Simulation of non-specific protein–mRNA interactions

James Magee and Jim Warwicker1,*

School of Chemical Engineering and Analytical Science, University of Manchester, Jackson’s Mill, PO Box 88, Sackville Street, Manchester M60 1QD, UK and 1Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

Received September 9, 2005; Revised and Accepted November 8, 2005

ABSTRACT

Protein–nucleic acid interactions exhibit varying degrees of specificity. Relatively high affinity, sequence-specific interactions, can be studied with structure determination, but lower affinity, non-specific interactions are also of biological importance. We report simulations that predict the population of nucleic acid paths around protein surfaces, and give binding constant differences for changes in the protein scaffold. The method is applied to the non-specific component of interactions between eIF4Es and messenger RNAs that are bound tightly at the cap site. Adding a fragment of eIF4G to the system changes both the population of mRNA paths and the protein–mRNA binding affinity, suggesting a potential role for non-specific interactions in modulating translational properties. Generally, the free energy simulation technique could work in harness with characterized tethering points to extend analysis of nucleic acid conformation, and its modulation by protein scaffolds.

INTRODUCTION

Non-specific interactions between protein and nucleic acid are important determinants of biological function, e.g. in combining with specific interactions to form distorted binding geometries (1) or facilitating reduced dimensionality searches for binding sites (2). Major questions remain to be resolved concerning packaging and unpackaging of nucleic acids in many systems, including chromatin (3) and viruses (4). Structural biology provides much information about specific protein–nucleic acid complexes (5), including the ribosome (6), whilst atomic-level molecular models can be used in relatively short simulations (7). It has been apparent since the first structure determinations for nucleic acid binding proteins that surface charge plays a significant role in binding. This property is currently being used to aid the identification of DNA-binding surfaces in proteins emerging from Structural Genomics initiatives (8,9), and in studies of DNA wrapping paths around proteins (10). Methods for computational analysis of protein–nucleic acid interactions at a reduced level of complexity are being introduced (11,12).

The mRNA cap-binding protein eIF4E has central roles in several aspects of post-transcriptional gene expression (13). In cap-dependent translation initiation, eIF4E binds the mRNA cap before ribosome docking, in a larger complex that includes eIF4G (14). Transport of mRNAs from the nucleus also make use of the capped mRNA-binding potential of eIF4E (15). The cap binds to a groove in eIF4E, an interaction which is modulated by a variety of binding partners (16) that contact a region on the opposite side of eIF4E to the cap site (17). The molecular mechanism of interaction between these binding sites is not yet fully understood. Binding of a fragment of eIF4G (393–490) triggers a coupled folding event with the eIF4E N-terminus (18). There is no major structural change at the cap site, although some chemical shift changes may relate to modulation of binding affinity (19). Interaction between the two sites can occur with cap in the absence of an mRNA body, demonstrated for eIF4G fragments (19), and with a substantial reduction in cap analogue binding upon eIF4E association with a number of factors, including promyelocytic leukemia protein (20). Interactions between mRNA and eIF4E have also been studied in the absence of additional factors. Although NMR measurements of eIF4E binding to capped trinucleotide RNAs show no detectable contacts beyond the first nucleotide following the cap structure (21), sequence alteration close to the cap causes variation in affinity (22), presumably due to non-specific protein–RNA interactions. Studies of eIF4E association with cap analogues of varying phosphate chain length, with changes in ionic strength and with engineered protein surface charge to mimic phosphorylation events (23), indicate that electrostatic interactions are important in this system (24). Brownian dynamics simulations have been used to interpret the salt-dependence of cap-eIF4E binding (25).

We report a Monte Carlo simulation approach, using a type of ‘phase switch’ algorithm (26), which allows prediction of the free energy difference between two systems, as opposed to the intractable problem of deriving free energy for a single
system in isolation using Monte Carlo. The ‘phase switch’ method is used with a coarse-grained RNA model and either free or eIF4G-bound eIF4E. The presence of the eIF4G fragment is predicted to have a small, but significant effect (via non-specific interactions) on capped mRNA-binding, accompanied by alteration of the preferred pathways of mRNA in the simulations. We further investigate alteration of surface charge for two modelled Schizosaccharomyces pombe eIF4Es (27).

