLE (mi32,mt6)	0.9222	0.9683	20	0.8959	0.9167	**7**	0.6068	0.7664	**3**
MUSCLE (nj,mt6)	0.9268	0.9707	15	0.8841	0.9110	16	0.5902	0.7503	11
MUSCLE (nj,mi32)	0.9268	0.9708	14	0.8855	0.9112	14	0.5897	0.7501	14
MUSCLE (nj,mi32,mt6)	0.9268	0.9708	13	0.8855	0.9112	13	0.5898	0.7501	13
PCMA	1.0030	0.9635	**3**	0.9196	0.9059	**3**	0.5339	0.6890	20
PCMA (agi20)	1.0030	0.9635	**2**	0.9255	0.9068	**1**	0.5621	0.7058	16
PCMA (agi60)	1.0030	0.9635	**1**	0.8938	0.8941	18	0.5270	0.6827	21
POA	0.8478	0.9644	33	0.7666	0.8739	29	0.4656	0.6740	27
POA (g)	0.9253	0.9722	17	0.8836	0.9130	15	0.5581	0.7543	15
POA (p)	0.8668	0.9656	31	0.7814	0.8814	28	0.5079	0.6964	22
POA (gp)	0.9444	0.9726	**10**	0.8929	0.9188	**10**	0.5843	0.7726	**6**
ProAlign (bw400)	0.9978	0.9631	**4**	0.9163	0.9072	**4**	0.6045	0.7490	**5**
Prrn	0.9364	0.9458	24	0.9036	0.9064	11	0.5903	0.7549	**10**
Prrn (S10)	0.9371	0.9458	23	0.8997	0.9086	12	0.5964	0.7596	**4**
T-Coffee	0.8867	0.9656	29	0.8129	0.8989	24	0.5322	0.7337	18
T-Coffee (c)	0.9201	0.9733	18	0.8956	0.9194	**6**	0.5972	0.7543	**7**
T-Coffee (f)	0.8867	0.9656	28	0.8129	0.8989	23	0.5322	0.7337	17
T-Coffee (s)	0.7892	0.9536	37	0.7151	0.8637	32	0.4447	0.6934	31

Dataset 1 has been divided into three homology groups: the high-homology group (75% ≤ seq. id), the medium-homology group (75% < seq. id ≤ 55%) and the low-homology group (seq. id < 55%). Rankings are computed from the product of SCI and SPS. The top 10 ranks are highlighted in boldface. Abbreviations of the parameter switches used to produce these results are shown in parentheses. Further details of these can be found in Supplementary Material.

with a variety of algorithm parameters. The results using default and any parameter combinations that increased algorithm performance are summarized in Figure 2. In order to measure relative algorithm performances, a ranking for each algorithm (and parameter setting) was calculated within each of the three homology ranges. The rank is based upon the product of mean SCI score and mean SPSs.

For the high-similarity datasets (sequence identity ⩾75%), there is little difference in accuracy across most of the algorithms considered here (see Table 1). Align-m and Handel rank well on this dataset, yet the relative performance of both these methods dropped rapidly with decreasing sequence homology. Interestingly, for Align-m, this is the opposite to what has been observed for protein-based results (43).

PCMA ranked well on both the high- and medium-similarity datasets; however, the relative performance of this method dropped on the low-similarity dataset.

ClustalW, MUSCLE, PCMA, POA [with both global and progressive modes—hereafter referred to as POA (gp)], ProAlign, Prrn and T-Coffee (when ClustalW is used to generate a library of pairwise alignments) perform comparatively well across all homology ranges, with little significant variation between these methods. Only ClustalW, ProAlign and POA (gp) consistently ranked in the top 10 across all the datasets. There is some redundancy in ranking algorithms over all the combinations of algorithm parameters we use here, obviously some combinations produce similar (or even identical) results. However, this has had little impact on our conclusions.

The results suggest that 60% sequence identity is a crude threshold, whereby the structural content of predicted sequence alignments diverges from reference structural alignments (see Figures 2 and 3).

Dataset 2: comparison of structural and sequence methods

Now, we contrast the relative performance of the comparatively good sequence-based methods identified in the previous section with structural alignment methods using a smaller tRNA dataset. The structure-based methods are generally computationally more intensive than the sequence-based methods—hence the small size (in terms of the number of sequences and the sequence length) of this dataset.

