Function and Structure of HIV-1 Rev

by

Matthew D. Daugherty

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Function and Structure of HIV-1 Rev

by

Matthew D. Daugherty

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

.....

.....

.....

.....

Committee in Charge

Acknowledgements

There are many people who made this thesis possible. First and foremost, I need to start by thanking Alan Frankel, who has been a constant source of enthusiasm and support. He willingly gives people in his lab the freedom to pursue their interests, which has allowed me to meander all over the scientific landscape and follow this project wherever it has taken me. And if he ever doubted that it would go somewhere meaningful, he never showed it. Alan has a razor sharp insight when it comes to identifying relevant issues, which also makes him an exceptional writer. I've learned a great deal from my time in his lab; this work would not have happened without him.

I have also been surrounded by great people in the Frankel lab. Although I've benefited enormously from the knowledge and generosity of everyone in the lab, there are four people I need to thank in particular. First, in order of appearance, is Steve Landt. This thesis work began on the coat-tails of his research, but more importantly, he offered some sage advice on how to navigate the waters of graduate school and the Frankel lab. Next is Rob Nakamura, with whom I've shared countless cups of coffee and great conversations. Iván D'Orso continues to inspire me with the way he thinks and works. And last but not least, David Booth has been a constant source of enthusiasm, a great collaborator, and a meticulous editor of this thesis.

I have been extremely fortunate to have the support and seemingly endless knowledge of John Gross available to me throughout graduate school. He has basically been a second advisor to me, in addition to being on my thesis committee. The other member of my thesis committee, Geeta Narlikar, has likewise been extremely helpful

with matters both technical and theoretical. I consider myself very lucky to have had these two exceptional scientists guide my time here and my thesis work.

I am also very grateful to have done this work within the extraordinary scientific community at UCSF. The people I have met and learned from here have been second to none in their generosity and knowledge. And some of them I hope to count as friends for years to come. In particular, the work in Chapters 3 and 4 could have never been accomplished without the help and expertise of so many people in the structure labs at UCSF. Hopefully I have individually acknowledged everyone in those chapters.

As a final scientific acknowledgement, I need to thank Andrei Osterman for inspiring me in ways both scientific and personal. His passion for problems and unusually broad way of thinking is something I continue to carry with me.

My family and friends have never wavered throughout this process, even at times when I have. In particular I need to thank Jenessa's parents, David and Shoshona, for their support all the way through. My own family, especially both of my brothers, have contributed to this work in ways they could never know. My parents have been part inspiration, part motivation and part support group. Combine my mom's eternal optimism with my dad's unflinching determination and you have an extremely effective duo. They are the people who have made me who I am, so they deserve a great deal of the credit for this thesis.

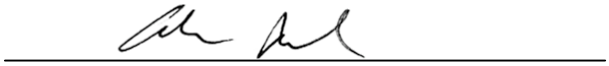
Finally, it is with immense gratitude that I thank my wife Jenessa for everything big and small that she has done in these last number of years. Her unconditional encouragement has provided a stable foundation in my life when everything else seemed up in the air. Without her, this thesis would never have been what it is.

Part of this thesis is a reproduction of material previously published. The text of Chapter 2 is reprinted from *Molecular Cell* (2008) Volume 31(6): pages 824-834 with permission from Elsevier.

Function and Structure of HIV-1 Rev

by

Matthew D. Daugherty



Alan D. Frankel, Ph.D.

Graduate advisor and thesis committee chair

Abstract

Replication of HIV requires the nuclear export of unspliced viral RNAs for the translation of structural proteins and viral packaging. Crucial to this process is oligomeric binding of the regulatory protein Rev to the Rev Response Element (RRE) within the viral RNA, and subsequent export through the nuclear pore complex. Despite two decades of intense scrutiny, the mechanistic importance of Rev oligomerization in RNA export was unknown. More strikingly, the molecular structures of the Rev protein and Rev-RRE complexes were unknown, in large part due to difficulties with aggregation and oligomerization of the Rev protein. Described here are my efforts to use biochemical, biophysical and structural techniques to understand the manner in which Rev oligomerizes on RNA and the role of assembly of the Rev-RRE complex in viral replication. We first established that Rev monomers bind cooperatively to multiple

discrete RNA sites within the RRE using an adaptable protein-RNA interface, forming a homooligomeric complex with 500-fold higher affinity than the tightest single interaction. Importantly, proper multimeric Rev-RRE assembly is strongly correlated with RNA export activity, indicating that oligomerization mediates high affinity complex formation and is thus required for *in vivo* function. With a better understanding of defined Rev-RNA complexes, we next sought to enhance the solubility and homogeneity of these complexes for structural studies. Using these techniques, we obtained low resolution structural information for a Rev dimer in complex with an essential portion of the RRE, as well as with a hexamer of Rev in complex with the RRE. This work thus provides the first structural glimpse of a functional Rev-RRE complex. As a final effort to understand Rev oligomerization, we crystallized the Rev dimer and solved the structure at 2.8 Å resolution. After nearly twenty years, this is the first molecular structure of the Rev protein and reveals the basis for monomeric folding, dimeric stability and cooperative RNA binding.

Table of Contents

Title page ..... *i*
Acknowledgements ..... *iii*
Abstract ..... *vi*
Table of Contents ..... *viii*

CHAPTER 1: INTRODUCTION 1

REFERENCES 11
FIGURE LEGEND 17
FIGURE 18

CHAPTER 2: A SOLUTION TO LIMITED GENOMIC CAPACITY: USING ADAPTABLE BINDING SURFACES TO ASSEMBLE THE FUNCTIONAL HIV REV OLIGOMER ON RNA 19

SUMMARY 20
INTRODUCTION 21
RESULTS 25
DISCUSSION 33
EXPERIMENTAL PROCEDURES 37
REFERENCES 43
ACKNOWLEDGEMENTS 49
FIGURE LEGENDS 50
FIGURES 54
SUPPLEMENTAL DATA 61

CHAPTER 3: ASSEMBLY AND STRUCTURAL INSIGHT INTO DISCRETE HIV REV-RNA COMPLEXES 66

ABSTRACT 67
INTRODUCTION 68
RESULTS 72
DISCUSSION 84
METHODS 87
REFERENCES 93
ACKNOWLEDGEMENTS 100
FIGURE LEGENDS 101
TABLES 105
FIGURES 107
SUPPLEMENTAL DATA 114

CHAPTER 4: STRUCTURE OF THE HIV-1 REV DIMER AT 2.8 Å RESOLUTION 116

SUMMARY 117
RESULTS AND DISCUSSION 118
METHODS 124
REFERENCES 128
ACKNOWLEDGEMENTS 132
FIGURE LEGENDS 133
FIGURES 136
SUPPLEMENTAL DATA 140

CHAPTER 5: IMPLICATIONS AND FUTURE DIRECTIONS 145

REFERENCES 154

APPENDIX 1: CRYSTALLIZATION AND PRELIMINARY X-RAY ANALYSIS OF HIV-1 REV: THE ROLE OF CRYOPROTECTANTS IN PHASE TRANSITIONS OF CRYSTAL SYMMETRY 159

RESULTS AND DISCUSSION 160
METHODS 171
REFERENCES 177
FIGURE LEGENDS 179
TABLES 180
FIGURES 182

List of Figures

CHAPTER 1

FIGURE 1 18

CHAPTER 2

FIGURE 1 54
FIGURE 2 55
FIGURE 3 56
FIGURE 4 57
FIGURE 5 58
FIGURE 6 59
FIGURE 7 60
FIGURE S1 63
FIGURE S2 65

CHAPTER 3

FIGURE 1 107
FIGURE 2 108
FIGURE 3 109
FIGURE 4 110
FIGURE 5 111
FIGURE 6 112
FIGURE 7 113
FIGURE S1 114
FIGURE S2 115

CHAPTER 4

FIGURE 1 136
FIGURE 2 137
FIGURE 3 138
FIGURE 4 139
FIGURE S1 141
FIGURE S2 142

APPENDIX 1

FIGURE 1 182
FIGURE 2 183

List of Tables

CHAPTER 3

TABLE 1 105

CHAPTER 4

TABLE S1 143

APPENDIX 1

TABLE 1 180
TABLE 2 181

Chapter 1

Introduction

Acquired Immune Deficiency Syndrome (AIDS) continues to be a global pandemic. In 2007, over 30 million people were living with AIDS or the virus that causes it, Human Immunodeficiency Virus (HIV)[1]. Despite over two decades of research and increased access to antiretroviral treatment, over 2 million people died from AIDS in 2007, bringing the total number of deaths from the disease to over 25 million people globally[1]. The severity of this disease makes continued research into the mechanisms of the viral replication essential.

One crucial step in the HIV-1 life cycle is the export of intron-containing viral RNAs from the nucleus to the cytoplasm. Initially, fully spliced RNAs are exported using the cellular TAP/Nxt pathway and encode the viral regulatory proteins Tat, Rev and Nef. However, replication requires the nuclear export of partially and fully unspliced viral RNAs for both translation of late viral genes and packaging of genomic RNA[2,3]. In order to overcome the nuclear retention of unspliced RNAs by host cell factors, the Rev protein is imported into the nucleus and binds to the Rev Response Element (RRE) found in the introns of unspliced viral RNAs[3-5]. To direct the transport of these RNAs to the cytoplasm, Rev binds to the human nuclear export receptor, Crm1, and forms a ternary complex stabilized by GTP-bound Ran[3,6]. After the complex docks to and translocates through the nuclear pore, RRE-containing RNAs are released in the cytoplasm and the protein components are shuttled back into the nucleus.

The Rev protein is 116 amino acids long and contains three important functional domains (Figure 1A). The nuclear export sequence (NES) found in the C-terminal region of the protein is a leucine rich motif that is required for direct interaction with Crm1 and nuclear export[7,8]. This sequence alone is sufficient to mediate nuclear export[8], and is

similar to numerous other export domains of Crm1 dependent shuttling proteins. The N-terminal region of the protein contains an arginine-rich motif (ARM) which mediates specific RNA binding as well as serving as a nuclear localization sequence. The ARM is immediately flanked by two oligomerization domains which facilitate protein-protein interactions between Rev monomers in the multimeric Rev-RRE complex.

Rev mediated RNA export requires the oligomerization of multiple Rev monomers on the RRE, a necessity that was initially established by the loss-of-function mutants in the oligomerization domains that flank the ARM[9]. RNA export and oligomerization of Rev on the RRE also requires a large portion of the highly structured 351-nt RRE (Figure 1B), such that progressively shorter RNAs have progressively lower Rev responsiveness[10]. Although essential for in vivo function, the mechanistic role of oligomerization had been unclear due to the lack of a clear biochemical function in RNA binding or RNA export. Early studies identified stem-loop IIB (Figure 1B) as the primary specific binding site, and a minimized IIB RNA hairpin binds a single Rev monomer with high specificity although it is insufficient for export activity[10,11]. Previous biochemical studies showed that 6-10 subunits can bind the RRE, but curiously, Rev bound to the isolated stem IIB hairpin with similar affinity as to the full RRE even though fewer subunits are bound[9,12-17]. Moreover, mutation of residues important for oligomerization had shown no loss in primary RNA-binding affinity [9,18,19] and thus the functional role of oligomerization in the HIV life cycle had remained uncertain.

Uncovering the mechanistic function of oligomerization in RNA export was the motivation behind the work described in Chapter 2. Our results demonstrate that Rev monomers bind cooperatively to multiple discrete RNA sites within the RRE using the

oligomerization domains and the inherent adaptability of the ARM (described below). This allows Rev to form a homooligomeric complex with 500-fold higher affinity than the tightest single interaction. Importantly, proper multimeric Rev-RRE assembly is strongly correlated with RNA export activity, indicating that oligomerization mediates high affinity complex formation and is thus required for in vivo function.

Interestingly, our results contrasted with numerous previous reports that found little to no increase in affinity of Rev binding to the full RRE over Rev binding to stem IIB alone[9,12-17]. The inconsistency appeared to be due to differences in preparation of the protein for biochemical studies; our studies used a fully native preparation of the protein that even retained endogenous *E. coli* RNA throughout purification, while many previous efforts had used denaturing preparations or Rev purified free of RNA. The implications of this will be discussed further below and in Chapter 3.

The arginine-rich motif of Rev: an adaptable RNA-binding surface

It has previously been shown that the isolated Rev ARM peptide, when properly stabilized as an alpha-helix, recapitulates the specificity and affinity of the interaction between Rev and stem IIB[20]. Mutagenesis and structural studies have revealed that the Rev peptide inserts deeply into the widened major groove of the RNA, and that there are numerous contacts between side chains of the peptide and specific bases of the RNA (Figure 1C). Most notably, the carboxamide moiety of Asn 40 forms hydrogen bond contacts to both partners of the G47:A73 base pair, and mutation of either the asparagine or the RNA bases involved in this interaction results in a severe loss of affinity and specificity[20-22]. More recent work in our lab has revealed that constrained

peptidomimetics using macrolactam constraints between amino acid side chains (i, i+4) can be used to recognize stem IIB. Importantly, the primary determinants of binding of these molecules to stem IIB are the placement and identity of the helix-stabilizing constraints and a single carboxamide containing sidechain[23]. Thus, even in the context of a non-natural molecule, interactions between the carboxamide moiety and the stem IIB RNA are the primary source of specificity of binding.

However, other binding modes are possible with the Rev ARM. Additional studies with the Rev ARM, as well as ARMs from other proteins, have established that this category of RNA-binding domain is able to bind to and is highly adaptable to different nucleic acid targets. Multiple selection experiments with the Rev ARM have been performed and revealed a large variety of nucleic acid structures that can be bound with high affinity and specificity, seemingly requiring only a disruption in the nucleic acid duplex[24-26]. Interestingly, not only do these peptides bind different types of structures, the same peptide adapts differently to each binding site, making specific contacts by utilizing different amino acid sidechains, different surfaces of the molecule and sometimes different backbone conformations of the peptide[26,27] (Landt et al., in preparation). This "chameleon" binding allows the adaptive binding of ARM peptides to different RNAs using a diverse set of amino acid contacts for each RNA[28,29].

My involvement in the some of the above mentioned studies[23,26,29] (Landt et al., in preparation) piqued my interest in this adaptability of the ARM peptide to its RNA binding partner. In particular, adaptable binding appeared to me to be an excellent method for viruses to evolve specificity in a protein-RNA interaction, a supposition supported by the use of ARMs in numerous viral proteins including retroviral Rev and

Tat. First, the ARM requires little protein sequence, which is important in the context of the constrained size of viral genomes. Second, adaptable recognition between an ARM and an RNA site might be more tolerant of the highly mutagenetic nature of RNA viruses, since a great deal of the binding affinity comes from electrostatic attraction between the numerous arginine side chains and the RNA backbone. This structural flexibility may allow for more sequence variability than a more rigid protein-RNA interaction. Indeed, it has very recently been suggested that the structural flexibility observed in many viral proteins enables them to function in the context of a dynamic genome and host[30].

Particularly intriguing to me was the observation that the Rev ARM binds numerous selected nucleic acid structures and that inspection of the RRE reveals numerous bulges and disruptions that would be ideal structures for binding by Rev based on these in vitro selections (Figure 1B). Although other regions of the RRE had been suggested to be involved in Rev binding[10,31,32], no other specific binding sites had been uncovered, leading to models in which binding to IIB nucleates oligomerization along the RNA[10,13,18,33], analogous to the coating of viral DNAs or RNAs during genome packaging[34]. However, in light of our observation that Rev binds with much higher affinity to the full RRE than to stem IIB alone, we hypothesized that secondary binding sites would be present in the RRE.

Crucial to the conclusions described in Chapter 2 regarding cooperative, high-affinity binding of the Rev oligomer to the RRE is the identification of an additional specific binding site for Rev in stem IA of the RRE. This represents the first specific binding site characterized for Rev outside of stem IIB. While the description of an

additional binding site was not completely unexpected, the interesting aspect of Rev binding to stem IA is how it contrasts with Rev binding to stem IIB. Not only is the sequence and structure of the stem IA site different than stem IIB, the binding mode utilized by Rev to bind to each site is different. The amino acids used to bind stem IA map to the opposite side of the ARM alpha-helix than is used to bind stem IIB. This strongly suggests that Rev employs the adaptability of the ARM, which had only been observed using in vitro selections, to form the functional Rev-RRE oligomer. This conclusion is supported by the functional consequences of mutations to non-IIB contacting residues in the ARM, as well as the high level of conservation of those same residues in all HIV-1 isolates.

HIV Rev structure

Another important conclusion of the work described in Chapter 2 is that Rev assembles into a biochemically more discrete complex with the RRE than had been previously proposed. That, coupled with the fact that we observed high affinity complex formation by using a more native protein preparation of Rev than had been used previously, hinted to us that we may be in a unique position to attack the problem of Rev structure.

Despite intense study for the last two decades, the molecular structure of Rev was unknown. In large part, this has been due to the low solubility of the protein in vitro at concentrations higher than 1-5 uM[13,35]. Although the structure of the ARM peptide with stem IIB was solved by NMR[22], no structures were available for the rest of the Rev protein or the Rev-RRE complex. Current models of the Rev protein and Rev-RRE

assembly have relied on biochemical and functional data. Circular dichroism studies suggest that the protein adopts a helix-turn-helix motif, with a relatively unstructured C-terminus[36,37]. In an elegant study, detailed biochemical mapping of the oligomerization domains led to a model for Rev structure and oligomerization[18]. However, without access to soluble species of homogenous Rev-RNA complexes, experimental structural determination had been unattainable.

It had also been observed that Rev could be induced to form filaments at approximately 5 uM by slowly warming the solution from 4°C to 25°C over several hours in the presence or absence of RNA[13,35]. Similar filaments form by slowly concentrating a Rev or Rev-RNA solution over several days at 4°C, and these fibril assemblies exhibit the predicted helix-turn-helix topology[38,39]. The relevance of these assemblies, however, has been difficult to reconcile due to the lack of biological significance[10,39].

Based on our observation that the manner in which the Rev protein is expressed and purified has a drastic effect on the biochemical activity of the protein in vitro, we hypothesized that it may also have different biophysical characteristics. Specifically, we wished to ask whether natively prepared Rev-RNA complexes might be more soluble and thus amenable to structural characterization. The work in Chapter 3 describes my efforts toward making homogenous, highly soluble species of discrete Rev-RNA complexes.

First, we were able to show that the Rev protein is highly soluble if the proper electrostatic considerations are provided. The Rev protein is overall highly charged, with 18 positively charged residues, 15 negatively charged residues and a predicted pI of 9.23. As a result, the protein is prone to aggregate without neutralization of these charges by an

'electrostatic sink'. We have found that binding the protein to the RRE or stem IIB derived RNAs is most effective at preventing Rev aggregation at concentrations needed for structural studies. However, non-specific RNA, such as tRNA, or oxyanionic counterions, such as sulfate and phosphate salts, also prevent the protein from aggregating. Both methods allow the protein to be concentrated to more than 1 mM, which represents a 100-1000 fold increase in the solubility limit of the protein.