MATERIALS AND METHODS

In our calculation scheme, eIF4E protein and eIF4E complexed to an eIF4G fragment (18) are modelled as fixed units with impenetrable volumes. The eIF4E protein alone was trimmed down to that part which is structured in the absence of eIF4G (17). Each protein unit is surrounded by a fixed electrostatic potential field, calculated with the Finite Difference Poisson Boltzmann (FDPB) method (28). Ionsensitive group charge was assigned at neutral pH with model compounds pK_s (29), and 0.15 M ionic strength was used. The mRNA chain is modelled as a chain of tangent hard spheres, diameter σ = 7 Å (i.e. nearest neighbour bond length fixed at 7 Å), each monomer carrying a charge of q = −1e, where e is the proton charge. Non-bonded monomers interact via the hard-core screened Yukawa potential, given by:

\[
u(r) = \begin{cases} \frac{-q^2 \sigma}{4 \pi \varepsilon_0 r} \exp \left(-\kappa \frac{r}{\sigma} \right) & r \leq \sigma \\ 0 & r > \sigma \end{cases}
\]

with an inverse screening length \(\kappa\) due to counterions that matches the FDPB calculations of protein electrostatic potential, and dielectric constant \(\varepsilon = 78.4\) for the solvent, with \(\varepsilon_0\) the vacuum permittivity.

We have performed simulations for single free hard core Yukawa chains to calculate the persistence length, \(L_p\), for comparison with experimental results. We have measured the mean end-to-end distance, \(R_e\), for these polymers, and estimated the persistence length via the relationship expected from the Worm-Like Chain model (30,31):

\[
\langle R_e^2 \rangle = 2L_p(1 - L_p(1 - \exp(-L/L_p)))/L
\]

where \(L\) is the contour length (end-to-end length at full extension) for the polymer. For chains at 311 K, this gives an estimate for the persistence length of 3.1 nm. By comparison, microscopy, single molecule stretching and spectroscopic methods yield experimental results from <1 nm to around 10 nm, for single-stranded RNA and DNA (31–36). The shorter persistence lengths mostly relate to single molecule experiments, and the larger values to bulk measurements.

For our non-specific binding simulations, the model mRNA chain, which is tethered to eIF4E at the cap site, moves within the protein electrostatic potential, discretized on a 2 Å grid. Protein excluded volume is also recognized with a grid map. Standard polymer Monte Carlo moves (37) (crankshaft, pivot and CCB) are used to generate mRNA chain configurations. In addition, every 1000 configurational move attempts a switch move (26) is attempted. This swaps the protein unit with which the mRNA chain is interacting, according to the standard Metropolis Monte Carlo acceptance probability (38), \(P_{acc} = \min[1, \exp(-\Delta E/k_B T)]\), where \(\Delta E\) is the change in energy for this move. At the end of the simulation, the free energy difference between the two systems can be estimated as \(\Delta G_{A→B} = -k_B T \ln(n_B/n_A)\), where \(k_B\) is the Boltzmann constant, \(T\) is the temperature and \(n_x\) is the number of sampled configurations where the mRNA was attached to protein unit \(x\). Since the system is held at constant volume (infinite dilution), the enthalpy difference between the two systems, \(\Delta H_{A→B}\), can be estimated as the difference in average interaction energy for each system. This also allows access to the entropy difference between the two systems, \(\Delta S_{A→B}\), since \(\Delta G_{A→B} = \Delta H_{A→B} - T\Delta S_{A→B}\). To avoid problems of metastability, simulations are performed within a parallel tempering framework (38), using ten simulation boxes distributed according to a power law across a range in temperature (39). After an initial equilibration run of \(5 \times 10^7\) move attempts, data collection runs are made over \(10^8\) move attempts, with data recorded every \(5 \times 10^4\) move attempts. During post-processing, a smooth interpolation of data across the studied temperature range is achieved using the multiple histogram method (40). Over the course of each simulation, we record how many times monomer centres lie within each potential map cell, which allows us to build up a probability density map for mRNA configuration.