We use dataset 2 to compare the relative performances of structure-based methods (e.g. Dynalign, Foldalign, PMcomp and Stemloc) to the 'better' sequence-based methods identified in the previous section (e.g. ClustalW, MUSCLE, PCMA, POA (gp), ProAlign and Prrn). We observe a dramatic divergence in relative performances below ∼60% sequence identity between the structure- and the sequence-based methods (see Figure 3).

The structural methods Dynalign, Foldalign and PMcomp show high conservation of structural information (SCI) across all homology ranges. However, the SPS of Dynalign and PMcomp are significantly lower than that of Foldalign. The difference is mostly marked in the high to medium homology range. This is possibly because the current versions of Dynalign and PMcomp optimize scores solely based on secondary structure information and hence are likely to produce rather different alignments to those used in the test dataset, where sequence information is also included. Stemloc performs comparatively well in terms of SCI for sequence identities >40%. In terms of SPS, however, Stemloc behaves much like the purely sequence-based methods. Given the apparent sophistication of this method and large computational resources required to run the algorithm (17,18) these results are rather
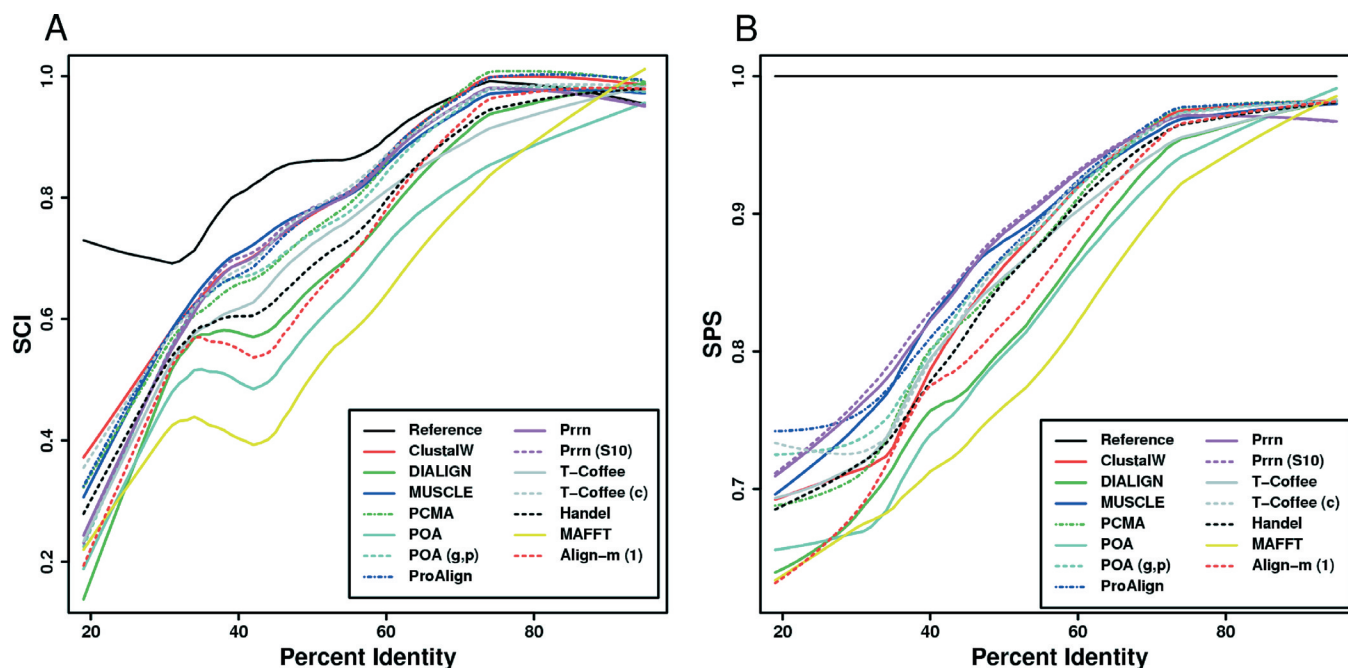