Interestingly, a commonly used fusion domain, GB1[40], also prevents aggregation of the protein when expressed in cis. Although assumed to be an inert solubility tag, we have shown that GB1 solubilizes Rev by interacting with it, and is completely dispensable once Rev is bound to RNA. With a predicted pI of 4.55 for GB1, these data suggest that GB1 is serving as an RNA surrogate for Rev. More broadly, GB1 may be a generalizable fusion domain for numerous nucleic acid binding proteins which are prone to aggregation. Indeed, at least two recent structural studies of RNA binding proteins have required GB1 for solubility, although both present evidence that GB1 is not interacting with the RNA-binding protein[41,42].

These methods have allowed us to obtain structural restraints on Rev-RNA complexes using NMR and SAXS, which are consistent with previous biochemical work. Notably, we have also viewed for the first time a functional Rev-RRE complex by EM. These results are described in Chapter 3, and represent the first of many studies that are sure to follow detailing the molecular structures of Rev with RNA.

These improvements in Rev solubility and homogeneity also enabled the use of X-ray crystallography to study Rev structure. In Chapter 4 and Appendix 1 of my thesis, I describe the determination of the crystal structure of the Rev dimer at 2.8 Å resolution.

Lacking only the disordered C-terminus of the protein, we present the atomic model of the Rev monomer, which has eluded structure determination for two decades. Crystallization of the protein as a dimer also allows us to provide a structural explanation for cooperative RNA binding and reveals the molecular details of the dimerization interface, which is a potential therapeutic target. Interestingly, aligning the Rev dimer structure with the structure of the stem IIB RNA further confirms that Rev uses of multiple binding modes to contact RNA.

Chapters 3 and 4 provide us with a much greater understanding of the structure and assembly of the Rev oligomer and the Rev-RRE complex. However, in many ways, these studies are just the beginning of what is sure to become a rapid expansion of our understanding of this essential viral RNP. A few possibilities for future developments in this area are described in the concluding chapter.

References

1. UNAIDS/WHO. AIDS epidemic update: December 2007. (UNAIDS, Geneva, 2007).
2. Cullen, B.R. Nuclear RNA export. *J Cell Sci* 116, 587-97 (2003).
3. Cullen, B.R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* 28, 419-24 (2003).
4. Cullen, B.R. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* 249, 203-210 (1998).
5. Pollard, V.W. & Malim, M.H. The HIV-1 Rev Protein. *Annu. Rev. Microbiol.* 52, 491-532 (1998).
6. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I.W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90, 1051-1060 (1997).
7. Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. & L√ºhrmann, R. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* 82, 475-83 (1995).
8. Wen, W., Meinkoth, J.L., Tsien, R.Y. & Taylor, S.S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* 82, 463-73 (1995).
9. Malim, M.H. & Cullen, B.R. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* 65, 241-48 (1991).

10. Mann, D.A. et al. A molecular rheostat: Co-operative rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J. Mol. Biol.* 241, 193-207 (1994).

11. Kjems, J. & Sharp, P.A. The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. *J Virol* 67, 4769-76 (1993).

12. Cook, K.S. et al. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* 19, 1577-83 (1991).

13. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. & Singh, M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. USA* 88, 7366-70 (1991).

14. Huang, X. et al. Minimal Rev-response element for Type 1 human immunodeficiency virus. *J. Virol.* 65, 2131-34 (1991).

15. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res.* 20, 6465-72 (1992).

16. Kjems, J., Calnan, B.J., Frankel, A.D. & Sharp, P.A. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J* 11, 1119-29 (1992).

17. Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. & Cullen, B.R. Identification of a high-affinity RNA-binding site for the human

immunodeficiency virus type 1 Rev protein. *Proc. Natl. Acad. Sci. USA* 89, 758-62 (1992).

18. Jain, C. & Belasco, J.G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* 7, 603-614 (2001).

19. Edgcomb, S.P. et al. Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. *Protein Sci* (2008).

20. Tan, R., Chen, L., Buettner, J.A., Hudson, D. & Frankel, A.D. RNA recognition by an isolated alpha helix. *Cell* 73, 1031-40 (1993).

21. Tan, R. & Frankel, A.D. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry* 33, 14579-14585 (1994).

22. Battiste, J.L. et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* 273, 1547-1551 (1996).

23. Mills, N.L., Daugherty, M.D., Frankel, A.D. & Guy, R.K. An alpha-helical peptidomimetic inhibitor of the HIV-1 Rev-RRE interaction. *J Am Chem Soc* 128, 3496-7 (2006).

24. Xu, W. & Ellington, A.D. Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc. Natl. Acad. Sci. USA* 93, 7475-80 (1996).

25. Bayer, T.S., Booth, L.N., Knudsen, S.M. & Ellington, A.D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* 11, 1848-57 (2005).

26. Landt, S.G., Ramirez, A., Daugherty, M.D. & Frankel, A.D. A simple motif for protein recognition in DNA secondary structures. *J. Mol. Biol.* 351, 982-94 (2005).

27. Ye, X. et al. RNA architecture dictates the conformations of a bound peptide. *Chem. Biol.* 6, 657-69 (1999).

28. Smith, C.A., Calabro, V. & Frankel, A.D. An RNA-binding chameleon. *Mol. Cell* 6, 1067-76 (2000).

29. Calabro, V., Daugherty, M.D. & Frankel, A.D. A single intermolecular contact mediates intramolecular stabilization of both RNA and protein. *Proc Natl Acad Sci U S A* 102, 6849-54 (2005).

30. Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I.N. & Tawfik, D.S. Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 34, 53-9 (2009).

31. Kjems, J., Brown, M., Chang, D.D. & Sharp, P.A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* 88, 683-7 (1991).

32. Zemmel, R.W., Kelley, A.C., Karn, J. & Butler, P.J. Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. *J. Mol. Biol.* 258, 763-77 (1996).

33. Malim, M.H., McCarn, D.F., Tiley, L.S. & Cullen, B.R. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J Virol* 65, 4248-54 (1991).

34. Harrison, S.C., Skehel, J.J. & Wiley, D.C. Virus structure. in *Virology*, Vol. 1 (ed. Fields, B.N.) 59-99 (Lippincott-Raven, Philadelphia, 1996).

35. Wingfield, P.T. et al. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. *Biochemistry* 30, 7527-34 (1991).

36. Daly, T.J., Rusche, J.R., Maione, T.E. & Frankel, A.D. Circular dichroism studies of the HIV-1 Rev protein and its specific RNA binding site. *Biochemistry* 29, 9791-5 (1990).

37. Auer, M. et al. Helix-loop-helix motif in HIV-1 Rev. *Biochemistry* 33, 2988-96 (1994).

38. Blanco, F.J., Hess, S., Pannell, L.K., Rizzo, N.W. & Tycko, R. Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J Mol Biol* 313, 845-59 (2001).

39. Havlin, R.H., Blanco, F.J. & Tycko, R. Constraints on protein structure in HIV-1 Rev and Rev-RNA supramolecular assemblies from two-dimensional solid state nuclear magnetic resonance. *Biochemistry* 46, 3586-93 (2007).

40. Huth, J.R. et al. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci* 6, 2359-64 (1997).

41. Hargous, Y. et al. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *Embo J* 25, 5126-37 (2006).

42. Deshmukh, M.V. et al. mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell* 29, 324-36 (2008).

43. Charpentier, B., Stutz, F. & Rosbash, M. A dynamic in vivo view of the HIV-I Rev-RRE interaction. *J. Mol. Biol.* 266, 950-62 (1997).

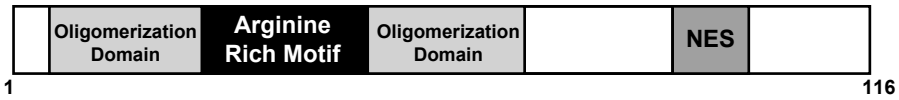
44. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V. & Cullen, B.R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 338, 254-7 (1989).

Figure Legend:

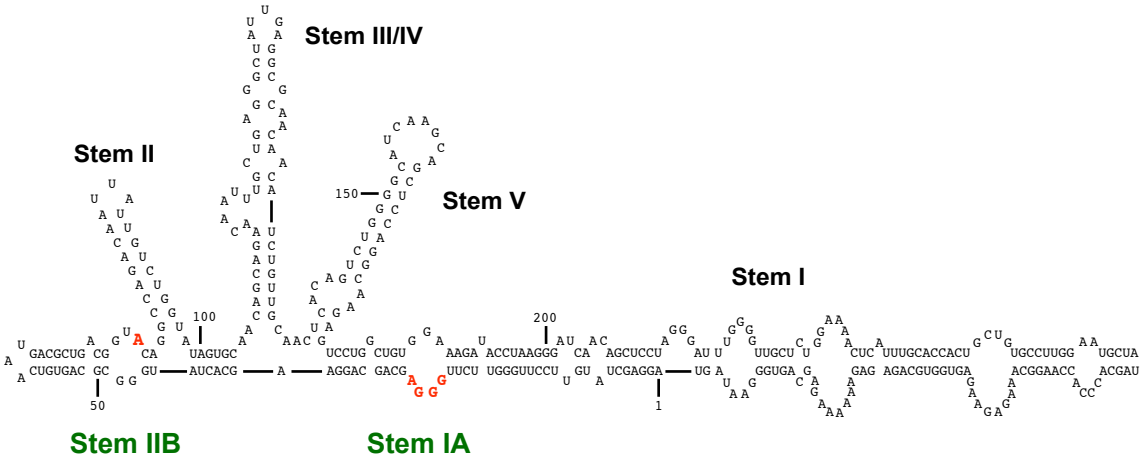
Figure 1: Rev, the RRE and the Rev ARM with stem IIB

A) Domain structure of Rev. B) Predicted secondary structure of the RRE with stem IIB and stem IA indicated. Stem nomenclature is from Charpentier et al.[43], and numbering corresponds to the originally described RRE structure[44]. C) Structural model of the helical ARM region of the Rev protein bound to stem IIB as determined by NMR (PDB code 1ETF)[22]. Shown is the side chain of Asn 40 (in broad lines), which makes essential contacts to the G47:A73 basepair (in thinner lines).

A


Oligomerization Domain
Arginine Rich Motif
Oligomerization Domain
NES
1
116

B


Stem III/IV
Stem II
Stem V
Stem I
Stem IIB
Stem IA

C

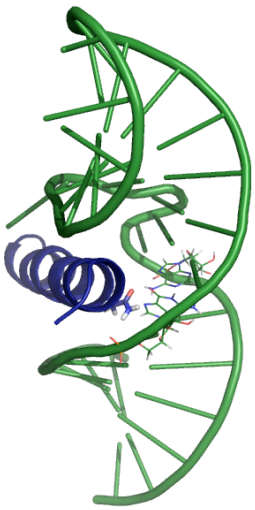


Figure 1

Chapter 2

A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA

Matthew D. Daugherty[1,2], Iván D'Orso[2], and Alan D. Frankel[2,*]

[1]Chemistry and Chemical Biology Graduate Program
[2]Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, CA 94158

[*]Contact: Alan Frankel
Department of Biochemistry and Biophysics
UCSF
600 16th St.
San Francisco, CA 94158
415-476-9994
frankel@cgl.ucsf.edu

Running title: Assembly of a high-affinity Rev-RNA complex

*This work has been previously published in *Molecular Cell*. 2008; 31(6): 824-834.

Summary

Many ribonucleoprotein (RNP) complexes assemble into large, organized structures in which protein subunits are positioned by interactions with RNA and other proteins. Here we demonstrate that HIV Rev, constrained in size by a limited viral genome, also forms an organized RNP by assembling a homo-oligomer on the Rev response element (RRE) RNA. Rev subunits bind cooperatively to discrete RNA sites using an oligomerization domain and an adaptable protein-RNA interface, forming a complex with 500-fold higher affinity than the tightest single interaction. High affinity binding correlates strongly with RNA export activity. Rev utilizes different surfaces of its α-helical RNA-binding domain to recognize several low-affinity binding sites, including the well-characterized stem IIB site and a newly discovered site in stem IA. We propose that adaptable RNA-binding surfaces allow the Rev oligomer to assemble economically into a discrete, stable RNP and provide a mechanistic role for Rev oligomerization during the HIV life cycle.

Introduction

Several key macromolecular machines are composed of ribonucleoprotein (RNP) assemblies, including the ribosome, spliceosome, signal recognition particle, and telomerase (Ban et al., 2000; Hainzl et al., 2002; Shen and Green, 2004; Zappulla and Cech, 2004). To organize such assemblies, proteins use several strategies to bind RNA with high affinity and specificity. In some cases, such as the ribosome, many hetero-oligomeric subunits recognize discrete portions of an RNA and become organized further by protein-protein interactions (Ban et al., 2000). In other cases, such as the Sex-lethal splicing protein, multiple RNA-binding domains are tethered on a single polypeptide to enable the protein to bind to an extended RNA site (Lunde et al., 2007). Both of these strategies utilize substantial protein coding capacity, either to encode multiple subunits or multi-domain proteins, typically using gene or domain duplication and subsequent evolution to generate new specificities. A more frugal coding strategy utilizes homo-oligomeric proteins to bind nucleic acids, most notably with DNA-binding proteins where each subunit typically recognizes a repeated or closely related binding site arrayed along the DNA (Marmorstein and Fitzgerald, 2003). Limiting the size of proteins is especially important to viruses, which often evolve mechanisms to cope with their limited coding capacity, such as organizing genes in overlapping reading frames. Here we describe a strategy used by HIV-1 to assemble a large RNP using a small, homo-oligomeric protein.

The 116-amino acid HIV-1 Rev protein binds to the ~350-nt Rev response element (RRE) RNA found in the introns of partially spliced and unspliced viral mRNAs, forming a large RNP that directs their transport to the cytoplasm before splicing is completed. These exported mRNAs either are translated into the viral structural proteins

or packaged as genomic RNA into virions (Cullen, 1998, 2003; Pollard and Malim, 1998). To export its bound RNA cargo, Rev binds to the human nuclear export receptor, Crm1, through its nuclear export sequence (NES; Figure 1A), forming a ternary complex stabilized by GTP-bound Ran (Cullen, 2003; Fornerod et al., 1997). After the complex docks to and translocates through the nuclear pore, RRE-containing RNAs are released in the cytoplasm and the protein components are shuttled back into the nucleus.

The function of the RNP relies on the binding of several Rev monomers to the RRE. This oligomeric assembly requires the RNA-binding arginine-rich motif (ARM) and flanking oligomerization domains of Rev (Figure 1A) in addition to most of the highly structured RRE (Figure 1B) (Hope et al., 1990; Huang et al., 1991; Malim and Cullen, 1991; Mann et al., 1994; Pollard and Malim, 1998). Although essential for in vivo function, the mechanistic role of oligomerization has been unclear due to the lack of a clear biochemical function in RNA binding or RNA export. Early studies identified stem-loop IIB (Figure 1B) as the primary specific binding site (Cook et al., 1991; Heaphy et al., 1991; Huang et al., 1991; Iwai et al., 1992; Kjems et al., 1991; Malim and Cullen, 1991; Tiley et al., 1992), and a minimized IIB RNA hairpin binds a single Rev monomer with high specificity although it is insufficient for export activity (Kjems and Sharp, 1993; Mann et al., 1994). Previous biochemical studies showed that 6-10 subunits can bind the RRE, but curiously, Rev binds to the isolated IIB hairpin with similar affinity as to the full RRE even though fewer subunits are bound (Cook et al., 1991; Heaphy et al., 1991; Huang et al., 1991; Iwai et al., 1992; Kjems et al., 1991; Malim and Cullen, 1991; Tiley et al., 1992). Furthermore, an isolated ARM peptide, properly stabilized as an α-helix, has been shown to recapitulate the affinity and specificity of the full complex but

does not oligomerize (Tan et al., 1993). The Rev peptide-IIB interaction has been well characterized: the peptide inserts deeply into a wide RNA major groove and makes numerous specific contacts, most notably forming hydrogen bonds between Asn40 and both partners of the G47:A73 base pair (Battiste et al., 1996; Tan et al., 1993; Tan and Frankel, 1994). Although other regions of the RRE have been suggested to be involved in Rev binding (Kjems et al., 1991; Mann et al., 1994; Zemmel et al., 1996), no other specific binding sites have yet been uncovered, leading to models in which binding to IIB nucleates oligomerization along the RNA (Heaphy et al., 1991; Jain and Belasco, 2001; Malim and Cullen, 1991; Mann et al., 1994), analogous to the coating of viral DNAs or RNAs during genome packaging (Harrison et al., 1996). Mutation of residues important for oligomerization have shown no loss in primary RNA-binding affinity (Edgcomb et al., 2008; Jain and Belasco, 2001; Malim and Cullen, 1991) and thus the functional role of oligomerization in the HIV life cycle has remained uncertain. It has been posited that the high concentrations of Rev needed for oligomerization might set an expression threshold such that Rev function would occur only during the later stages of viral replication (Malim and Cullen, 1991; Mann et al., 1994).

The Rev ARM, like ARMs of other RNA-binding proteins, is highly adaptable to different nucleic acid targets. Selection experiments have identified both RNAs and DNAs with imperfect double helices that bind the Rev ARM with high affinity and specificity (Bayer et al., 2005; Landt et al., 2005; Xu and Ellington, 1996), utilizing different amino acid side chains to contact the RNA and, in some cases, different surfaces of the helical ARM or different peptide backbone conformations (Landt et al., 2005; Ye et al., 1999) (S.G. Landt, unpublished data). This “chameleon” binding behavior of

ARMs in principle can allow a protein to readily adapt to a new RNA site, particularly in a rapidly evolving virus (Bayer et al., 2005; Smith et al., 2000), but no natural example has been described. Inspection of the RRE reveals numerous intriguing bulge structures and other helical disruptions suggestive of the types of sites found in the selection experiments (Figure 1B). Indeed, these disruptions appear to be important for higher order Rev assembly on the RRE (Kjems et al., 1991; Zemmel et al., 1996), while amino acids in the ARM other than those used for IIB recognition are highly conserved among viral isolates (Figure 1A), suggesting that other binding sites and perhaps alternative binding modes may be used in assembling the functional Rev RNP.

Here we present evidence that Rev forms a cooperative, high-affinity (~100 pM) oligomeric complex with the RRE, nearly three orders of magnitude tighter than the single binding site interaction previously reported. Assembly requires the Rev oligomerization domains and multiple specific RNA sites that bind Rev with moderate affinities, including IIB and a new discovered site in stem IA. Interestingly, Rev shows chameleon-like binding behavior in which different Rev monomers of the oligomer bind in different modes to at least one other site. We suggest that to overcome a limited genome size, the virus has evolved a mechanism to form a large, defined RNP using only a single protein with an ARM domain flexible enough to adapt to each site. Further, we have shown for the first time a strong correlation between the affinity of the oligomer and RNA export, suggesting that Rev oligomerization is needed to create a stable export-competent RNP that helps regulate the timing of the HIV life cycle.