RESULTS

Non-specific RNA–protein interactions in the eIF4E/4G system

One end of RNA polymer is fixed to eIF4E in all simulations, modelling cap binding (Figure 1). Monte Carlo sampling of RNA polymer conformation is made for two related systems, with a switch move allowing evaluation of the protein–RNA binding energy difference, \(\Delta G\), between the two systems. Figure 2A shows the calculated \(\Delta G\) for mRNA-binding upon addition of the eIF4G fragment (18), with variation in RNA length and temperature. The binding differential saturates as mRNA length is increased to 40 monomers, with a value of 1.5 kJ/mol at 310 K. Although this represents a relatively small impact on a binding equilibrium (a factor of ~1.8), it is accompanied by a substantial change in RNA paths brought about by eIF4G fragment binding to eIF4E.

![Figure 1](https://academic.oup.com/nar/article-abstract/33/21/6694/2401154)

Figure 1. Simulation schematic. Modelled mRNA (yellow beads) moves within the protein excluded volumes and electrostatic potential fields (red, blue) of the eIF4E and eIF4E/eIF4G fragment systems. Switch moves test the energetics of swapping between these systems.
A relatively small free energy change coupled to an observable redistribution in mRNA paths is consistent with a delicate balance of enthalpic and entropic components (Figure 2A).

**Differences between S. pombe eIF4E1 and eIF4E2 interactions with mRNA**

Our analysis has been extended to include two eIF4E homologues from *S. pombe*, for which we have previously made structural models (41). In this case we use just that part of eIF4E that is structured in the absence of eIF4G, and the binding energy difference is now calculated between the two *S. pombe* eIF4E isoforms. The binding differential is again saturated at a length of 40 monomers, with binding more favourable for eIF4E1 by about 5 kJ/mol at 300 K, corresponding to a change in binding of around 7-fold (Figure 3A). The extent to which enthalpic and entropic components balance to give a small change in free energy is also evident. Monomer density plots (Figure 3B) suggest that mRNA occupies significantly different conformational space in the two systems. These densities follow the positive protein electrostatic potential, whether this be extended (eIF4E1) or more concentrated in a surface patch (eIF4E2). The mRNA model includes screened monomer–monomer interactions, which generates a stiffness (calculated persistence length of 3.1 nm for free mRNA) that probably contributes to mRNA following elongated paths where these are available (eIF4E1).

**DISCUSSION**

The alteration of surface charge mediated by phosphorylation of eIF4E on S209 may play a role in the control of translation initiation (23), and appears to be involved in regulation of transformation and mRNA transport properties of eIF4E (42). The scale of binding changes that are likely to underpin these physiological effects can be assessed from *in vitro*...
measurements. Phosphorylation at S209 gives about a 2.5-fold change for cap analogue binding and 5-fold for capped mRNA (23). Charge mutation at S209 has not been simulated in the current study since this is close to the cap site and would be a less effective probe of non-specific interactions. However the alteration of binding equilibria that we predict for capped mRNA interactions, about 2-fold for eIF4G/eIF4E compared to eIF4E and about 7-fold difference between S.pombe eIF4E1 and eIF4E2, are of the same order as those measured for S209 phosphorylation. We therefore suggest that modulation of protein surface charge could provide a physiologically relevant mechanism for fine-tuning the mRNA-binding affinities of eIF4E variants.