Figure 2. Both measures of structural RNA alignment correctness, SCI (**A**) and SPS (**B**), are plotted as functions of the mean pairwise sequence identity (calculated using the reference alignments). The curves are fit to dataset 1 (see text for details) using lowess (local weighted regression) smoothing. At most, two curves are plotted for each alignment package—one corresponding to the default parameters, the other corresponds to the best parameter combination we could identify.
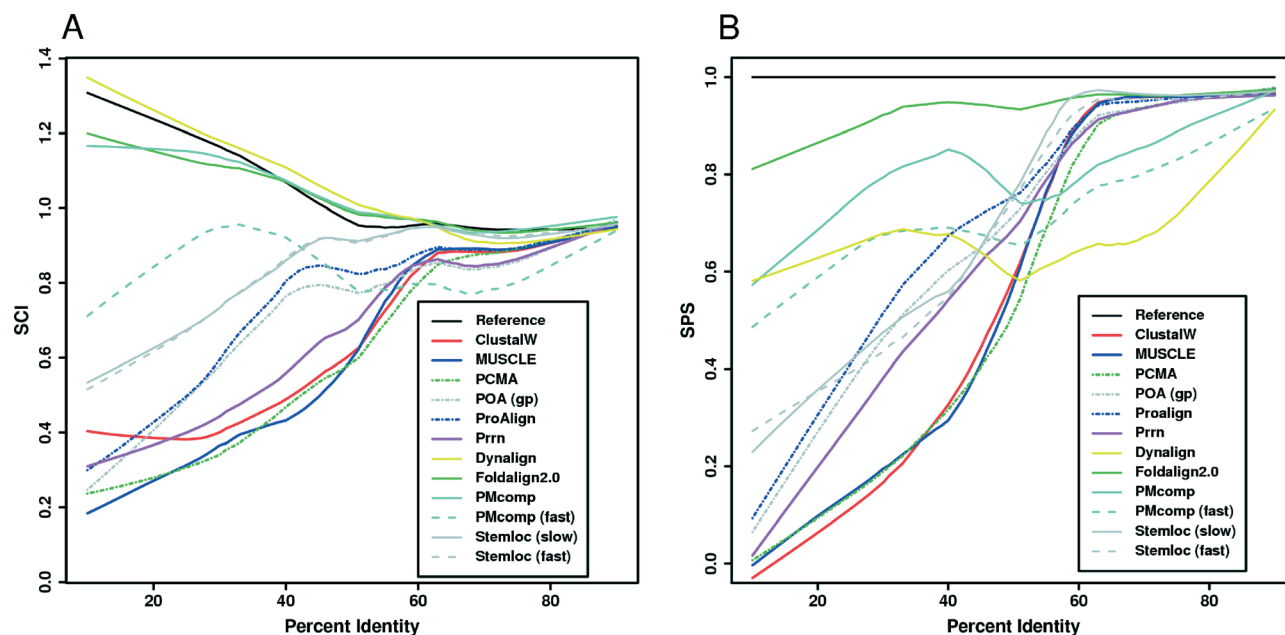
Figure 3. SCI (**A**) and SPS (**B**) as functions of the sequence identity for dataset 2 (see text for details). Five structural algorithms are shown: the Sankoff-based methods Dynalign, Foldalign, PMcomp and Stemloc, and the base pair probability profile alignment method implemented in PMcomp (fast). These are in contrast with the hand-curated structural alignments and six of the better sequence-based alignment algorithms (ClustalW, MUSCLE, PCMA, POA (gp), ProAlign and Prrn).

disappointing. We hypothesize that perhaps these are due to an overemphasis upon the sequence-based component of the algorithm.

All of the above datasets are freely available from http://www.binf.ku.dk/users/pgardner/bralibase/. As novel and updated algorithms become available, updated results will also be made available from the web page.

DISCUSSION

An aim of this work is to determine the boundaries between when pure sequence alignment methods perform well and when augmentation of the alignment with structure is necessary. We wish to highlight that the benchmarks based purely upon structural protein alignments do not adequately test all the uses of sequence alignment. In addition, we are pleased to note that our two independent measures of alignment fitness (SCI and SPS) produce similar results.

In some cases, we found that altering algorithm parameters produced a dramatic improvement over the defaults (e.g. T-Coffee performance improves using Clustal to generate a library of pairwise alignments and POA performance improves dramatically using a combination of the global and the progressive modes).