Results

A high-affinity, cooperative oligomeric Rev-RRE complex

Previous in vitro binding experiments have shown that the number of Rev monomers bound to the RRE increases as the RNA is lengthened from the IIB hairpin alone to a complete 250-350-nt element but that its affinity remains approximately constant at 20-100 nM (Cook et al., 1991; Heaphy et al., 1991; Huang et al., 1991; Iwai et al., 1992; Malim and Cullen, 1991; Mann et al., 1994; Tiley et al., 1992). The affinity of a minimal Rev ARM peptide-IIB complex also is similar (Tan et al., 1993), leading to models in which one Rev subunit binds to the IIB site and then nucleates binding of additional subunits through protein-protein and nonspecific RNA interactions (Heaphy et al., 1991; Jain and Belasco, 2001; Malim and Cullen, 1991; Mann et al., 1994). In contrast to those previous studies, we have found that Rev purified under native conditions, in which Rev is maintained bound to RNA throughout the purification, has much higher affinity for the full RRE. As the RRE is lengthened from the IIB hairpin to a IIABC three-helix junction to a 242-nt RRE, we observe a large, progressive increase in Rev binding affinity, with 500-fold higher affinity for the full RRE element versus IIB (Figures 2A, 2B, and 2E). Moreover, distinctly cooperative behavior is seen with the larger fragments, with a >3-fold increase in the Hill coefficient and little monomer or dimer observed with the full RRE.

Mutation of the Rev oligomerization domains disrupts function but not its ability to recognize IIB (Edgcomb et al., 2008; Jain and Belasco, 2001; Malim and Cullen, 1991), and we wished to determine whether the loss of function might reflect a loss of

high affinity, cooperative RRE binding. We first compared full-length Rev binding to the 17-amino acid ARM alone, expected to bind non-cooperatively, by fusing the Rev ARM peptide to the B1 domain of streptococcal protein G (GB1) to increase its molecular mass to ~11 kDa and thereby allow us to observe a mobility shift with the full RRE. The GB1-fused Rev ARM binds to the isolated IIB hairpin with similar affinity (30-50 nM) as Rev (data not shown), consistent with previous observations with the ARM peptide. However, in marked contrast to Rev, the ARM binds non-cooperatively to the full RRE, and with little increase in binding affinity (Figures 2C and 2D). The small increase in binding affinity likely reflects the entropic benefit of an increased number of binding sites for individual subunits or small differences in the folding of the IIB hairpin in the larger RRE context. Interestingly, previously described single point mutations in the oligomerization domains, L18Q and L60R (Jain and Belasco, 2001), show similarly deficient binding characteristics (Figures 2C and 2D, and data not shown), underscoring the importance of proper protein-protein interactions for cooperative, high affinity binding to the RRE.

Even a minimized binding site, such as an extended IIB hairpin, can cooperatively bind two Rev subunits, provided that the RNA sites are correctly juxtaposed and the oligomerization domains are intact. Placing a second helical disruption in an extended IIB hairpin permits binding of a second Rev subunit (IIB 40mer, Figure 2A) as has been observed previously (Zemmel et al., 1996), but binding is non-cooperative unless the disruption is configured to better mimic the junction between stems IIA, IIB and IIC, for example by adding two nucleotides to reorient a stem. The experiments reported in the supplemental data (see Figure S1) demonstrate that this 42mer RNA has higher binding

affinity, accumulates less monomer, and has a higher Hill coefficient than the 40mer. Furthermore, cooperative binding is abrogated with an oligomerization domain mutant (Figure S1C). Thus, the greatly enhanced affinity and cooperativity of the Rev-RRE complex requires protein-protein interactions between Rev subunits as well as properly oriented binding sites that provide an RNA scaffold compatible with these interfaces.

RNA-binding specificity of the Rev oligomer

Current models suggest that Rev binds with high specificity to the IIB site and then oligomerizes along the RRE (Jain and Belasco, 2001; Mann et al., 1994; Zemmel et al., 1996), however the high affinity for the full RRE compared to IIB alone suggested that regions outside of IIB also may contribute to Rev recognition. Indeed, we find that introducing an A73G mutation into the isolated IIB hairpin reduces Rev affinity 100-fold and ablates specificity relative to an antisense hairpin (Figures 3A and 3C), as previously observed (Jain and Belasco, 1996), whereas in the context of the full RRE, affinity is reduced only 40-fold and still is 100-fold tighter than an antisense control (Figures 3B and 3C). Moreover, the affinity for the mutated RRE is still 10-fold higher than for the wild type IIB hairpin alone (Figure 3C) and shows reduced levels of monomer and dimer accumulation (see below and Figure S2). Importantly, the extended RRE not only permits high affinity binding but also allows the Rev protein to distinguish the viral RNA from other RNAs, with a 40-fold increase in specificity compared to IIB alone. Thus, binding to IIB accounts for only a fraction of the RRE affinity and specificity, suggesting that Rev also may recognize additional sites.

Identification of a second specific Rev binding site in the RRE

In vitro selection experiments have identified nucleic acid motifs other than IIB that bind Rev ARM peptides with high affinity and specificity (Landt et al., 2005; Xu and Ellington, 1996). Bulge, internal loop, and stem junctions generally appear to be favorable for binding, and inspection of the RRE reveals many potential binding sites (Figure 1B). Some of these regions are protected from nucleases upon Rev binding (Kjems et al., 1991), and the junctions formed along stem I are important for Rev oligomerization (Mann et al., 1994; Zemmel et al., 1996), but no specific sites have been characterized. To identify such additional specific binding sites, we generated small RNA models attempting to mimic the junction regions between stems I, II, III/IV, and V (Figure 1B), and found that an internal loop region in stem IA bound the Rev ARM with only 5-fold lower affinity than IIB (Figure 4A). We observed conformational rearrangements in a three-helix model but no tight binding to the stem IIC, III/IV, or V models (data not shown). It is likely that some of these molecules fold incorrectly in the small RNA contexts, and do not accurately mimic the binding conformations present in the full RRE.

Like IIB, the stem IA binding site contains an asymmetric purine-rich internal loop that we surmised would be important for binding, and indeed, deleting the loop reduced binding of the Rev ARM 5-fold to nonspecific levels (Figure 4A). This reduction is not due to a gross perturbation of IA structure, as conservative mutations that preserve the purine content but change their sequence either on the 5' side, 3' side, or both sides reduces binding to the same extent as the deletion (Figure 4A, and data not shown). The same 5' side substitution engineered into the full RRE significantly reduces

affinity of the Rev oligomer even though the IIB site is completely intact (Figures 4B and 4C). Similar to the IIB binding site, the 5-fold decrease in affinity of Rev for the IA-mutated RRE mirrors the 5-fold decrease observed with minimal ARM peptide binding to the mutated IA hairpin. Thus, the stem IA site represents the first characterized specific Rev binding site outside of stem IIB and is needed in conjunction with IIB, and probably other sites, to generate the high-affinity Rev oligomeric RNP.

The Rev peptide binds to stem IA differently than to stem IIB

ARM peptides generally can adapt to their nucleic acid binding sites, and the Rev peptide in particular has been shown to use different amino acids and, in some cases, different surfaces of the α-helix to make essential contacts to binding sites selected in vitro (Landt et al., 2005; Xu and Ellington, 1996; Ye et al., 1999) (S.G. Landt, unpublished data). This binding mode adaptability has not been functionally characterized in a biological setting, but the presence of at least one other specific binding site in the RRE raised the possibility that Rev may bind to IA in a manner different than IIB. When binding IIB, Rev uses the carboxamide moiety of Asn40 to hydrogen bond to the two purines of the A73:G47 base pair, and mutation at that position causes a 40-fold reduction in binding affinity (Tan et al., 1993). Given, the purine-rich internal loop of IA, we asked if this amino acid also is essential for IA binding. Interestingly, mutation of Asn40 had no effect on IA binding. Conversely, mutation of Arg41, which does not affect IIB binding (Tan et al., 1993), showed a reproducible 2-fold decrease in IA affinity (Figure 5A). R38A and R46A mutants showed similar 2-fold reductions in IA affinity but had no effect on IIB, while R43A, R44A, and W45A

mutants bound IA like the wild type peptide. Helical wheel projections clearly show that different surfaces of the helix are used to recognize IIB and IA RNAs (Figure 5B).

We further characterized the differences in IA and IIB binding modes using amide proton chemical shifts from ^{15}N HSQC NMR spectra to monitor changes in the peptide-RNA interfaces. There is little peak dispersion of peptide resonances in the absence of RNA (Figure 5C), indicative of a largely unstructured molecule. However, we observe substantial amide peak dispersion in the presence of either IIB (red) or IA (blue) RNAs (Figure 5C), including the Hε protons of arginine side chains (Figure 5C inset). The overall upfield shifts observed in both ^{1}H and ^{15}N dimensions is consistent with stabilization of peptide α-helical structure upon binding (Wang and Jardetzky, 2002). Strikingly, the chemical shift patterns are quite different when bound to stem IA or IIB, including peaks that are perturbed in the IA but not in the IIB complex. Most notable is Hε of Trp45, which is positioned opposite the binding interface in the IIB complex (Battiste et al., 1996) (Figure 5B) and thus is not shifted, while it probably is at or near the interface in the IA complex. This face of the α-helix also includes Arg38, Arg41, and Arg46, identified by mutagenesis as important for the IA interaction (Figure 5B), further highlighting the chameleon-like binding behavior of the Rev peptide when bound to the two sites.

Binding mode adaptability is important for Rev RNA export activity

Based on the in vitro peptide binding behavior, we wished to test whether Rev acts as a chameleon in which a combination of binding modes is used to assemble its oligomeric RNP. If so, we would expect that amino acids apparently located at the IA

interface (Figure 5B) might be important for oligomeric binding and Rev function, while other residues important for IIB binding would not be as crucial as previously thought. Indeed, R41A, R42A, and W45A mutations in the full-length Rev context, which do not affect IIB binding, reduce oligomeric Rev affinity 5-100-fold (Figure 6A), while N40A still binds the RRE with 6 nM affinity even though it is essential for IIB recognition (Jain and Belasco, 1996; Tan et al., 1993). Interestingly, the N40A mutant, like the IIB A73G mutant, shows reduced levels of monomer and dimer species (Figure S2), suggesting that these are intermediate Rev-IIB complexes that are not absolutely required to form the complete Rev-RRE RNP.

Given earlier studies showing that oligomeric binding is essential for RNA export (Malim and Cullen, 1991; Pollard and Malim, 1998), we asked whether formation of the high affinity complex in vitro correlates with formation of an export competent complex in vivo. Using a Rev reporter assay in which the RRE has been engineered in the intron of a chloramphenicol acetyltransferase (CAT) gene such that Rev-mediated export of the unspliced mRNA is required for expression (Huang et al., 1991), we observed a striking correlation between oligomeric RNA-binding affinity in vitro and Rev function in vivo (Figures 6B and 6F). Mutation at the crucial N40 position, which binds IIB with 80-350-fold lower affinity (Jain and Belasco, 1996), in fact retains ~40% activity, a value similar to the R41A mutant that binds IA weakly. Moreover, the W45A mutant that forms the lowest affinity oligomeric complex is functionally the most deficient. We next tested the activities of these mutants in a more native RRE context, using a pCMV-GagPol-RRE reporter in which HIV-1 Gag expression requires RRE-mediated mRNA export (Srinivasakumar et al., 1997). Western blot analyses of p55 and p24 Gag (Figure 6D)

show similar effects with the Rev mutants as the CAT assays and correlate well with in vitro binding affinities (Figure 6F). These data are consistent with the hypothesis that Rev binds the RRE using multiple binding modes.

As further confirmation that high affinity, oligomeric binding is relevant to Rev function, we found that oligomerization mutations (L18Q and L60R) that ablate high affinity complex formation also have lower (3-fold) export activities using CAT and Gag reporters (Figure 6B and 6D). Moreover, a mutation in IIB still retained ~30% activity (Figure 6C and 6E), as previously observed (Jain and Belasco, 1996), suggesting that other binding sites still support partial function. A mutant with purine substitutions in the 5' side of the stem IA internal loop showed a slight decrease (~10%) in activity (Figure 6C), consistent with its slight decrease in affinity (see Figure 3), and a double mutant in both IIB and IA showed a slight additive effect, consistent with a functional role for both sites. Thus, mutations in RNA-binding residues of the ARM, the oligomerization domains, and the two identified RNA-binding sites all show an excellent correlation between high affinity, oligomeric RNA binding in vitro and activity in vivo (Figure 6F), supporting the hypothesis that Rev oligomerization is needed to create a stable, export-competent RNP.

Discussion

The Rev protein has been intensively studied for many years, and while much has been learned about the interaction of the helical ARM peptide with the RRE IIB site, the mechanism of assembly and role of the larger oligomeric complex have been less well defined. Previous results suggested that the ARM-IIB interaction contributed most of the affinity and specificity to complex formation, leading to models in which homo-oligomerization is required for steps in RNA export other than assembly of the Rev-RRE RNP, such as enhancing the Rev-Crm1 interaction (Askjaer et al., 1999). However, we now provide evidence that oligomerization on the RRE is required to assemble a complex nearly three orders of magnitude tighter than the IIB interaction, cooperatively assembling a few weakly binding subunits (K_D >20 nM) at specific sites on the RNA scaffold, including stem IIB and a previously uncharacterized site in stem IA. It also was shown previously that a large portion of the RRE is required for RNA export, as well as Rev subunit interactions (Hope et al., 1990; Huang et al., 1991; Jain and Belasco, 2001; Malim and Cullen, 1991; Mann et al., 1994; Pollard and Malim, 1998); we have found that these same requirements are essential for high affinity RNP formation. The tight correlation between in vitro binding and RNA export data (Figure 6F) suggests that the primary role of Rev oligomerization may simply be to enhance RNA-binding affinity. Consistent with this view are the findings that a single Rev NES is sufficient to mediate nuclear export (Wen et al., 1995) and that unspliced mRNAs can be exported even without Rev oligomerization if a second RRE-binding domain is covalently appended to Rev (Furnes et al., 2005) or if protein-RNA affinity is enhanced using the MS2 coat protein-RNA interaction (Nam et al., 2001). This manner of forming an RNP may also

be utilized by Rev for functions other than RNA export, such as translation and genome packaging.

Although high affinity complexes may be achieved through a single protein-RNA interaction, assembling a multi-subunit RNP may help establish the timing of Rev function in the viral life cycle. It has been proposed that oligomeric binding sets a threshold protein concentration for Rev-mediated RNA export and progression to late stages of the viral life cycle (Malim and Cullen, 1991; Mann et al., 1994). Our results suggest that this control is mediated at the level of RNA binding affinity, and that by requiring a high stoichiometry of Rev monomers bound to achieve high affinity export-competent complexes rather than a single high affinity protein-RNA interaction, the proper viral timing may be achieved.

One especially interesting facet of the Rev-RRE assembly is the recognition of different RNA-binding sites using different binding modes of the ARM. In particular, the stem IIB and IA sites appear to be recognized by distinct faces of the α-helix (Figure 5B), explaining why residues involved in binding to both sites are highly conserved between all HIV-1 isolates (Figure 1A). Moreover, the pronounced effects of mutations within the Rev ARM on full complex assembly, which exceed the effects on individual RNA sites (for example, R41A), suggest that there are additional sites in the RRE, beyond IIB and IA that may be recognized using these additional binding modes. This chameleon-like behavior is reminiscent of HIV-1 nucleocapsid and the Jembrana Disease Virus Tat ARM domain, where the RNA-binding domains are able to adopt different conformations in the context of different RNA sites, arguing that adaptable binding may be well suited to rapidly evolving new RNA specificity in viruses (Amarasinghe et al., 2000; Smith et al.,

2000). In the Rev case, this property has been biologically exploited such that different portions of the RRE structure may have been "selected" during viral evolution to bind different surfaces of the ARM, which overlaps with the coding frame of the *env* gene, much like in vitro selection can identify a variety of tight-binding aptamers. The large RRE scaffold, which also is constrained by the *env* gene, presumably then presents these multiple binding sites in an arrangement compatible with the subunit orientations defined by the oligomeric protein-protein interfaces (Jain and Belasco, 2001). The importance of a proper binding site arrangement is underscored by the observations that concatenated arrays of a IIB hairpin or aptamer do not function as well as the wild type RRE despite the high affinity of the individual sites (Kjems and Sharp, 1993; Symensma et al., 1999). Thus, the proper configuration of binding sites within a larger RNA context allows the Rev oligomer to recognize the RRE and form an organized RNP that exports RRE containing viral RNAs over non-RRE containing RNAs with a very high degree of specificity.

The homo-oligomeric Rev complex demonstrates that even a small, 14kDa protein, constrained by the limited genome size of a virus, can generate a rather substantial RNP (200-300 kDa) with the diversity of protein-RNA contacts more characteristic of multi-domain or hetero-oligomeric complexes (Figure 7). It long has been known that covalently tethering RNA binding domains, such as RRMs, within a single protein can increase RNA-binding affinity and specificity (Lunde et al., 2007). Individual contributions can be dissected for each RRM (Perez et al., 1997; Shamoo et al., 1995), and the entropic benefits and costs of tethering may be determined by domain orientations that are fixed in some cases by protein-protein contacts and in others only

when bound to an RNA scaffold. Similar effects are seen with DNA-binding proteins, especially for dimeric arrangements in which subunits are spatially oriented to match the array of binding sites on the DNA, such as with nuclear receptors or GAL4-related proteins (Marmorstein and Fitzgerald, 2003). In contrast, large RNP assemblies, such as the ribosome, typically rely on more complex sets of protein-protein interactions between hetero-oligomers, as well as a more complex RNA scaffold (Ban et al., 2000; Hainzl et al., 2002; Shen and Green, 2004). Rev appears to have achieved a rather extraordinary level of structural organization by flexibly utilizing its 116-amino acid chain to recognize multiple RNA sites in a homo-oligomeric RNP context.

Experimental Procedures

Plasmids, RNAs, and peptides

Proteins were expressed in *E. coli* using pHGB1, a plasmid derived from pHis-GB1-parallel1, which introduces an N-terminal His_6 tag, followed by the B1 domain of streptococcal protein G (GB1) to increase solubility, followed by a TEV cleavage site before the coding region (Harper et al., 2003). Details of pHGB1 construction can be found in Supplemental Experimental Procedures. Full length Rev (1-116) and the Rev ARM (34-50, with a C-terminal pentapeptide AAAAR to stabilize the α-helix) were cloned into pHGB1 immediately downstream of the TEV site. Mutants were generated using standard site-directed mutagenesis (Stratagene).