The model excludes several factors that play roles in eIF4E–mRNA interactions, such as possible allosteric changes associated with effector binding to eIF4E (18), additional (sequence-specific) mRNA tethering interactions mediated by some eIF4E-binding proteins (16,43,44), the extent of secondary structure within the 5'-untranslated region (5'-UTR) and RNA helicase activity associated with eIF4A, and poly(A)-binding protein (13). With regard to RNA secondary structure, it is possible that eIF4E surface charge could interact differentially according to the density of RNA negative charge. For example, the less linear concentration of positive charge on S.pombe eIF4E2 relative to eIF4E1 may interact preferentially with more compact/structured mRNA segments. This would be consistent with the observation that the presence of eIF4E2 modifies the competence of S.pombe ribosomes to translate mRNAs with structured leaders (27). Our calculations also suggest that mRNA binding affinity will vary with changes of intracellular salt concentrations that may result from externally applied salt stress (45).

Figure 3. Binding of mRNA to S.pombe eIF4E1/E2. (A) Temperature-dependence of binding differential, with ΔG_E2→E1 plotted for variable polymer length. Simulations have been performed up to high (non-physiological) temperatures to facilitate the parallel tempering algorithm in allowing configurations to escape metastable ‘traps’. More favourable binding to S.pombe eIF4E1 is denoted by negative ΔG. (B) eIF4E surfaces with mRNA monomer density plots, as detailed in Figure 2, with spacefilled cap analogue (green).
Since we are examining the effects of non-specific mRNA–protein interactions, the complexities of individual nucleotides are replaced with a chain of identical, spherical monomers. The internal (to the nucleotide) degrees of freedom ignored by this model will be approximately the same in the systems between which the free energy differences are being taken; as such, they should make only a small contribution to the measured binding differences. The switch algorithm used to find the free energy differences is entirely separate from the approximations made in the modelling process. In order for the algorithm to function, it is necessary only that the switch move is accepted often enough to give statistically significant sampling of both systems. This type of algorithm has proven useful in the field of chemical physics (26), and has other possible applications in computational molecular biology. The simulations reported here, and related work (12), add quantitative prediction to qualitative arguments based solely on matching of positively-charged protein regions with negatively-charged phosphate backbone. The details of protein surface charge can be incorporated with FDPB calculations, whilst the entropic cost of restraining nucleic acid conformation is included with polymer modelling. This technique is suitable for study of other systems where a nucleic acid molecule is tethered to a protein unit, such as mRNA decapping enzymes. The method can also be extended, (with appropriate revision of nucleic acid model), to duplex DNA-binding systems in which a sequence-specific recognition site is flanked by regions that can form non-sequence-specific contacts with protein. Incorporation of duplex modelling will also allow estimation of the role of nucleic acid secondary structure for interacting systems such as eIF4E and mRNA. Our modelling suggests that a degree of tuning may exist between charged surface shapes of protein and nucleic acid. Bioinformatics analysis of the database of complexes could go beyond positive/negative charge recognition to study underlying patterning. These further developments generally have the potential to enhance computational analyses of proteins of unknown function, particularly when combined with searches for specific binding/tethering sites.

ACKNOWLEDGEMENTS

We thank Professor John McCarthy and Drs Leo Lue, Tobias von der Haar and Marina Pushkina for advice and discussion. The work was funded by grant number 36/B17005 from the UK BBSCR. Funding to pay the Open Access publication charges for this article was provided by the UK BBSCR. Conflict of interest statement. None declared.