We find that the conclusions of previous studies based upon structural protein alignments do not necessarily hold for the alignment of structural ncRNA. For example, DIALIGN, identified as a method which performed well for low-homology protein alignment did not generally improve (relative to the alternative methods) on low-homology datasets (32). Another surprising discovery was that T-Coffee, touted as an excellent method for high-homology datasets, did not perform well (again, relative to the alternative methods) on the ncRNA

datasets (32). Another surprise was that the supposedly outdated, yet still widely used method ClustalW, performed consistently well across all homology datasets. This is possibly a consequence of the fact that more recent algorithms are heavily optimized for protein alignment. The relatively new methods ProAlign, POA (gp) and MUSCLE also performed consistently well. ProAlign, in particular, produced (comparatively) reliable alignments and ranked in the top 5 across all homology ranges. This is possibly due to the fact that ProAlign is one of the few algorithms to use a scoring scheme derived from reliable nucleic acid sequence alignments. The performance of POA (gp) is also remarkable, not only because it employs a very fast method [said to accurately align 5000 EST sequences in 4 h on a Pentium II (44)] but also because it performed consistently well over all test sets.

Another conclusion of this work is that the 'twilight zone' of ncRNA alignment—the homology range where little to no structural information of predicted alignments (using the current state of the art algorithms) for structurally homologous sequences is retained—is in the 50–60% sequence-identity range. This is dramatically higher than that of the protein sequences which is 10–20% (29). Much of this difference is, of course, due to the different alphabet sizes and the generally limited models and the score matrices for nucleotide alignment.

It is interesting to note that three of the structural methods (Dynalign, Foldalign and PMcomp), for a short homology range (40–60% sequence identity), have higher SCI scores than the reference alignment and that in the same regions there is a dip in the performance when Dynalign, Foldalign and PMcomp performance is measured using SPS. This suggests that the reference alignments themselves may be improved upon in this homology range.

Based upon these results the Foldalign score routines seem to have optimized the delicate balance between the sequence and the structure-based scores. This implementation of Sankoff's algorithm employs a light-weight energy model (13,41,45,46) in concert with the substitution matrices similar to those of RIBOSUM (47) and BLOSUM (48), which seem to produce excellent predictions. However, the computational complexity of this algorithm is still an issue, global alignment is restricted to sequences of ∼200 nt or less, in practice. Further optimization may increase this bound, however.

The profile-based approach of Hofacker *et al.* [pmcomp - -fast (15,49)], holds promise for producing fast and reasonably accurate alignments in satisfactory time across all homology ranges. It by no means produces 'optimal' alignments in terms of sequence or structure, but is a reasonable compromise between the sequence- and the structure-based methods in terms of improved accuracy for the former and dramatically reduced computational requirements for the latter. This method is in the process of being re-implemented in C with affine gap costs and an adjustable sequence-weighting parameter. This is available as 'RNApaln' with the Vienna package version 1.5 or greater (I. Hofacker, personal communication).

SUMMARY

The results and main conclusions of our study can be summarized as follows:

 (i) The two independent measures of global alignment accuracy SPS and SCI are generally in agreement. These measures only differ significantly on methods, such as Dynalign and PMcomp, that perform only structural alignment. The SCI is, therefore, a useful score for assessing the accuracy of structural RNA alignments.
 (ii) The relative performance of multiple sequence alignment programs on RNA alignments can differ remarkably from the performance observed on protein alignments.
(iii) The multiple sequence alignment algorithms, such as ClustalW, MUSCLE, PCMA, POA (gp), ProAlign and Prrn, perform well on high- to medium-homology datasets.
(iv) ClustalW, ProAlign and POA (gp) consistently ranked in the top 10 across all homology ranges.
 (v) The 'twilight zone' of ncRNA alignment is in the 50–60% sequence-identity range.
(vi) Below this limit, algorithms incorporating structural information (Dynalign, Foldalign, PMcomp and Stemloc) outperform pure sequence-based methods. However, these algorithms are computationally demanding which severely limits their use in practice.