To generate RRE RNAs, a 242 nt portion of the RRE from pDM128 (Huang et al., 1991) was cloned between the NotI and EcoRI sites of a pBluescript-KS+ vector downstream of the T7 promoter. RNAs were produced by T7 RNA polymerase in vitro run off transcription from plasmid templates linearized by EcoRI or, for shorter fragments, using synthetic DNA templates as described in Supplemental Experimental Procedures. For radiolabeling, RNAs were dephosphorylated using alkaline phosphatase, 5' end-labeled using T4 polynucleotide kinase and γ-^{32}P ATP (New England BioLabs), and separated from unincorporated ATP using NucAway spin columns (Ambion). Long and short RNAs were purified on denaturing 4% and 10% or 12% polyacrylamide/8 M urea gels respectively and annealed by heating at 95°C for 2 min and slow cooling to room temperature in renaturation buffer (20 mM Tris pH 7.5, 100 mM NaCl).

The Rev ARM peptide (residues 34-50), capped at the N-terminus by a succinyl group and at the C-terminus by AAAAR was synthesized using Rink amide resin on a Applied Biosystems Model 433A peptide synthesizer and standard Fmoc chemistry. Wild type peptide (succ-TRQARRNRRRRWRERQRAAAAR-am) or mutants were cleaved from the resin, deprotected and purified by HPLC on a C_{18} reverse-phase column using an acetonitrile gradient in 0.1% trifluoroacetic acid (TFA).

Rev protein and peptide purification

Proteins were expressed in *E. coli* strain BL21/DE3 from pHGB1-derived vectors as N-terminal fusions with a His_6 tag, GB1 domain and a TEV-protease cleavage site. Cells were grown to $OD_{600} = 0.8$ at 37°C in LB medium with 100 μg/ml ampicillin. Isopropyl-β-D-thiogalactopyranoside was added to 1 mM to induce expression by shaking 4 h at 37°C, and harvested cells were resuspended in lysis buffer [25 mM HEPES pH 7.5, 200 mM NaCl, 0.1% Tween-20, 2 mM β-mercaptoethanol (β-ME), 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor cocktail (Roche)]. The suspension was incubated on ice for 20 min with 1 mg/ml lysozyme and frozen in liquid nitrogen. After thawing and sonicating, cell debris was removed by centrifugation at 14,000 rpm for 30 min.

Full length Rev was purified by Ni-NTA affinity chromatography and cleaved by TEV using standard procedures. Briefly, supernatant from the cell lysate was applied to Ni-NTA superflow resin (Qiagen) equilibrated with buffer A+ (50 mM Tris 8.0, 250 mM NaCl, 0.1% Tween-20, 2 mM β-ME, 10 mM imidazole). The resin was rinsed with buffer A+ then buffer A (buffer A+ without Tween-20). A stepwise elution was

performed using buffer A with increasing concentrations of imidazole. Fractions were analyzed by SDS-PAGE, pooled, and dialyzed against buffer B [40 mM Tris pH 8.0, 200 mM NaCl, 0.5 mM dithiothreitol (DTT)] at 4°C. TEV protease was added and incubated at room temperature for 1 h. The reaction was loaded on Ni-NTA resin equilibrated with buffer B to remove the His_6-tagged TEV protease, the free His_6-GB1 tag, and any uncleaved protein. Full length Rev, with an N-terminal GA dipeptide, was collected in the flow through, DTT was added to 2 mM, and protein was quantified by SDS-PAGE (generally 0.05-0.2 mg/ml) and immediately aliquoted and frozen in liquid nitrogen. A significant absorption peak at 260 nm and bands observed on ethidium bromide stained agarose gels indicated that large amounts of non-specific *E. coli* RNA co-purified with Rev under these native conditions.

A His_6-GB1 fused Rev ARM peptide was purified similarly, with a modification to remove non-specific RNA. RNase A (50 μg/ml) and T1 (50 U/ml) (Roche) and NaCl to 2 M were added to the cleared lysate and buffer A+ contained 2 M NaCl. Buffer A and elution buffers contained 250 mM NaCl, as above. Following elution, fractions were analyzed by SDS-PAGE, pooled, dialyzed in buffer B with 2 mM DTT, aliquoted, and frozen. The resulting protein retains the N-terminal His_6-GB1 tag, GAMA appended to the N-terminus of the Rev ARM, followed by AAAAR at the C-terminus. The binding affinity for RRE IIB was comparable to the synthetic Rev ARM peptide described above (data not shown).

For NMR experiments, the His_6-GB1 fused Rev ARM peptide was grown in M9 minimal media supplemented with trace minerals, thiamine, and $^{15}NH_4Cl$ as the sole nitrogen source. The protein was purified as above and incubated with TEV protease for

1 hour. The resulting peptide (GAMATRQARRNRRRRWRERQRAAAAR) was purified by HPLC as above.

RNA binding gel shift assays

Electrophoretic mobility shift assays were performed in binding buffer [10 mM HEPES pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 10% glycerol, and 50 μg/mL yeast tRNA (Sigma-Aldrich)]. Full length Rev proteins and His_6-GB1 tagged Rev ARM were diluted serially in buffer (5 mM Tris pH 8.0, 20 mM NaCl, and 200 μg/ml bovine serum albumin) and combined with an equal volume of <10 pM radiolabeled RNA. Reactions were incubated at room temperature for 20 min and loaded onto continuously running 8% or 10% polyacrylamide (37.5:1 mono:bis, 0.5x TBE) gels. Gels were run at room temperature for 1-4 h depending on the RNA size, dried and exposed to a PhosphorImaging plate for >12 h. Binding assays of synthetic Rev ARM to stem IA were performed similarly except that the peptide was serially diluted in water, and binding reactions and gel separation were performed at 4°C. Bands were quantified with a Molecular Dynamics PhosphorImager and Imagequant software. Binding constants were calculated by measuring the fraction of all bound RNA species compared to total RNA, and fitting the data to binding curves using Kaleidagraph software (Synergy Software, Reading, PA) and are reported as the mean +/- standard deviation of two or more replicates.

NMR

All ^{15}N-resolved HSQC NMR experiments were conducted at 283K on a Bruker BioSpin DRX 800 MHz spectrometer equipped with a cryoprobe. All samples contained 200 μM HPLC-purified uniformly ^{15}N labeled Rev ARM peptide in 25 mM HEPES pH 6.5, 100 mM KCl, 1 mM $MgCl_2$, 0.5 mM EDTA, and 10% D_2O, and contained stem IA or IIB RNAs at 220 μM. Spectra were processed with NMRPIPE (Delaglio et al., 1995) and analyzed with SPARKY (Goddard and Kneller, 2007).

RNA export reporter assays

For CAT assays, HeLa cells were co-transfected in 24 well plates with Polyfect (Qiagen) and 25 ng pDM128 CAT reporter plasmid (Huang et al., 1991), 5 ng pSV2-Rev expression vector, and 2.5 ng pRL-CMV luciferase plasmid (Promega) as a transfection control. After 48 h, cells were lysed with passive lysis buffer and CAT activity was measured and normalized to luciferase levels (Promega). Relative fold activation was calculated as CAT values normalized to the luciferase values, relative to reporter without Rev, and setting the activation level of wild type Rev and RRE to 100. Assays were performed in triplicate and data are presented as the mean +/- standard deviation.

For Gag Western blot analyses, HeLa cells were co-transfected with 100 ng pCMV-GagPol-RRE reporter plasmid (Srinivasakumar et al., 1997) and 100 ng pSV2-Rev expression vector. After 48 h, cells were lysed and samples were separated by SDS-PAGE, transferred to nitrocellulose and probed overnight with a mouse anti-p24 monoclonal antibody (AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH) or mouse anti-β-actin followed by goat anti-mouse IgG-HRP (Santa

Cruz Biotechnology), and developed with SuperSignal West Pico chemiluminescence (Thermo Scientific).

References

Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R., and Summers, M.F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. J. Mol. Biol. *301*, 491-511.

Askjaer, P., Bachi, A., Wilm, M., Bischoff, F.R., Weels, D.L., Ogniewski, V., Ohno, M., Niehrs, C., Kjems, J., Mattaj, I.W., and Fornerod, M. (1999). RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. Mol. Cell. Biol. *19*, 6276-6285.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.

Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. Science *273*, 1547-1551.

Bayer, T.S., Booth, L.N., Knudsen, S.M., and Ellington, A.D. (2005). Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA *11*, 1848-1857.

Charpentier, B., Stutz, F., and Rosbash, M. (1997). A dynamic in vivo view of the HIV-I Rev-RRE interaction. J. Mol. Biol. *266*, 950-962.

Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J., and Rusche, J.R. (1991). Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. Nucleic Acids Res. *19*, 1577-1583.

Cullen, B.R. (1998). Retroviruses as model systems for the study of nuclear RNA export pathways. Virology *249*, 203-210.

Cullen, B.R. (2003). Nuclear mRNA export: insights from virology. Trends Biochem. Sci. *28*, 419-424.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277-293.

Edgcomb, S.P., Aschrafi, A., Kompfner, E., Williamson, J.R., Gerace, L., and Hennig, M. (2008). Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. Protein Sci.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051-1060.

Furnes, C., Arnesen, T., Askjaer, P., Kjems, J., and Szilvay, A.M. (2005). HIV-1 Rev oligomerization is not obligatory in the presence of an extra basic domain. Retrovirology *2*, 39.

Goddard, T.D., and Kneller, D.G. (2007). SPARKY 3.112, University of California, San Francisco.

Hainzl, T., Huang, S., and Sauer-Eriksson, A.E. (2002). Structure of the SRP19 RNA complex and implications for signal recognition particle assembly. Nature *417*, 767-771.

Harper, S.M., Neil, L.C., and Gardner, K.H. (2003). Structural basis of a phototropin light switch. Science *301*, 1541-1544.

Harrison, S.C., Skehel, J.J., and Wiley, D.C. (1996). Virus structure. In Virology, B.N. Fields, ed. (Philadelphia: Lippincott-Raven), pp. 59-99.

Heaphy, S., Finch, J.T., Gait, M.J., Karn, J., and Singh, M. (1991). Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. Proc. Natl. Acad. Sci. USA *88*, 7366-7370.

Hope, T.J., McDonald, D., Huang, X.J., Low, J., and Parslow, T.G. (1990). Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. J. Virol. *64*, 5360-5366.

Huang, X., Hope, T.J., Bond, B.L., McDonald, D., Grahl, K., and Parslow, T.G. (1991). Minimal Rev-response element for Type 1 human immunodeficiency virus. J. Virol. *65*, 2131-2134.

Iwai, S., Pritchard, C., Mann, D.A., Karn, J., and Gait, M.J. (1992). Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. Nucleic Acids Res. *20*, 6465-6472.

Jain, C., and Belasco, J.G. (1996). A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. Cell *87*, 115-125.

Jain, C., and Belasco, J.G. (2001). Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Mol. Cell *7*, 603-614.

Kjems, J., Brown, M., Chang, D.D., and Sharp, P.A. (1991). Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. Proc. Natl. Acad. Sci. USA *88*, 683-687.

Kjems, J., and Sharp, P.A. (1993). The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. J Virol *67*, 4769-4776.

Landt, S.G., Ramirez, A., Daugherty, M.D., and Frankel, A.D. (2005). A simple motif for protein recognition in DNA secondary structures. J. Mol. Biol. *351*, 982-994.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479-490.

Malim, M.H., and Cullen, B.R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. Cell *65*, 241-248.

Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature *338*, 254-257.

Mann, D.A., Mikaelian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Butler, P.J., Gait, M.J., and Karn, J. (1994). A molecular rheostat: Co-operative rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. J. Mol. Biol. *241*, 193-207.

Marmorstein, R., and Fitzgerald, M.X. (2003). Modulation of DNA-binding domains for sequence-specific DNA recognition. Gene *304*, 1-12.

Nam, Y.S., Petrovic, A., Jeong, K.S., and Venkatesan, S. (2001). Exchange of the basic domain of human immunodeficiency virus type 1 Rev for a polyarginine stretch expands the RNA binding specificity, and a minimal arginine cluster is required for optimal RRE

RNA binding affinity, nuclear accumulation, and trans-activation. J. Virol. *75*, 2957-2971.

Perez, I., McAfee, J.G., and Patton, J.G. (1997). Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. Biochemistry *36*, 11881-11890.

Pollard, V.W., and Malim, M.H. (1998). The HIV-1 Rev Protein. Annu. Rev. Microbiol. *52*, 491-532.

Shamoo, Y., Abdul-Manan, N., and Williams, K.R. (1995). Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res. *23*, 725-728.

Shen, H., and Green, M.R. (2004). A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. Mol. Cell *16*, 363-373.

Smith, C.A., Calabro, V., and Frankel, A.D. (2000). An RNA-binding chameleon. Mol. Cell *6*, 1067-1076.

Srinivasakumar, N., Chazal, N., Helga-Maria, C., Prasad, S., Hammarskjold, M.L., and Rekosh, D. (1997). The effect of viral regulatory protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. J. Virol. *71*, 5841-5848.

Symensma, T.L., Baskerville, S., Yan, A., and Ellington, A.D. (1999). Polyvalent Rev decoys act as artificial Rev-responsive elements. J. Virol. *73*, 4341-4349.

Tan, R., Chen, L., Buettner, J.A., Hudson, D., and Frankel, A.D. (1993). RNA recognition by an isolated alpha helix. Cell *73*, 1031-1040.

Tan, R., and Frankel, A.D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. Biochemistry *33*, 14579-14585.

Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G., and Cullen, B.R. (1992). Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. Proc. Natl. Acad. Sci. USA *89*, 758-762.

Wang, Y., and Jardetzky, O. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci. *11*, 852-861.

Wen, W., Meinkoth, J.L., Tsien, R.Y., and Taylor, S.S. (1995). Identification of a signal for rapid export of proteins from the nucleus. Cell *82*, 463-473.

Xu, W., and Ellington, A.D. (1996). Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. Proc. Natl. Acad. Sci. USA *93*, 7475-7480.

Ye, X., Gorin, A., Frederick, R., Hu, W., Majumdar, A., Xu, W., McLendon, G., Ellington, A., and Patel, D.J. (1999). RNA architecture dictates the conformations of a bound peptide. Chem. Biol. *6*, 657-669.

Zappulla, D.C., and Cech, T.R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc. Natl. Acad. Sci. USA *101*, 10024-10029.

Zemmel, R.W., Kelley, A.C., Karn, J., and Butler, P.J. (1996). Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. J. Mol. Biol. *258*, 763-777.

Acknowledgments

We thank John Gross, Mark Kelly and members of the Frankel and Gross laboratories for helpful suggestions, and John Gross, Geeta Narlikar, JJ Miranda and David Booth for critical reading of the manuscript. M.D.D. is a Howard Hughes Medical Institute predoctoral fellow. I.D. is a postdoctoral fellow of the Human Frontiers Science Program and Fundación Antorchas. This work was supported by NIH grants P50GM82250 and P0GM56531 to A.D.F.

Figure Legends

Figure 1: Rev protein and RRE RNA

A) Domain structure of Rev. The sequence of the ARM (arginine-rich motif) is shown, with residues that contact the IIB site indicated in red. Shown below is the extent of conservation of ARM residues based on 747 HIV-1 sequences (Los Alamos HIV sequence database (http://www.hiv.lanl.gov/)) and graphed as percent identity at each position, with position 39 allowing either Arg or Lys (+). Green bars highlight positions with >95% conservation. B) Predicted secondary structure of the RRE used in this study. Stem nomenclature is from Charpentier et al. (Charpentier et al., 1997), and numbering corresponds to the originally described RRE structure (Malim et al., 1989). The dashed line indicates the boundaries of the 242-nt fragment for in vitro binding assays. Red bases indicate mutated positions used as specificity controls.

Figure 2: Effects of RRE length and oligomerization mutations on Rev-RRE assembly

A) Gel shift assays using purified Rev at the concentrations indicated and <10 pM radiolabeled RRE fragments. Nucleotide lengths and schematics of the RNA structures are shown. The species indicated are: F, Free RNA; M, Rev monomer; D, Rev dimer; Mu, Rev multimer. B) Binding curves calculated from the gel shifts as fraction of RNA bound (total bound RNA in all bands divided by the sum of bound and unbound RNAs). C) Gel shift assays using purified His-GB1 fused Rev ARM peptide or L18Q mutant Rev protein at the concentrations indicated. The red X in the protein schematic shows the location of the L18Q oligomerization domain mutant. D) Binding curves for

oligomerization-deficient protein or peptide as calculated in B. E) Dissociation constants and Hill coefficients (*n*) for Rev or mutants and different size RRE fragments of the RRE, calculated using the equation: FractionBound = $[Rev]^n/(K_D{}^n + [Rev]^n)$. Data are presented as the mean +/- standard deviation from at least two replicates. We note that the Hill coefficients for the shorter RNA fragments are less than 1. Such behavior can be modeled if competing reactions exist that affect apparent concentrations, such as protein aggregation. Nonetheless, the Hill coefficients increase with increasing RNA length, and the calculated value of 1.6 for the full-length RRE may underestimate the degree of cooperativity if competing reactions occur under these conditions.

Figure 3: Extended Rev binding specificity within the RRE

A) Binding curves for Rev and a minimal 34-nt stem IIB RNA, using wild type stem IIB, an A73G mutant, and antisense stem IIB and calculated as in Figure 2. B) Binding curves as in A, but using full-length RRE RNAs or mutants as indicated. C) Calculated dissociation constants, as in Figure 2, and specificity values, calculated as the ratio of observed dissociation constants to that of the same length antisense RNA.

Figure 4: Rev binding to a site in stem IA

A) Gel shift assays of the Rev ARM peptide at the concentrations indicated with radiolabeled stem IA wild type or mutant RNAs. Mutated bases are shown in red. Apparent binding constants (K_D in nM) are: Stem IA = 270 ± 9.7, ΔBulge = 1200 ± 62, and 5' Purine Mutant = 1500 ± 57. B) Gel shift assays of the full-length Rev protein with stem IA or the 5' purine mutant RNA. The species indicated are: F, Free RNA; M, Rev

monomer; D, Rev dimer; Mu, Rev multimer. C) Binding curves for Rev and the full-length RRE or stem IA mutant. Apparent binding constants (K_D in nM) are: Wild type RRE = 0.12 ± 0.0067 and Stem IA 5' Purine Mutant RRE = 0.61 ± 0.072.