REFERENCES

1. Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science, 253, 1001–1007.
2. Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res., 32, 3040–3052.
3. Horn, P.J. and Peterson, C.L. (2002) Molecular biology. Chromatin higher order folding: wrapping up transcription. Science, 297, 1824–1827.
4. Conway, J.F., Wikoff, W.R., Cheng, N., Duda, R.L., Hendrix, R.W., Johnson, J.E. and Steven, A.C. (2001) Virus maturation involving large subunit rotations and local refolding. Science, 292, 744–748.
5. Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2001) An overview of the structures of protein–DNA complexes. Genome Biol., 1, reviews 1–37.
6. Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. Trends Biochem. Sci., 28, 411–418.
7. Dixit, S.B., Andrews, D.Q. and Beveridge, D.L. (2005) Induced fit and the entropy of structural adaptation in the complexation of CAP and λ-repressor with cognate DNA sequences. Biophys. J., 88, 3147–3157.
8. Shanahan, H.P., Garcia, M.A., Jones, S. and Thornton, J.M. (2004) Identifying DNA-binding proteins using structural motifs and the electrostatic potential. Nucleic Acids Res., 32, 4732–4741.
9. Stawiski, E.W., Gregorek, L.M. and Mandel-Gutfreund, Y. (2003) Annotating nucleic acid-binding function based on protein structure. J. Mol. Biol., 326, 1065–1079.
10. Saecker, R.M. and Record, M.T. Jr (2002) Protein surface salt bridges and paths for DNA wrapping. Curr. Opin. Struct. Biol., 12, 311–319.
11. Beard, D.A. and Schlick, T. (2001) Modeling salt-mediated electrostatics of macromolecules: the discrete surface charge optimization algorithm and its application to the nucleosome. Biopolymers, 58, 106–115.
12. Zhang, D., Konecny, R., Baker, N.A. and McCammon, J.A. (2004) Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus; simulation by a coarse-grained RNA model and a Monte Carlo approach. Biopolymers, 75, 325–337.
13. von der Haar, T., Gross, J.D., Wagner, G. and McCarthy, J.E. (2004) The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. Nature Struct. Biol., 11, 503–511.
14. Sonenberg, N. and Dever, T.E. (2003) Eukaryotic translation initiation factors and regulators. Curr. Opin. Struct. Biol., 13, 56–63.
15. Strudwick, D. and Borden, K.L. (2002) The emerging roles of translation factor eIF4E in the nucleus. Differentiation, 70, 10–22.
16. Richter, J.D. and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature, 433, 477–480.
17. Marcorotijano, J.M., Gingras, A.C., Sonenberg, N. and Burley, S.K. (1999) Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. Mol. Cell, 3, 707–716.
18. Gross, J.D., Moerke, N.J., von der Haar, T., Lagovskoy, A.A., Sachs, A.B., McCarthy, J.E. and Wagner, G. (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell, 115, 739–750.
19. Von der Haar, T., Ball, P.D. and McCarthy, J.E. (2000) Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5′-cap by domains of eIF4G. J. Biol. Chem., 275, 30551–30555.
20. Kentsis, A., Dwyer, E.C., Perez, J.M., Sharma, M., Chen, A., Pan, Z.Q. and Borden, K.L. (2001) The RING domains of the promyelocytic protein PML and the barrier protein Z repress translation by directly inhibiting translation initiation factor eIF4E. J. Mol. Biol., 312, 609–623.
21. Matsuo, H., Moriguchi, T., Takagi, K., Kusakabe, T., Buratowski, S., Sekine, M., Kyogoku, Y. and Wagner, G. (2000) Efficient synthesis of 13C, 15N-labelled RNA containing the cap structure m7GpppA. J. Am. Chem. Soc., 122, 2417–2421.
22. Carberry, S.E., Friedland, D.E., Rhoads, R.E. and Goss, D.J. (1992) Binding of protein synthesis initiation factor 4E to oligoribonucleotides: effects of cap accessibility and secondary structure. Biochemistry, 31, 1427–1432.
23. Schepers, G.C. and Proud, C.G. (2002) Does phosphorylation of the cap-binding protein eIF4E play a role in translation inhibition? Eur. J. Biochem., 269, 5350–5359.
24. Zuberek, J., Jenieliity, J., Jablonsowska, A., Stepinski, J., Dadlez, M., Stolarski, R. and Darzyynkiewicz, E. (2004) Influence of electric charge variation at residues 209 and 159 on the interaction of eIF4E with the mRNA 5′ terminus. Biochemistry, 43, 5370–5379.
25. Blachut-Okrasinska, E., Bojarska, E., Niedzwiecka, A., Chlebicka, L., Darzyynkiewicz, E., Stolarski, R., Stepinski, J. and Antosiewicz, J.M. (2000) Stopped-flow and Brownian dynamics studies of electrostatic effects in the kinetics of binding of 7-methyl-GpppG to the protein eIF4E. Eur. Biophys. J., 29, 487–498.
26. Bruce, A.D., Jackson, A.N., Ackland, G.J. and Wilding, N.B. (2000) Lattice-switch monte carlo method. Phys. Rev. E. Stat. Phys. Plasmas Fluids. Relat. Interdiscip. Topics, 61, 906–919.
27. Pushkina, M., Malys, N. and von der Haar, M.J. (2004) eIF4E isorom 2 in Schizosaccharomyces pombe is a novel stress-response factor. EMBO Rep., 5, 311–316.
28. Warwicker, J. (1986) Continuum dielectric modelling of the protein-solvent system, and calculation of the long-range electrostatic field of the enzyme phosphoglycerate mutase. *J. Theor. Biol.*, 121, 199–210.