Future directions

One rather interesting result of this study is that the structure profile alignment method (pmcomp - -fast) produces reasonable structural alignments across all homology ranges in a dramatically short time. This method in combination with, as yet undeveloped, iterative alignment refinement strategies seems poised to become a method of choice for RNA researchers in the near future. This also has interesting implications for the notoriously difficult problem of ncRNA homology search. A combination of a database of locally stable regions (50)

(analogous to the index creation of the BLAST procedure) and the profile alignment method is likely to produce a superior homology detection tool. This application and the extension of this method to multiple alignments is an area of active research.

Other potentially fruitful research areas to explore are as follows: (i) The implementation of light-weight Sankoff-like algorithms, which produce reasonable alignments in a short time-frame and use score routines combining energy and sequence scores similar to those of Foldalign. (ii) In analogy to the improvement of structure prediction accuracy by including stacking parameters (nearest-neighbour model), perhaps the alignment of RNA sequences over a dinucleotide alphabet could produce improvements in sequence-based alignment. (iii) The current score matrices for nucleotide alignment are generally *ad hoc*, it is likely that significant improvements could be gained by using RIBOSUM-like matrices (47) for scoring alignment. (iv) 'Intelligent' alignment algorithms which employ sequence information when this is reasonable or structure alignment when this is better.

NOTE ADDED TO PROOF

Since embarking on this project the alignment algorithms Align-m, Handel and MAFFT have been updated. Preliminary analysis of the updated algorithms suggests that improvements to Align-m and Handel have resulted in modest performance increases across the high, medium and low similarity groups of data-set 1. The improvements to MAFFT however have resulted in major performance increases across all similarity groups of data-set 1. In fact, across all similarity groups MAFFT (ver. 5) now ranks second only to \proalign. However, gap-parameters for this algorithm have been estimated directly from data-set 1, this bias could be alleviated by determining optimal gap-parameters for all methods prior to benchmarking. Preliminary work in this direction shows algorithm performances on RNAs can in some cases be enhanced by optimising gap-parameters (personal communication K. Katoh).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES

1. Chiu,D.K. and Kolodziejczak,T. (1991) Inferring consensus structure from nucleic acid sequences. *Comput. Appl. Biosci.*, **7**, 347–352.
2. Gutell,R.R., Power,A., Hertz,G.Z., Putz,E.J. and Stormo,G.D. (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.*, **20**, 5785–5795.
3. Gorodkin,J., Heyer,L., Brunak,S. and Stormo,G. (1997) Displaying the information contents of structural RNA alignments. *CABIOS*, **13**, 583–586.
4. Knudsen,B. and Hein,J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.*, **31**, 3423–3428.
5. Hofacker,I., Fekete,M. and Stadler,P. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
6. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
7. Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A. and Eddy,S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.
8. Rivas,E. and Eddy,S. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, **2**, 8.
9. Washietl,S., Hofacker,I. and Stadler,P. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
10. Woese,C. and Fox,G. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA*, **74**, 5088–5090.
11. Hudelot,C., Gowri-Shankar,V., Jow,H., Rattray,M. and Higgs,P. (2003) RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. *Mol. Phylogenet. Evol.*, **28**, 241–252.
12. Sankoff,D. (1985) Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J. Appl. Math.*, **45**, 810–825.
13. Mathews,D. and Turner,D. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, **317**, 191–203.
14. Mathews,D. (2005) Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics*, in press.
15. Hofacker,I., Bernhart,S. and Stadler,P. (2004) Alignment of RNA base pairing probability matrices. *Bioinformatics*, **20**, 2222–2227.
16. Hull Havgaard,J., Lyngsø,R., Stormo,G. and Gorodkin,J. (2005) Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%. *Bioinformatics*, in press.
17. Holmes,I. (2004) A probabilistic model for the evolution of RNA structure. *BMC Bioinformatics*, **5**, 166.
18. Holmes,I. (2005) Accelerated probabilistic inference of RNA structure evolution. *BMC Bioinformatics*, **6**, 73.
19. Sczyrba,A., Kruger,J., Mersch,H., Kurtz,S. and Giegerich,R. (2003) RNA-related tools on the Bielefeld Bioinformatics Server. *Nucleic Acids Res.*, **31**, 3767–3770.
20. Höchsmann,M., Töller,T., Giegerich,R. and Kurtz,S. (2003) Local similarity of RNA secondary structures. In *2nd IEEE Computer Society Bioinformatics Conference (CSB 2003), 11–14 August 2003, Stanford, CA, USA. ISBN 0-7695-2000-6*, pp. 159–168.
21. Siebert,S. and Backofen,R. (2003) MARNA: a server for multiple alignment of RNAs. In *Proceedings of the German Conference on Bioinformatics, GCB 2003, October 12–14, 2003, Neuherberg/Garching near Munich, Germany*, pp. 135–140.
22. Konings,D. and Gutell,R. (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA*, **1**, 559–574.
23. Fields,D. and Gutell,R. (1996) An analysis of large rRNA sequences folded by a thermodynamic method. *Fold Des.*, **1**, 419–430.
24. Doshi,K., Cannone,J., Cobaugh,C. and Gutell,R. (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*, **5**, 105.
25. Mathews,D. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178–1190.
26. Gardner,P. and Giegerich,R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*, **5**, 140.
27. Zuker,M. and Sankoff,D. (1984) RNA secondary structures and their prediction. *Bull. Math. Biol.*, **46**, 591–621.
28. Hofacker,I.L., Fontana,W., Bonhoeffer,S. and Stadler,P.F. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie*, **125**, 167–188.
29. Thompson,J., Plewniak,F. and Poch,O. (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res.*, **27**.
30. Thompson,J., Plewniak,F. and Poch,O. (1999) BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. *Bioinformatics*, **15**, 87–88.
31. Bahr,A., Thompson,J., Thierry,J. and Poch,O. (2001) BAliBASE (Benchmark Alignment dataBASE): enhancements for repeats, transmembrane sequences and circular permutations. *Nucleic Acids Res.*, **29**, 323–326.
32. Lassmann,T. and Sonnhammer,E. (2002) Quality assessment of multiple alignment programs. *FEBS Lett.*, **529**, 126–130.
33. Pollard,D., Bergman,C., Stoye,J., Celniker,S. and Eisen,M. (2004) Benchmarking tools for the alignment of functional noncoding DNA. *BMC Bioinformatics*, **5**, 6.
34. Needleman,S. and Wunsch,C. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
35. Smith,T. and Waterman,M. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
36. Feng,D. and Doolittle,R. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, **25**, 351–360.
37. Durbin,R., Eddy,S., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis: Probabilistic Models of Protein and Nucleic Acids*. Cambridge University Press, Cambridge.
38. Holmes,I. and Bruno,W.J. (2001) Evolutionary HMMs: a Bayesian approach to multiple alignment. *Bioinformatics*, **17**, 803–820.
39. Löytynoja,A. and Milinkovitch,M.C. (2003) A hidden Markov model for progressive multiple alignment. *Bioinformatics*, **19**, 1505–1513.
40. Nussinov,R., Piecznik,G., Grigg,J.R. and Kleitman,D.J. (1978) Algorithms for loop matchings. *SIAM J. Appl. Math.*, **35**, 68–82.
41. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
42. Washietl,S. and Hofacker,I. (2004) Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J. Mol. Biol.*, **342**, 19–30.
43. Van Walle,I., Lasters,I. and Wyns,L. (2004) Align-m: a new algorithm for multiple alignment of highly divergent sequences. *Bioinformatics*, **20**, 1428–1435.
44. Lee,C., Grasso,C. and Sharlow,M.F. (2002) Multiple sequence alignment using partial order graphs. *Bioinformatics*, **18**, 452–464.
45. Mathews,D., Sabina,J., Zuker,M. and Turner,H. (1999) Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
46. Xia,T., SantaLucia,J., Burkard,M., Kierzek,R., Schroeder,S., Jiao,X., Cox,C. and Turner,D. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **4**, 14719–14735.
47. Klein,R. and Eddy,S. (2003) RSEARCH: finding homologs of single structured RNA sequences. *BMC Bioinformatics*, **4**, 44.
48. Henikoff,S. and Henikoff,J. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
49. Bonhoeffer,S., McCaskill,J., Stadler,P. and Schuster,P. (1993) RNA multi-structure landscapes. A study based on temperature dependent partition functions. *Eur. Biophys. J.*, **22**, 13–24.
50. Hofacker,I., Priwitzer,B. and Stadler,P.F. (2004) Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics*, **20**, 186–190.