Figure 5: Differences in Rev peptide binding modes to stems IA and IIB

A) Gel shift assays of Rev ARM peptide mutants at the concentrations indicated with radiolabeled stem IA RNA. Apparent binding constants (K_D in nM) are: Rev ARM N40A = 260 ± 9.8 and Rev ARM R41A = 550 ± 30. B) Helical wheel projections of the Rev ARM α-helix, with positions that reduce affinity by >2-fold (stem IA) or >10-fold (stem IIB) upon mutation to alanine circled in red. Positions circled in black (stem IA) indicate no effect of the mutation on binding affinity. Approximate RNA interfaces, indicated in red, were constructed from mutagenesis and HSQC NMR data (see below) for stem IA and from NMR structural data (Battiste et al., 1996) for stem IIB. C) ^{15}N HSQC spectra of the Rev ARM peptide alone (green), bound to stem IIB (red), or bound to stem IA (blue). The Nε proton resonance of W45 is indicated. Inset shows the region corresponding to the Nε protons of arginines in the folded portion of the spectrum.

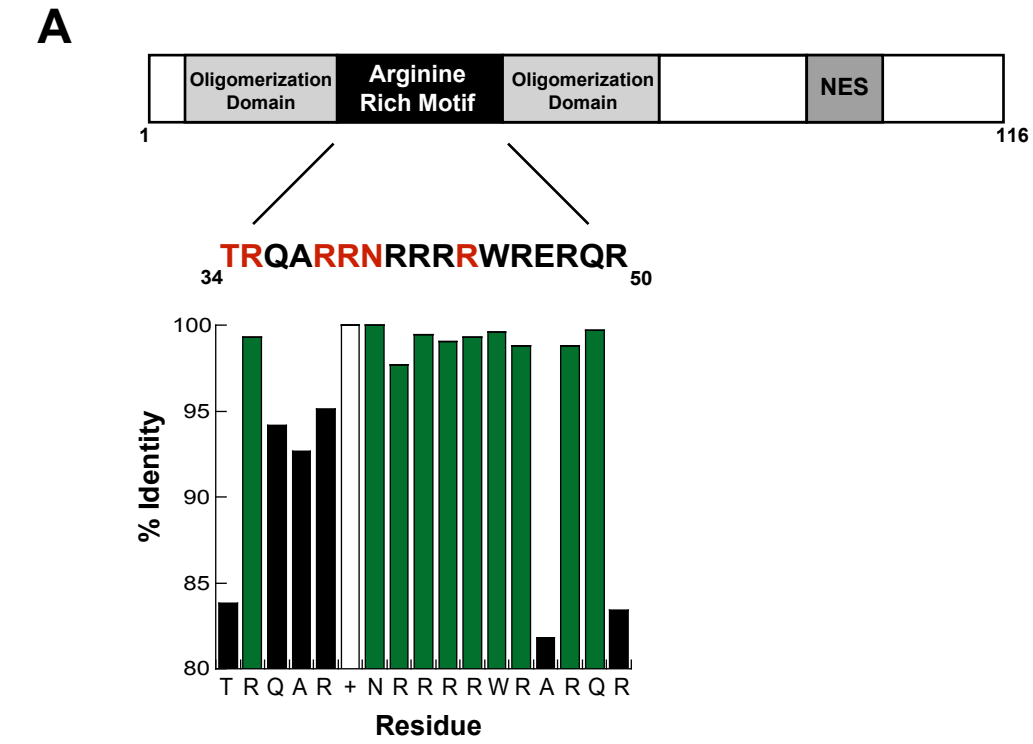
Figure 6: Oligomeric RNA-binding affinity and export activity correlate for protein or RRE mutants

A) Binding curves derived from gel shift assays with wild type Rev protein or the mutants indicated and full-length RRE, calculated as in Figure 2. B) RNA export activity of wild type (black) and mutant (grey) proteins in the CAT reporter assay. Values are normalized to the activity of wild type Rev and presented as the mean +/- standard

deviation. C) RNA export activity with wild type (black) or mutant (grey) RREs, engineered with stem IIB (A73G), stem IA (5' purine) or double (IIB and IA) mutations. D) Gag expression from pCMV-GagPol-RRE using wild type or mutant Rev proteins as indicated, measured by Western blot. The Gag precursor (p55) and proteolyzed products (p41/39 and p24) are labeled. Bottom, β-actin controls. E) Same as in D, but using wild type Rev and wild type or mutant RREs as indicated. F) Correlation between export activities measured by CAT assays (black bars) and dissociation constants determined by gel shift assays (grey bars) for the wild type and mutant Rev-RRE complexes indicated.

Figure 7: Strategies for assembling protein-RNA complexes

Schematic representation of three strategies for forming protein-RNA assemblies and aspects of each that contribute to forming the homo-oligomeric Rev-RRE RNP. Solid blue lines represent RNA, with protein binding sites highlighted in red. Different shading of the binding sites indicates that the sequence or structural motifs differ between sites. Colored ovals represent different proteins or protein domains. In the Rev-RRE complex, different shading of the Rev subunits represents different binding modes at several sites in the RRE.


A
Oligomerization Domain
Arginine Rich Motif
Oligomerization Domain
NES
1
116
TRQARRNRRRRWRERQR
34
50
100
95
90
85
80
% Identity
T R Q A R + N R R R R W R A R Q R
Residue

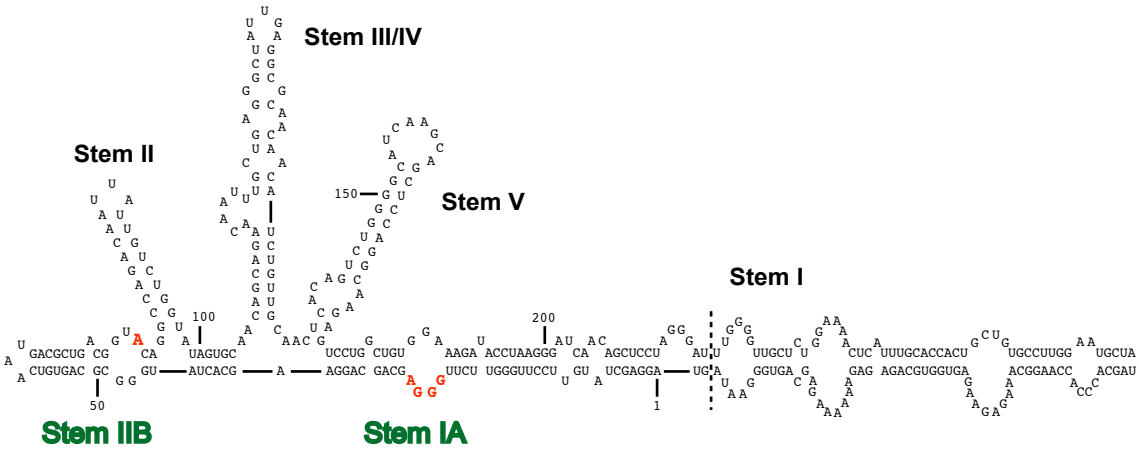
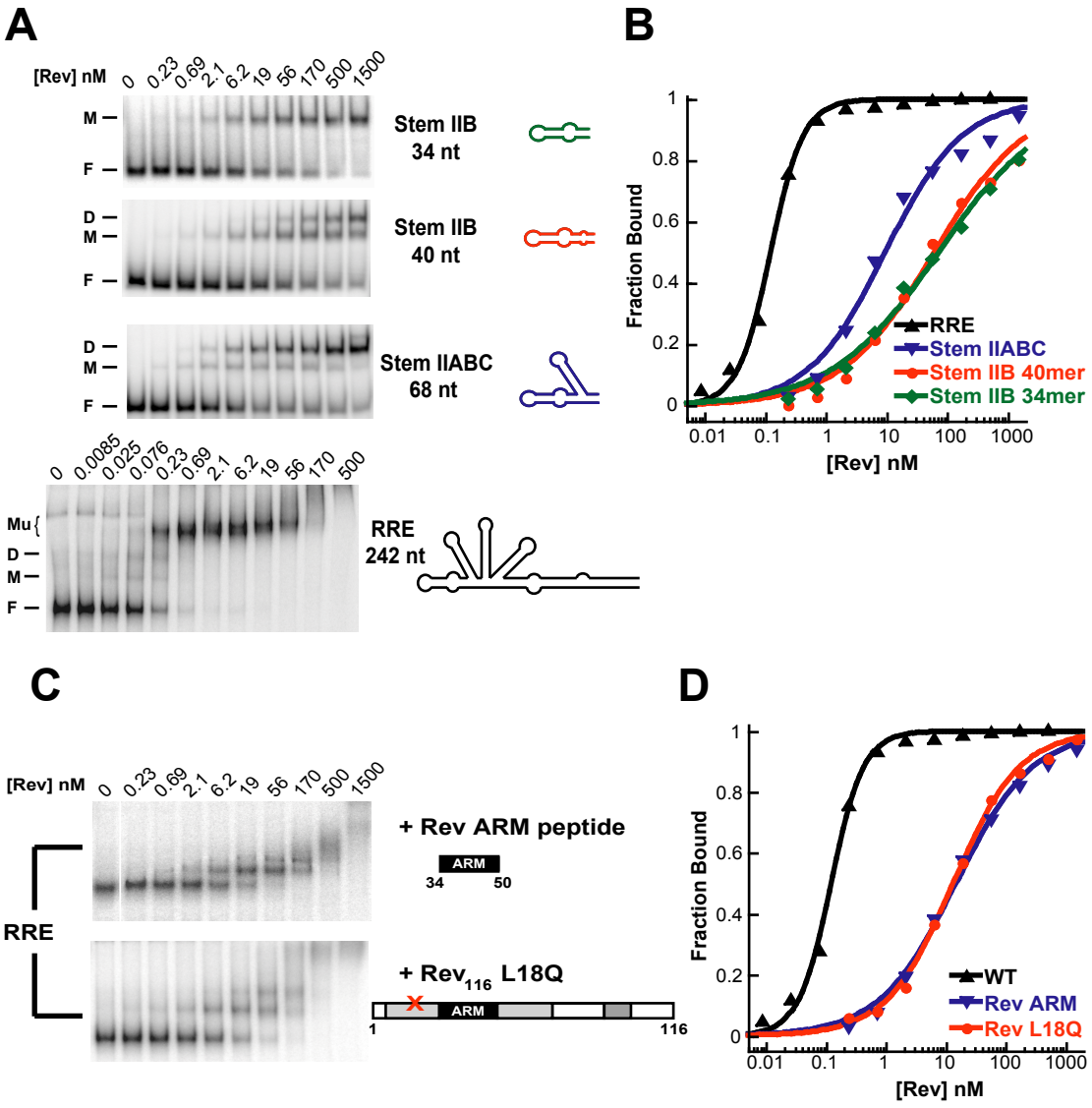

B
Stem III/IV
Stem II
Stem V
Stem I
Stem IIB
Stem IA

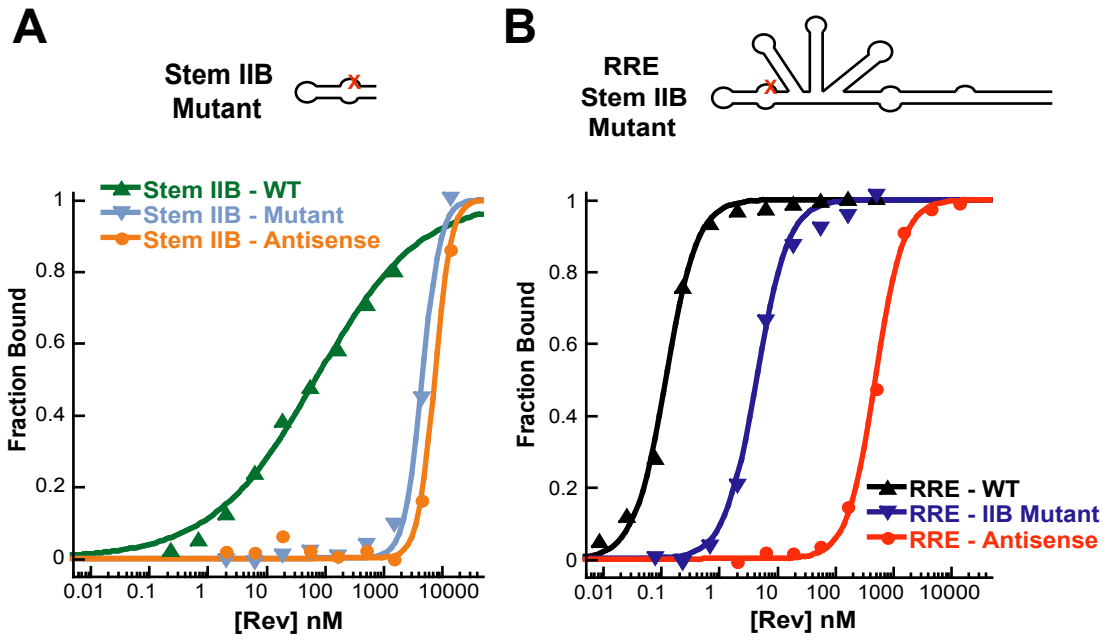
Figure 1



E

Protein	RNA	K_D (nM)	Hill Coefficient (n)
WT	RRE	0.12 ± 0.00062	1.6 ± 0.061
	IIABC	9.4 ± 0.11	0.65 ± 0.047
	IIB 40mer	54 ± 9.9	0.57 ± 0.0046
	IIB 34mer	73 ± 1.3	0.49 ± 0.0011
Rev ARM	RRE	13 ± 2.0	0.73 ± 0.064
L18Q	RRE	15 ± 1.7	0.79 ± 0.042

Figure 2



C

RNA		K_D (nM)		Specificity
IIB - 34mer	WT	73	± 1.3	110
	IIB Mutant	5000	± 380	1.6
	Antisense	8000	± 520	
RRE	WT	0.12	± 0.00062	4400
	IIB Mutant	4.8	± 0.48	110
	Antisense	530	± 38	

Figure 3

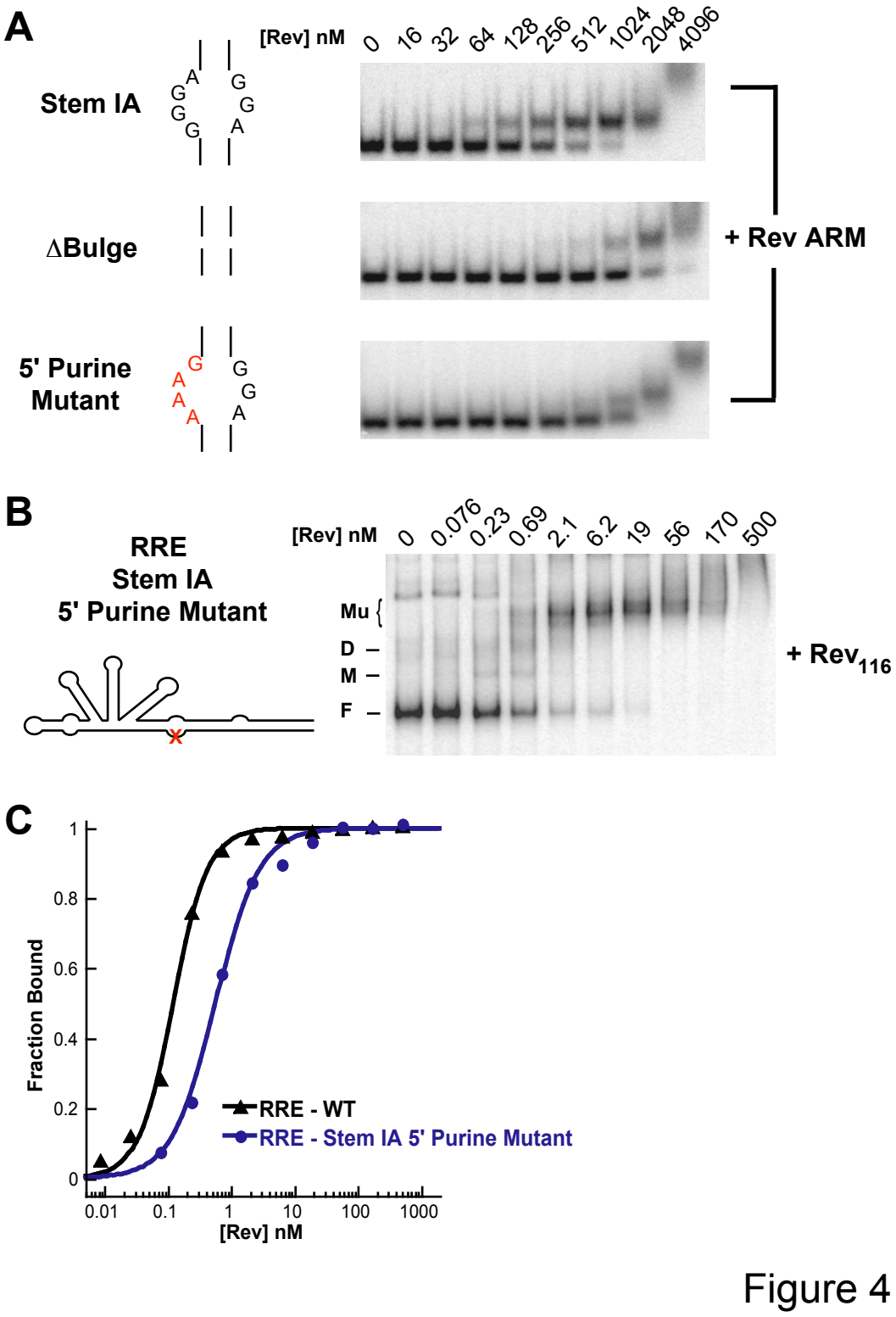

A
[Rev] nM
0
16
32
64
128
256
512
1024
2048
4096
Stem IA
ΔBulge
5' Purine Mutant
+ Rev ARM
B
RRE
Stem IA
5' Purine Mutant
[Rev] nM
0
0.076
0.23
0.69
2.1
6.2
19
56
170
500
Mu
D
M
F
+ Rev116
C
Fraction Bound
0
0.2
0.4
0.6
0.8
1
0.01
0.1
1
10
100
1000
[Rev] nM
RRE - WT
RRE - Stem IA 5' Purine Mutant