29. Warwicker, J. (2004) Improved pK_a calculations through flexibility-based sampling of a water-dominated interaction scheme. *Protein Sci.*, 13, 2793–2805.

30. Grosberg, A.Yu. and Khoklov, A.R. (1993) *Statistical Physics of Macromolecules*. AIP Press, NY.

31. Rivetti, C., Walker, C. and Bustamante, C. (1998) Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.*, 280, 41–59.

32. Mills, J.B., Vacano, E. and Hagerman, P.J. (1999) Flexibility of single-stranded DNA: use of gapped supliex helices to determine the persistence lengths of poly(dT) and poly(dA). *J. Mol. Biol.*, 285, 245–257.

33. Hagerman, P.J. (1997) Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct.*, 26, 139–156.

34. Vanzu, F., Takagi, Y., Shuman, H., Cooperman, B.S. and Goldman, Y.E. (2005) Mechanical studies of single ribosome/mRNA complexes. *Biophys. J.*, 89, 1909–1919.

35. Seol, Y., Skinner, G.M. and Visscher, K. (2004) Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev. Lett.*, 93, 118102.

36. Murphy, M.C., Rasnik, L., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, 86, 2530–2537.

37. Sadus, R.J. (2002) *Molecular Simulation of Fluids*. Elsevier, Amsterdam.

38. Frenkel, D. and Smit, B. *Understanding Molecular Simulation*. 2nd edn. Academic Press, San Diego.

39. Kofoe, D.A. (2002) On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.*, 117, 6911–6914.

40. Ferrenberg, A.M. and Swendsen, R.H. (1989) Optimized monte carlo data analysis. *Phys. Rev. Letters*, 63, 1195–1198.

41. Pushkina, M., Berthelot, K., von der Haar, T., Geffers, L., Warwicker, J. and McCarthy, J.E.G. (2001) A second eIF4E protein in *Schizosaccharomyces pombe* has distinct eIF4G-binding properties. *Nucleic Acids Res.*, 29, 4561–4569.

42. Topisirovic, I., Ruiz-Gutierrez, M. and Borden, K.L.B. (2004) Phosphorylation of the eukaryotic translation initiation factor eIF4E contributes to its transformation and mRNA transport properties. *Cancer Res.*, 64, 8639–8642.

43. Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P. and Sonenberg, N. (2005) A new paradigm for translational control: Inhibition via 5'-3' mRNA tethering by bicoid and the eIF4E cognate 4EHP. *Cell*, 121, 411–423.

44. Culjkovic, B., Topisirovic, I., Skrabak, L., Ruiz-Gutierrez, M. and Borden, K.L.B. (2005) eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'-UTR. *J. Cell. Biol.*, 169, 245–256.

45. Yagi, T. (1992) Regulation of intracellular osmotic pressure during the initial stages of salt stress in a salt-tolerant yeast, *Zygosaccharomyces rouxii*. *Microbiol.*, 70, 932–102.

46. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PDB Viewer: an environment for comparative protein modeling. *Electrophoresis*, 18, 2714–2723.