Figure 4

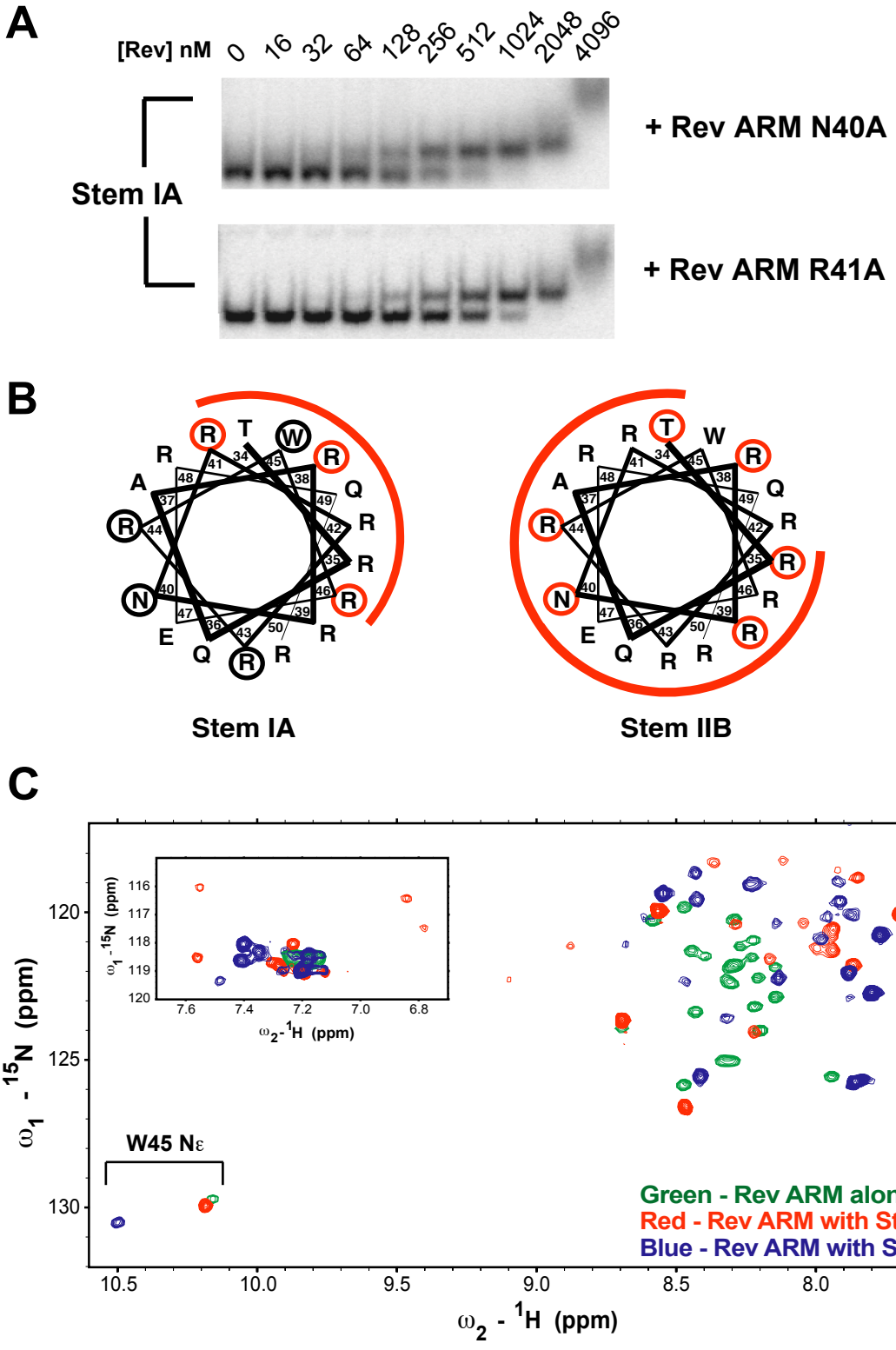

A
[Rev] nM 0 16 32 64 128 256 512 1024 2048 4096
Stem IA
+ Rev ARM N40A
+ Rev ARM R41A
B
Stem IA
Stem IIB
C
W45 Nε
Green - Rev ARM alone
Red - Rev ARM with Stem IIB
Blue - Rev ARM with Stem IA
ω1 - 15N (ppm)
ω2 - 1H (ppm)

Figure 5

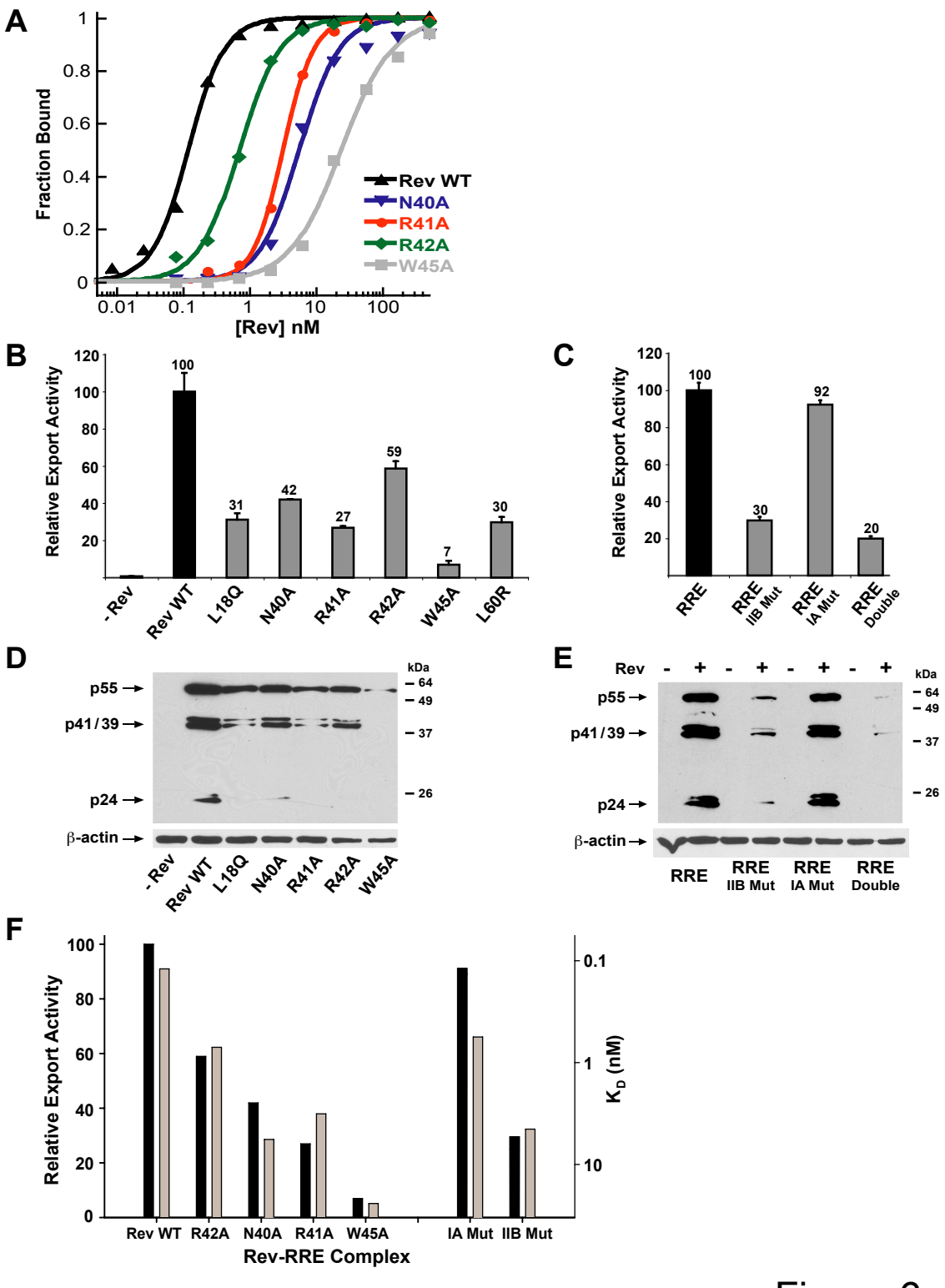

A
Fraction Bound
[Rev] nM
Rev WT
N40A
R41A
R42A
W45A
B
Relative Export Activity
- Rev
Rev WT
L18Q
N40A
R41A
R42A
W45A
L60R
C
Relative Export Activity
RRE
RRE IIB Mut
RRE IA Mut
RRE Double
D
kDa
p55
p41/39
p24
β-actin
E
Rev
kDa
p55
p41/39
p24
β-actin
RRE
RRE IIB Mut
RRE IA Mut
RRE Double
F
Relative Export Activity
K_D (nM)
Rev WT
R42A
N40A
R41A
W45A
IA Mut
IIB Mut
Rev-RRE Complex

Figure 6

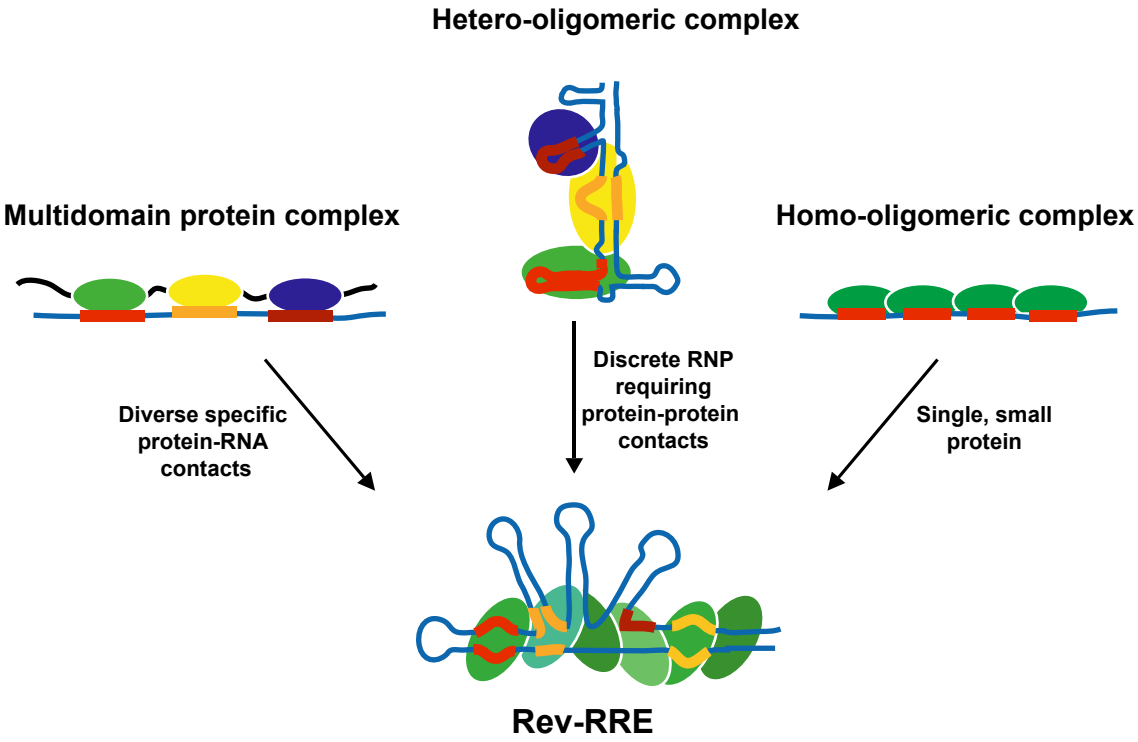

Hetero-oligomeric complex
Multidomain protein complex
Homo-oligomeric complex
Diverse specific protein-RNA contacts
Discrete RNP requiring protein-protein contacts
Single, small protein
Rev-RRE

Figure 7

Supplemental Data

A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA

Matthew D. Daugherty, Iván D'Orso, and Alan D. Frankel*

Supplemental Experimental Procedures

Construction of pHGB1

To increase solubility of expressed proteins, pHGB1 was constructed from the original pHis-GB1-parallel1 plasmid (Harper et al., 2003). The new plasmid replaces a linker between the His_6 tag and the GB1 domain from pHis-GB1-parallel1 with a shorter, more hydrophilic sequence for additional solubility and was constructed by inserting a DNA oligonucleotide (CATATGGGCCACCATCACCATCACCATAGCAGCG-GCGGTATGCAGTACAAGCTT) encoding the His_6 tag and linker into NdeI and HindIII sites upstream of the GB1 domain.

RNA transcription

RNAs shorter than 70 nts were produced using synthetic DNA templates, following standard protocols (Milligan and Uhlenbeck, 1989). All DNA templates were annealed in 56mM $MgCl_2$ with an equal concentration of T7 promoter DNA (TAATACGACTCACTATAG) by heating to 95°C and slow cooling to room

temperature. The following DNA oligonucleotide sequences were used as templates for RNA transcription:

IIB-34mer:

GGCCTGTACCGTCAGCTTGCGCTGCGCCCAGACCCTATAGTGAGTCGTATTA

IIB-34mer A73G:

GGCCTGcACCGTCAGCTTGCGCTGCGCCCAGACCCTATAGTGAGTCGTATTA

IIB-34mer Antisense:

GGTCTGGGCGCAGCGCAAGCTGACGGTACAGAGGCTATAGTGAGTCGTATTA

IIB-40mer:

GGAGGCCTGTACCGTCAGCTTGCGCTGCGCCCATACCTCCCTATAGTGAGTCG TATTA

IIB-42mer:

GGAGGCCTGTACCGTCAGCTTGCGCTGCGCCCATatACCTCCCTATAGTG AGTCGTATTA

IIB-68mer:

GGCACTATACCAGACAATAATTGTCTGGCCTGTACCGTCAGCGTCAATGACA CTGCGCCCATAGTGCCCTATAGTGAGTCGTATTA

Supplemental References

Harper, S.M., Neil, L.C., and Gardner, K.H. (2003). Structural basis of a phototropin light switch. Science *301*, 1541-1544

Milligan, J.F., and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Meth. Enzymol. *180*, 51-62.

Supplemental Figures

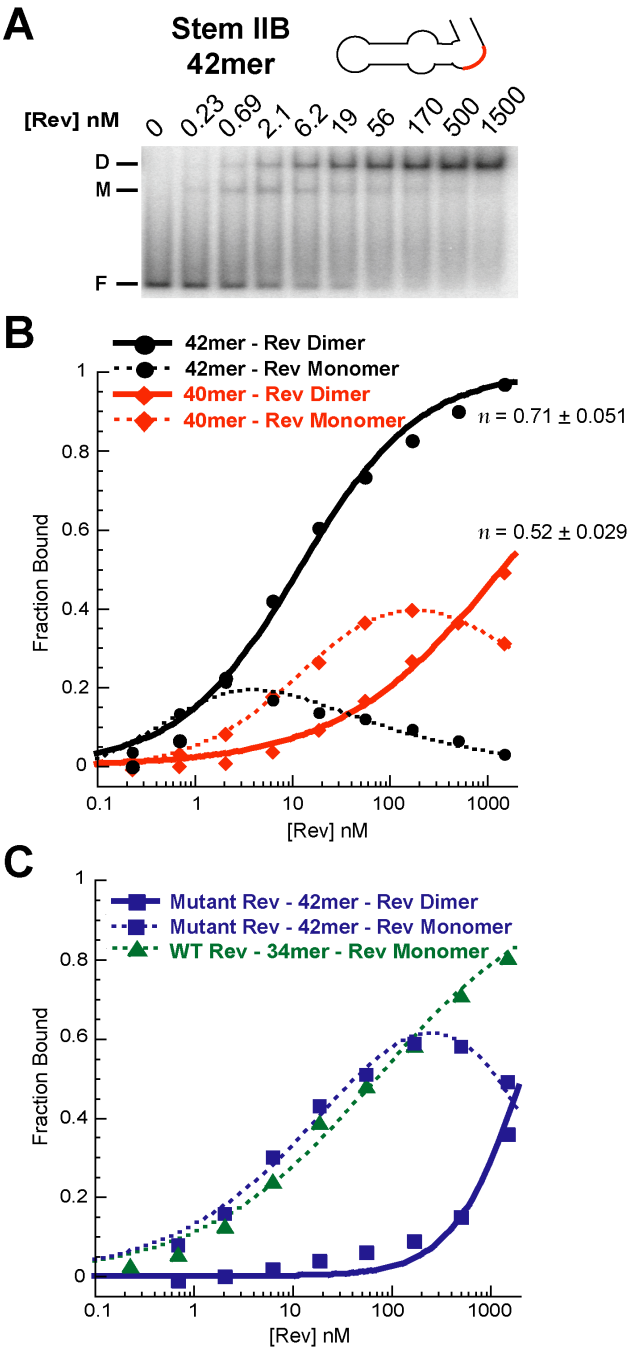


Figure S1: Cooperative binding of Rev depends on the RNA structure

A) Gel shift assays with Rev and a IIB 42mer RNA in which two nucleotides (highlighted in red) were added into an internal loop of the 40mer RNA shown in Figure

2A. Abbreviations: F, Free RNA; M, Rev monomer; D, Rev dimer. B) Binding curves calculated from the gel shifts in which monomer and dimer complexes with the 42mer were quantified separately. For comparison, binding to the IIB 40mer RNA (Figure 2A) was analyzed similarly. Hill coefficients (*n*) for the dimer species are shown. C) Binding curves for the L18Q/L60R Rev oligomerization mutant and IIB 42mer, compared to wild type Rev and the IIB 34mer (Figure 2A) which forms a single, non-cooperative complex.

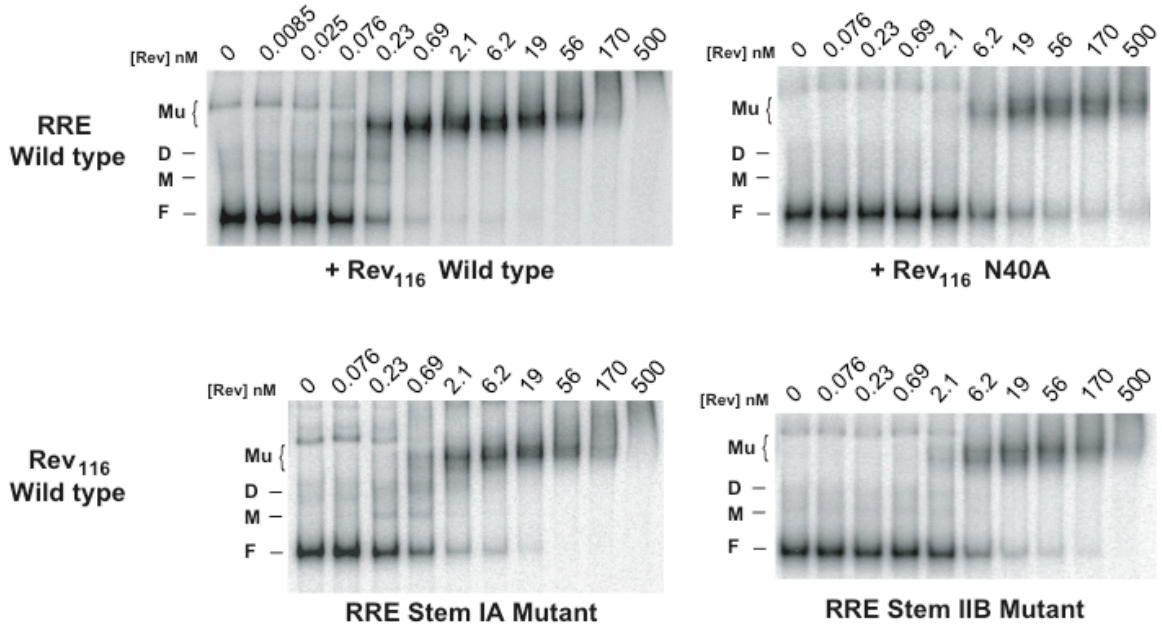


Figure S2: Reduced accumulation of intermediates in mutant Rev-RRE complexes deficient in IIB binding

Gel shift assays with the wild type and mutant Rev-RRE complexes indicated. The two right panels contain Rev or RRE mutations that reduce binding to the IIB binding site and show substantially lower levels of intermediate complexes, with expected positions of monomer (M) and dimer (D) shown. The two left panels are from Figures 2A and 4B for comparison. Abbreviations: F, Free RNA; Mu, Rev multimer.

Chapter 3

Assembly and structural insight into discrete HIV Rev-RNA complexes

Matthew D. Daugherty[1,3], David S. Booth[2,3], Yifan Cheng[3], and Alan D. Frankel[3,*]

[1]Chemistry and Chemical Biology Graduate Program
[2]Graduate Group in Biophysics
[3]Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, CA 94158

[*]Contact: Alan Frankel
Department of Biochemistry and Biophysics
UCSF
600 16th St.
San Francisco, CA 94158
415-476-9994
frankel@cgl.ucsf.edu

Abstract:

Replication of HIV requires the assembly of the viral Rev protein oligomer on the Rev Response Element (RRE) RNA. Biophysical and structural characterization of these complexes has remained elusive due to the low solubility and inability to control the oligomerization state of Rev. Here we show that high solubility of Rev relies on an 'electrostatic sink', which can be in the form of specific or non-specific RNAs, salt counterions or, interestingly, a negatively charged fusion domain. Moreover, specific RRE derived RNAs can precisely control the oligomerization state of Rev, allowing characterization of defined multimeric Rev-RNA complexes. Using these methods, we assemble homogenous particles of Rev with fragments of the RRE for analysis by NMR and small angle X-ray scattering (SAXS), and with the RRE for electron microscopy imaging. These results provide structural restraints on Rev-RNA complexes, and demonstrate that the functional Rev-RRE complex is a discrete particle similar to other structurally characterized RNPs.

Introduction:

Complex retroviruses, such as HIV, encode the essential regulatory protein Rev to facilitate nuclear export of the viral RNA genome. Rev binds to and oligomerizes on the Rev Response Element (RRE) RNA found in the introns of partially and fully unspliced viral mRNAs, and directs their transport to the cytoplasm before splicing is completed[1-3]. In this way, assembly of the Rev-RRE ribonucleoprotein (RNP) is coupled to progression through the viral life cycle.

Specific binding and oligomerization of the Rev protein on the RRE is required for functional RNA export. Initial studies defined one site within the RRE, known as stem IIB, that is required for in vivo function as well specific binding to the RRE in vitro[4-10]. Rev binding to stem IIB is mediated by the arginine-rich motif (ARM) within the protein, and the interaction between the ARM peptide and stem IIB is well understood at the biochemical and structural level[11-13]. However, binding to stem IIB is not sufficient for RNA export[14,15]. Full RNA export requires a large portion of the ~350 nucleotide RRE, as well as the oligomerization domains of Rev (Figure 1A)[3,6,15-17]. We have recently shown that these same requirements are necessary for Rev-RRE assembly in vitro[18]. The assembly of the Rev homooligomer on the RRE is a highly cooperative and high affinity process, nearly 500-fold higher affinity than binding to stem IIB. Moreover, there is a strong correlation between in vitro Rev-RRE complex assembly and RNA export activity in vivo, demonstrating that proper formation of the RNP is essential for its biological function[18].

Despite this wealth of functional data on Rev-RRE complexes, molecular structures of the Rev protein and Rev-RRE complexes have remained elusive due to the low solubility of the protein in vitro at concentrations higher than 1-5 uM[5,19]. Circular dichroism studies[20,21] suggest that the protein adopts a helix-turn-helix motif, with a relatively unstructured C-terminus. Attempts to characterize the oligomerization state of Rev, either in the absence or presence of RNA, have led to inconsistent results[19,22] although biochemical mapping of the oligomerization domains have led to models for Rev structure and oligomerization[23,24], and recently a model for EIAV Rev has been computationally derived[25]. However, without access to soluble species of homogenous Rev-RNA complexes, experimental structural determination has been unattainable. It has also been observed that Rev could be induced to form filaments at approximately 5 uM by slowly warming the solution from 4°C to 25°C over several hours in the presence or absence of RNA[5,19]. Similar filaments form by slowly concentrating a Rev or Rev-RNA solution over several days at 4°C, and these fibril assemblies exhibit the helix-turn-helix topology previously predicted for Rev[26,27]. The relevance of these assemblies, however, has been difficult to reconcile due to the lack of biological significance[15,27].

Our previous biochemical and functional characterization of the high affinity Rev-RRE complex contrasted with numerous previous reports that found little to no increase in affinity of Rev binding to the full RRE over Rev binding to stem IIB alone. The inconsistency appeared to be due to differences in preparation of the protein for biochemical studies; our studies used a fully native preparation of the protein that even retained endogenous *E. coli* RNA throughout purification, while many previous efforts had used denaturing preparations or Rev purified free of RNA. Our supposition thus has

been that retaining RNA during purification under non-denaturing conditions keeps the protein in a more native state that allows full use of the oligomerization domains to assemble a highly cooperative and high affinity complex on its RNA. Importantly, this biochemical activity also closely reflects RNA export activity, further supporting the importance of a fully native preparation of the Rev protein for in vitro studies[18].

Due to the fact that the manner in which the Rev protein is expressed and purified has a drastic effect on the biochemical activity of the protein in vitro, we hypothesized that it may also have different biophysical characteristics. Specifically, we wished to ask whether natively prepared Rev-RNA complexes might be more soluble and thus amenable to structural characterization. Here we show that natively purified Rev is soluble up to concentrations that are two to three orders of magnitude higher than previously reported. Solubility of Rev relies on the presence of a negatively charged 'electrostatic sink' that can be presented to Rev in the form of either non-specific or specific RNA, oxyanionic salt counterions or a negatively charged fusion domain, GB1. Interestingly, this fusion domain, from the B1 domain of streptococcal protein G[28], interacts with the Rev protein to stabilize it in the absence either RNA or oxyanionic salts. These interactions keep the protein from irreversibly precipitating, and may thus be a generalizable solution to preventing aggregation of highly charged proteins, especially those that bind to nucleic acids. We also show that addition of specific RRE derived RNAs can control the oligomerization state of Rev. Importantly, combining protein mutants with these specific RNAs yields discrete complexes of Rev-RNA that have a defined stoichiometry and are highly soluble in solution. Analysis of these complexes by NMR demonstrates that the C-terminus is disordered, although additional structural

characterization by NMR is limited by exchange broadening of numerous peaks. We also employ small-angle X-ray scattering (SAXS) to determine a molecular envelope for the Rev dimer bound to RNA which is consistent both in size and shape with the recent crystal structure of the Rev dimer (Daugherty et al., in preparation) and the NMR structure of the bound IIB RNA[11]. Finally, we assemble discrete complexes of Rev with the entire RRE for imaging by electron microscopy. Individual Rev-RRE particles display intriguing lobed features and a size consistent with the measured mass calculations of six Rev monomers bound to the RNA. These particles clearly demonstrate that functional Rev-RRE complexes, which have been shown to cooperatively assemble with high affinity, are discrete in nature similar to other structurally characterized RNPs.

Results:

Rev requirements for high solubility

We recently described the biochemical characterization of Rev-RRE complexes that assemble with high cooperativity and high affinity[18]. These complexes exhibit requirements for formation that are the same as requirements for Rev-mediated RNA export, suggesting that they are related to complexes that form in vivo. Since this biochemical activity had not been described previously, we reasoned that our method of preparation of the protein might be important for this activity, and that this might represent a more native state of the protein. As natively prepared proteins are usually more soluble than unfolded proteins, we also surmised that this same method of preparation might allow for higher solubility of the Rev protein. Rev has been a particular challenge, as previous methods of preparation have allowed an upper solubility limit of only approximately 1-5 uM[5,19].

Our described method[18] of preparation differs in two important aspects than previously described methods. First, the protein is expressed as a fusion construct with GB1, a commonly used expression tag derived from the well folded B1 domain of streptococcal protein G[28]. Second, after Rev was cleaved away from GB1, the protein was never purified free of RNA, including endogenous RNAs from the host *E. coli* from which it was purified. To determine whether GB1 alone could increase the solubility of Rev, GB1-fused Rev (termed HG-Rev) was purified natively from *E. coli* lysates with all endogenous *E. coli* RNA removed. High yields of soluble HG-Rev protein (typically 50-200 uM) were eluted from standard Ni-NTA purification (Figure 1C, lane 2),

representing a increase in the upper solubility limit of Rev of over 100-fold, even in the absence of RNA. Interestingly, at 100uM when the GB1 domain was cleaved away from Rev and no RNA was present, visible precipitate began forming almost immediately and increased over time. Centrifugation of the precipitate identified the aggregated protein as cleaved Rev (Figure 1C, lane 5). Thus, fusion with the GB1 domain greatly increases the upper solubility limit of Rev, but this solubility is dependent on GB1 being present *in cis* with Rev.

As our previous biochemical data relied on cleaved Rev, but with RNA present, we wished to ask whether RNA could support high Rev solubility when added *in trans*. When HG-Rev was cleaved in the presence of approximately equimolar amounts of non-specific RNA, the GB1 domain was dispensable for high Rev solubility (Figure 1C, lane 17). These results demonstrate that either RNA added *in trans* or the GB1 domain *in cis* helps to solubilize the Rev protein. One possible explanation of these similar solubility enhancing effects from different macromolecules relies on the overall electrostatic properties of the molecules involved. Rev is overall positively charged due to the large number of arginine and lysine residues, with a net charge of +3 at neutral pH and a predicted pI of 9.23. In contrast, both GB1 and RNA contain a large overall negative charge, with the GB1 domain having a net charge of –4 at neutral pH due to glutamic and aspartic acid side chains and a predicted pI of 4.55. We thus surmised that charge neutralization by oxyanionic counterions might be an important aspect of this stability of the protein, and asked whether appropriate salt counterions might also support high Rev solubility. Although increasing the amount of NaCl to 500 mM did not prevent precipitation of Rev (data not shown), the addition of either 100 mM potassium

phosphate or sodium sulfate supports Rev solubility in the absence of the fused GB1 domain (Figure 1C, lanes 9 and 13 respectively). Thus, the solubility of Rev relies on charge neutralization by an 'electrostatic sink', which may be presented in the form of a fused protein domain, non-specific RNA or oxyanionic salt counterions. Importantly, these observations allow for Rev to reach concentrations that enable previously unfeasible studies of Rev assembly and structure.

Specific RNAs control the oligomerization state of Rev

One outstanding question regarding Rev structure and function is the oligomerization state of the protein. Numerous attempts have been made to define the state of the protein, both with and without RNA[19,22]. The oligomerization state of HG-Rev in the absence of RNA is highly heterogeneous and is concentration dependent as measured by analytical size exclusion chromatography (SEC). At a concentration of 500 uM (11 mg/ml), HG-Rev elutes from the column as a wide peak, indicative of a broad distribution of oligomeric states (Figure 2A). Coupled multi-angle light scattering (MALS) and refractive index measurements of the maximal absorption peak establish a molar mass of 210kDa (Figure 2B and Table 1). This mass corresponds to 9-10 HG-Rev monomers. At a concentration of 200 uM, the protein still elutes as a broad distribution of sizes, but with a slightly smaller peak mass of 150 kDa, or 6-7 HG-Rev monomers (Figure 2A and 2B, and Table 1).

Previous biochemical data has revealed that at concentrations in which the Rev protein can multimerize on the entire RRE, only monomers or dimers form on hairpin elements of the RRE[7,15,18]. We also noted that the kinetics of TEV cleavage of the GB1

domain from Rev is accelerated in the sample containing tRNA (Figure 1C), perhaps indicating a greater accessibility of the TEV cleavage site to the protease that would result from a smaller HG-Rev complex. These data prompted us to consider the possibility that short elements of the RRE may be able to define the oligomerization state of Rev even at high concentrations of protein. We therefore performed the same analytical SEC experiments with HG-Rev, including increasing amounts of a minimal stem IIB element from the RRE (labeled IIB 34). As the protein:RNA stoichiometry is adjusted from 4:1 to 1:1, a drastic alteration of elution peak and shape is observed for protein-RNA complexes (Figure 2C). Importantly, SEC with equal stoichiometry of protein and RNA (Figure 2C, bottom) displays the longest retention on the column and the sharpest elution profile, indicative of the smallest and most defined complex. Native gel analysis of fractions from these experiments supports the increasingly distinct nature of complexes as the concentration of RNA is increased (Figure 2D). To confirm that increased concentrations of RNA yield smaller complexes, the molar masses of prominent peaks from SEC were measured. As the concentration of IIB 34 RNA is increased, the size of the Rev-RNA complex is reduced from 240 kDa to 63 kDa (Figure 2E and Table 1). Although we can not accurately measure the fraction of this mass that corresponds to RNA, the measured molar mass, native gel migration and significant free RNA peak in the SEC peak suggests that at a 1:1 stoichiometry, two to three HG-Rev monomers are bound to a single IIB RNA element (Table 1).

Assembly of discrete monomeric and dimeric Rev-RNA complexes

To further understand the molecular basis for Rev oligomerization and to proceed with structural studies, we wished to further define the stoichiometrys of Rev-RNA complexes. Previous biochemical dissection of the oligomerization domains defined two important classes of mutations that disrupt Rev multimerization, which suggested that there are two separable surfaces for Rev oligomerization[24]. Mutation of residues in one of these classes, represented by the L60R mutation, reduces formation of multimeric species larger than a dimer on the RRE. Mutation of residues in another class, represented by the L18Q mutation, reduces formation of species higher than a monomer on the RRE. Thus, we hypothesized that mutations in the oligomerization surfaces of Rev, in combination with the minimal stem IIB element, could yield a single discrete species of monomeric Rev on RNA.

Mutation of both oligomerization surfaces (L18Q/L60R) in the absence of RNA did not specifically define a single monomeric species of HG-Rev, as may have been predicted. Instead, a distribution of species was seen from analytical SEC (Figure 3A, top), as is seen above for wildtype protein in the absence of RNA (Figure 2A). Molar mass measurements of the two primary species yield masses of approximately 100 kDa and 45 kDa, corresponding to 4-5 monomers and 2-3 monomers respectively (Figure 3B, bottom and Table 1). Likewise, the wildtype Rev protein, in the absence of the GB1 domain but in the presence of an excess of the minimal stem IIB RNA, yielded a distribution of species by SEC (Figure 3A, middle). Native gel analysis and molar mass calculations show a mixture of species corresponding to 2-3 Rev monomers per RNA (Figure 3B and Table 1), similar to what was observed with the wildtype HG-Rev in the

presence of an equal stoichiometry of RNA (Figure 2D and 2E, and Table 1). Although the above described individual changes did not define a single Rev species, when mutations in both oligomerization surfaces were combined with an excess of stem IIB RNA, a single species was observed by SEC and native gel analysis (Figure 3A and 3B). The molar mass of this species is 24.5 kDa, which is consistent with a single IIB RNA with a single Rev monomer (Figure 3B and Table 1).

We next asked if these same methods could be used to define a complex in which multiple Rev monomers were bound to a single RNA. Again, we took advantage of the L60R mutation, which was shown to disrupt formation of oligomers greater than a dimer on RRE[24], but did not combine it with any other mutations, so that dimerization would be unaffected. We also used a slightly longer variant of stem IIB that was previously shown to cooperatively bind a Rev dimer (labeled IIB 42)[18]. Similar to the results described above, neither the oligomerization mutant alone nor addition of RNA to the wildtype protein resolved the large masses and heterogeneity of the complexes (Figures 2C and 2D and Table 1). However, when the oligomerization mutant and RNA were combined, a peak eluted from SEC with a molar mass of 39.8 kDa, consistent with a Rev dimer on a single RNA (Figures 3C and 3D and Table 1). Thus, using proper mutations in the oligomerization domains and binding to proper RNAs, specific, defined complexes of Rev with RNA can be obtained.

The C-terminus of Rev is disordered

The discrete nature and high solubility of the above described Rev-RNA complexes allows them to be investigated using structural techniques that previously

have been unavailable for the study of Rev, such as NMR. To begin, we performed heteronuclear HSQC experiments, which can be a powerful source of information about individual residues of a protein, as well as the feasibility of structure determination by NMR.

The ^{15}N HSQC spectrum of the 40 kDa Rev dimer on RNA displayed a paucity of peaks (Figure 4A, left). Only 40-50 amide resonances could be identified, suggesting that other resonances were in intermediate exchange and broadened beyond detection. Moreover, the peaks that were apparent in the spectra displayed chemical shifts consistent with a lack of structure of these residues[29]. The same set of peaks were observed in the spectrum of the monomeric Rev on IIB (data not shown).

It has been previously suggested that the C-terminus of Rev is disordered[21], so we wished to ask if the resonances visible in the NMR spectrum of Rev-IIB arise from the C-terminus of the protein. A truncated form of the protein ending at residue 70 showed substantially fewer peaks in the ^{15}N HSQC spectrum (Figure 4A, top), supporting the assignment of most of the peaks in the full-length Rev spectrum to the C-terminus of the protein. Thus, the resonances due to the backbone amides from the ARM and oligomerization domains appear to almost all be broadened beyond detection. Consistent with this region of the protein being involved in dynamic exchange is the notable absence of the imino peak expected from the tryptophan in the ARM of Rev, as well as the substantial broadening of resonances of the RNA in 1D spectra (data not shown). We surmise that transient interactions between Rev dimers may be responsible for the broadening of this entire region of the protein (see GB1 results below).

GB1 interacts with Rev to enhance stability

The GB1 domain greatly enhances the solubility of the Rev protein when it is fused *in cis*, allowing concentration of the protein to more than 500uM. Fusion to GB1 is a common method to enhance expression and solubility of a target protein, and is generally believed to act as other solubilizing domains by being well-folded and inert[28,30]. Interestingly, with Rev, the solubility enhancement from GB1 fusion is similar to that observed when RNA is added *in trans*, which is likely acting by binding to Rev. RNA binding to Rev may shield the large positive charge of the ARM from inappropriately aggregating with other negatively charged regions of the protein, explaining why salt counterions also solubilize the protein. This suggests that GB1 enhances the solubility of Rev not by acting as a well-folded inert domain, but rather by binding to Rev.

A prediction of this model for GB1 action is that GB1 should not behave as an inert domain in NMR experiments, as has been observed previously[28,31,32]. Rather, we expect that GB1 would bind to Rev when no RNA is present, and show reduced or no binding to Rev when RNA is present. As the ^{15}N HSQC resonances for most of the Rev protein are broadened in our NMR experiments, we expected that GB1 resonances would become broadened upon Rev binding and become visible and sharper as the domain becomes free.

Indeed, in the absence of RNA, HG-Rev shows a complete lack of resonances characteristic of the GB1 domain (Figure 4B). In fact, the only peaks that are present are the same 40-50 peaks arising from the disordered C-terminus of Rev (Figure 4A). Although the molecular mass of this HG-Rev complex is not determined, if GB1 were a freely rotating, inert domain, we would expect peaks on a ^{15}N HSQC spectrum, as has

been shown recently for mobile domains in intact ribosomes[33,34]. To further confirm that GB1 is interacting with Rev in the absence of RNA, we added RNA to the HG-Rev sample and observed the appearance of GB1 resonances in the ^{15}N HSQC spectrum (Figure 4B). Interestingly, at low temperature, most resonances showed some persistent exchange broadening, but not all resonances were broadened equally (Figure 4B, right). However, with increased temperature, the broadening of the GB1 resonances was lessened, consistent with a temperature dependence of the GB1 interaction with Rev. No remaining broadening was observed when GB1 was cleaved from Rev (Figure 4B). These results indicate that when expressed as a fusion protein, GB1 interacts with Rev and is needed for solubility, but that this interaction is lessened when RNA is present and is no longer necessary for solubility. Importantly, this suggests that GB1 is solubilizing Rev by acting as protein domain to replace RNA, and indicate that GB1 could be a rather generalizable solution for nucleic acid binding proteins that are insoluble in the absence of their binding partner.

Structural features of a Rev-RNA complex

We next wished to investigate the overall shape and domain organization of the Rev dimeric complex with RNA using small-angle X-ray scattering (SAXS). Small-angle scattering can be used to generate low resolution models of the shape and size of complexes, and in combination with other structural techniques can be a powerful method to understand the overall organization of multidomain assemblies in solution[35,36].

Scattering from the dimeric Rev-IIB 42 complex displays no indication of aggregation, as indicated by the similar shape of the scattering curves at three

concentrations (Figure 5A), as well as the linearity of the Guinier plot (Figure S2). The interatomic distance probability curve P(r) displays a wide maximum, suggestive of an elongated or lobed structure. Indeed, *ab initio* models generated using either Gasbor or DAMMIN[37,38] demonstrates that the Rev dimer in complex with IIB 42 adopts a three-lobed structure. To confirm that the low resolution Rev-RNA model generated from SAXS represents feasible orientations of the protein and RNA, the crystal structure of the Rev dimer (Daugherty et al., in preparation) and NMR structure of the bound IIB 34 RNA element[11] were aligned and fit into the density. Alignment of the ARM of the Rev dimer with the ARM in the NMR structure allows for an unambiguous placement of the Rev dimer relative to the RNA. That model, which contains the Rev dimer on IIB 34 rather than IIB 42, was fit into the SAXS density allowing for additional density that would arise from the longer IIB 42 RNA. The quality of the fit of the Rev dimer and IIB 34 RNA models validates the shape and orientation of the lobed structure generated by SAXS data.

Rev-RRE is a discrete complex

Rev-mediated RNA export in vivo relies on Rev binding to a large portion of the RRE[15], which has also been shown to be important for high affinity assembly of the functional RNP[18]. We thus wished to ask whether we could purify these larger functional Rev-RRE complexes using the same methods employed to purify defined monomeric and dimeric complexes of Rev with fragments of the RRE. The RRE alone is predominantly a single species by SEC with a calculated molar mass of 86 kDa (Figure 6A and 6C and Table 1). A minor species of dimeric RNA is also apparent, which is present in various

amounts depending on the RNA annealing conditions (Figure 6A, peak A). When an 8-fold excess of HG-Rev is incubated with the RRE, the predominant species elutes as a narrow peak from SEC with a calculated molar mass of 216 kDa (Figure 6A and 6C and Table 1). This mass corresponds to the addition of 6 HG-Rev monomers to the mass of the RRE, which is consistent with previous estimates of 6-8 Rev monomers binding to this length of RRE[15]. Native PAGE analysis also reveals that this species migrates as a distinct band, consistent with a discrete complex. Likewise, when an 8-fold excess of purified Rev is incubated with the RRE, the result is also a narrow elution peak from SEC that migrates as a distinct band of a native PAGE gels (Figure 6A, bottom and 6B). The molar mass of this peak is calculated to be 160 kDa, which also corresponds to the addition of 6 Rev monomers to the mass of the RRE (Table 1).

We also wished to confirm that these complexes were the result of cooperative, high affinity assembly of Rev on the RRE. Previous work demonstrates that oligomerization mutants disrupt high affinity, cooperative assembly on the RRE, as well as RNA export[18,23,24], so we wished to ask if these same mutants would disrupt discrete complex assembly. When an 8-fold excess of the Rev oligomerization mutant L18Q/L60R is incubated with the RRE and analyzed by MALS, the predominant peak elutes later and is broader than the wildtype protein and has a mass consistent with 3-4 Rev monomers (Figure S2A). Analysis by native PAGE reveals no distinct band, consistent with a more heterogeneous population of protein-RNA complexes (Figure S2B). These same results are observed with same oligomerization mutant with the appended GB1 tag, further confirming that interfering with Rev oligomerization disrupts formation of discrete Rev-RRE complexes (Figure S2A and S2B).

Electron microscopy of Rev-RRE complexes

An ongoing debate about Rev structure has been what the overall topology of the complex with RNA is. The inability to obtain soluble complexes of Rev with RNA has hampered this effort greatly. Though denaturing-purified Rev can be induced to form filaments[5,26,27], the cooperativity and high affinity assembly of Rev on the RRE observed biochemically[18] suggests that Rev-RRE complexes have a more discrete structure than a filament would predict. We thus hypothesized that soluble complexes formed between Rev and the RRE would display structural properties of a discrete complex, rather than a filament.

EM micrographs of the Rev-RRE complex reveal distinct particles with a consistent diameter of approximately 10 nm (Figure 7A). These dimensions are in agreement with the measured mass of the Rev-RRE particle of 160 kDa from SEC and MALS (Table 1). Interestingly, we observe lobed features within individual particles that are further highlighted when we categorize the particles into 40 class averages (Figure 7A, lower). EM of HG-Rev-RRE complexes reveals particles which are also quite consistent in size, although slightly larger than the Rev-RRE complexes (Figure 7B). This added electron density is likely due to the 6 additional GB1 domains, accounting for the calculated 56 kDa of added mass to the particles (Table 1). Lobed features are also prominent in individual particles of HG-Rev-RRE, as well as the class averages (Figure 7B, bottom), further confirming the existence of defined structural attributes of Rev-RRE complexes.

Discussion:

For nearly two decades, the structure of the Rev protein from HIV has remained elusive, due to low solubility in solution and an inability to control the oligomerization state of the protein. The problems encountered with Rev are common in the study of large molecular assemblies, especially RNP complexes. One particular challenge is the definition and purification of soluble, discrete, functional complexes. This is true of both heterooligomeric and homooligomeric complexes, although recent successes with the spliceosome and multiple viral nucleocapsid proteins illustrates that these problems can be surmounted[39-42]. We have now shown that native purification and consideration of Rev solubility requirements yield concentrations of soluble Rev protein that are 100-1000 fold higher than previously reported. Using non-denaturing purification is essential to this solubility, but interestingly, preventing aggregation of natively purified Rev appears to rely on the addition of negatively charged molecules, whether protein domains, nucleic acids or salts. This suggests that improper aggregation and precipitation of Rev is primarily mediated by charged interactions, rather than the hydrophobic interactions that are mediated by the oligomerization domains.

One especially interesting result of these studies with Rev is on the role of the protein domain, GB1, in Rev solubility. While numerous studies have utilized GB1, the mechanism by which this domain promotes the solubility of its fusion partners is mostly unexplored. In fact, most solubility tags have unknown mechanisms for preventing aggregation or misfolding[30,43]. Our data demonstrate that in this case, GB1 is acting not as an inert domain, but rather as a binding partner of Rev in the absence of its natural

binding partner, RNA. The negative charge of GB1 is presumably contributing to charge neutralization of the highly charged Rev protein, and thus serving as an 'electrostatic sink'. That GB1 serves this role with Rev raises the possibility that this may be a more generalizable approach to solubilizing proteins. Specifically, charge neutralization with domains such as GB1 may be a powerful tool in helping to solubilize charged fusion partners, such as nucleic acid binding proteins, that are prone to aggregation. Indeed, GB1 has already been shown to be indispensable for solubility of the RNA-binding domain of the decapping enzyme holoenzyme Dcp2 and the SR protein Srp20, although those studies indicate that the RNA-binding protein and GB1 do not physically interact[31,32].

Even in non-aggregating Rev or Rev-RNA samples, the protein multimerizes using the oligomerization domains that flank the ARM. Two sets of hydrophobic residues that characterize the oligomerization surfaces have been previously defined[24]. Mutation of residues in the first set (e.g. L60R) reduce oligomerization of the protein on the RRE to states higher than a dimer, while mutation of residues in the second set (e.g. L18Q) reduce oligomerization of Rev on the RRE to states higher than a monomer. Utilizing this targeted disruption of oligomerization surfaces in Rev-RNA complexes allows for the purification of discrete assemblies with defined protein and RNA stoichiometry. These Rev-RRE subcomplexes, which are homogeneous in composition and soluble to near millimolar concentrations (>10 mg/ml), serve as excellent starting points for structural studies of monomeric or dimeric Rev with RNA by NMR, SAXS or X-ray crystallography. Indeed, our preliminary NMR data have already demonstrated that the C-terminus of the protein is disordered, although further structural determination

by NMR will required optimization to reduce line broadening. Moreover, the SAXS data presented above suggest that the oligomerization domains present the RNA binding surfaces to the RNA in a manner consistent with bipartite, cooperative binding of the two Rev monomers to a single RNA. Importantly, these same methods of preparation of the protein have recently allowed determination of the crystal structure of the Rev dimer at 2.8 Å resolution (Daugherty et. al., in preparation).

In addition to the subcomplexes that we have defined, we have also been able to purify discrete complexes of the wildtype Rev protein assembled on the RRE. This complex has been shown to cooperatively assemble with high affinity, and is strongly correlated with RNA export function[18]. Moreover, proper complex formation in these assays requires intact oligomerization domains (see figure S2), further supporting the relevance of the assembled complexes. Thus, using electron microscopy, we have been able to obtain the first structural information on a functional Rev-RRE complex. Importantly, EM confirms the discrete nature of these complexes, and strongly indicates that functional Rev-RRE complexes are not filamentous, as has been proposed. In fact, as the biochemical data suggest[18,44], Rev-RRE complexes appear ordered and defined, with obvious structural features, similar to other ordered RNPs such as the ribosome and spliceosome. Although the identity of these features is difficult to assign without further biochemistry, it is intriguing to speculate that these lobed represent dimeric Rev subunits on individual stemloops of the RRE, as the recent strcture of the Rev dimer indicates that cooperative dimerization is crucial to RNA recognition (Daugherty et al., in preparation). Further structural studies will be required to supply the molecular details of how this Rev oligomerization and RNA recognition facilitates assembly of this essential viral RNP.

Methods:

Plasmids and RNAs

Proteins were expressed in E. coli using pHGB1, a plasmid derived from pHis-GB1-parallel1, which introduces an N-terminal His6 tag, followed by the B1 domain of streptococcal protein G (GB1) to increase solubility, followed by a TEV cleavage site before the coding region[18]. Full length Rev (1-116) and C-terminally truncated Rev (1-70) were cloned into pHGB1 immediately downstream of the TEV site. Mutants were generated using standard site-directed mutagenesis (Stratagene). To generate RRE RNA, a 242 nucleotide portion of the RRE from pDM128[6] was cloned between the NotI and EcoRI sites of a pBluescript-KS+ vector downstream of the T7 promoter. RNA was produced by T7 RNA polymerase in vitro run off transcription from plasmid templates linearized by EcoRI. Stem IIB RNAs were prepared using synthetic DNA templates[45]. Long and short RNAs were purified on denaturing 4% and 10% polyacrylamide/8 M urea gels respectively. RNAs were annealed by heating at 95°C for 2 min and slow cooling to room temperature in minimal buffer (0.5 mM HEPES, pH 7.0) at concentrations less than 50 uM. For high RNA concentrations, the annealed RNA was lyophilized and resuspended in reduced volume, rather than annealing at higher concentrations, to avoid duplex strand formation.

Native purification of Rev

Rev protein was expressed in *E. coli* strain BL21/DE3 from pHGB1-derived vectors as N-terminal fusions with a His_6 tag, GB1 domain and a TEV-protease cleavage

site. Cells were grown with 100 ug/ml ampicillin to $O.D._{600}$ = 0.8 at 37°C in LB medium or, for NMR experiments, M9 minimal media supplemented with trace minerals, thiamine and $^{15}NH_4Cl$ as the sole nitrogen source. Isopropyl-β-D-thiogalactopyranoside was added to 1 mM to induce expression by shaking 4-5 h at 37°C. Cell pellets were frozen in liquid nitrogen and resuspended in lysis buffer [25 mM HEPES pH 7.5, 200 mM NaCl, 100 mM Na_2SO_4, 0.1% Tween-20, 2 mM β-mercaptoethanol (β-ME) supplemented with 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor cocktail (Roche)]. The suspension was incubated on ice for 20 min with 1 mg/ml lysozyme, sonicated, and centrifuged at 14,000 rpm for 30 min to remove cell debris. RNase A (50 mg/ml) and T1 (50 U/ml) (Roche) and NaCl to 2 M were added to the cleared lysate to remove endogenous *E. coli* RNA.

Ni-NTA affinity chromatography was performed using standard procedures. Briefly, supernatant from the cell lysate was applied to Ni-NTA superflow resin (Qiagen) that had been equilibrated with buffer A+ (50 mM Tris 8.0, 2 M NaCl, 0.1% Tween-20, 2 mM β-ME, 10 mM imidazole). The resin was rinsed thoroughly with buffer A+ then buffer A (same as buffer A+ but with 250 mM NaCl and no Tween-20). A stepwise elution was performed using buffer A with increasing concentrations of imidazole. Fractions were analyzed by SDS-PAGE, pooled, and dialyzed against buffer B [40 mM Tris pH 8.0, 200 mM NaCl, 2 mM β-ME] at 4°C.

To remove the GB1 domain, TEV protease was added and incubated at room temperature for 1-2 h. For preparative reactions, either specific RNA or 100 mM Na_2SO and 400 mM $(NH_4)_2SO_4$ were added prior to TEV proteolysis to prevent Rev aggregation. The reaction was loaded on Ni-NTA resin equilibrated with buffer B containing 20 mM

imidazole to remove the His-tagged TEV protease, the free His-GB1 tag, and any uncleaved protein. Full length Rev or Rev-RNA complexes were collected in the flow through and stored at 4 degrees. Care was taken to use the protein soon after purification to prevent aggregation.

Size exclusion chromatography (SEC) and multi-angle light scattering (MALS)

Analytical size exclusion chromatography was performed using a Ettan LC system (GE Life Sciences) with a silica gel KW803 column (Shodex) equilibrated in buffer B at a flow rate of 0.35 ml/min. The system was coupled on-line to an 18-angle MALS detector (DAWN HELEOS II, Wyatt Technology) and a differential refractometer (Optilab rEX, Wyatt Technology). Molar mass determination was performed by the ASTRA 5.3.1.5 software.

For NMR and SAXS samples, preparative size exclusion chromatography was performed on full length Rev L60R in 2:1 stoichiometry with IIB 42 on an AKTApurifier system (GE Life Sciences) with a Superdex 200 10/300 GL column equilibrated in buffer B at a flow rate of 0.5 ml/min. Only the peak corresponding to 2:1 Rev-RNA complexes was collected and concentrated for analysis.

Native gel analysis

Fractions from size exclusion chromatography were supplemented with glycerol to 10% and loaded onto 6% (for RRE) or 10% polyacrylamide (37.5:1 mono:bis, 0.5x TBE) gels. Gels were run for 1-2 hours depending on RNA size and stained with ethidium bromide (for RRE) or toluidine blue.

NMR

All ^{15}N-resolved HSQC and heteronuclear NOE NMR experiments were conducted at 283K (unless otherwise noted) on a Bruker BioSpin DRX 800 MHz spectrometer equipped with a cryoprobe. All samples contained uniformly ^{15}N labeled Rev (L60R, full-length or C-terminally truncated) in complex with IIB 42. Complexes were purified by size exclusion chromatography at 2:1 protein:RNA stoichiometry and were in 40 mM Tris pH 7.0, 200 mM NaCl, 2 mM β-ME, and 10% D2O. Spectra were processed with NMRPIPE[46] and analyzed with SPARKY[47].

Small-Angle X-Ray Scattering (SAXS)

SAXS data on the dimeric Rev-RNA complex at concentrations of 2.5, 5 and 10 mg/ml were acquired at the Advanced Light Source (ALS) SIBYLS beamline (12.3.1) using a MarCCD165 detector. Exposures of 0.5 s, 5 s and 0.5 s were analyzed for radiation damage and background scattering from the buffer (40 mM Tris, 200 mM NaCl, 2 mM β-ME) was subtracted. Intensity data, I(Q), were scaled and merged using PRIMUS[48]. To ensure that no aggregation was observed, the scattering curves were calculated separately for each concentration and were nearly identical in shape (Figure 5A). Scattering data from all three concentrations were averaged together and the interatomic distance distribution, P(r), was calculated using GNOM[49] using a D_{max} value of 110 Å. Further indicating a lack of aggregation, the Guinier plot was linear for $Q \bullet R_g < 1.3$ (Figure S2) with a calculated radius of gyration (R_g) of 32 Å.

Twenty *ab initio* models were calculated using Gasbor and DAMMIN[37,38] using default parameters. For each set of twenty, an average model was calculated using DAMMAVER[50]. Individual models and both average models showed a similar overall three-lobed shape although the orientation of each lobe differed slightly in different individual models.

To fit into the model density, two molecular models were utilized. First, the recently solved crystal structure of the dimer of the Rev protein was used (Chains A and B) to model the oligomerization and ARM domains of the protein dimer (Daugherty et al., in preparation). The C-terminus, which is not in the crystal structure, was assumed to be disordered and was therefore not modeled. This model was aligned with the NMR structure of the Rev ARM with the IIB 34 RNA (PDB code 1ETF)[11] using the ARMs of each model (all atom RMSD = 0.78). The overall Rev dimer on IIB 34 was fit into the SAXS density and visualized using PyMOL (http://www.pymol.org).

Electron Microscopy

Samples were prepared for negative stain EM as previously described[51]. Briefly, 2.5 ul of SEC purified Rev-RRE or HG-Rev-RRE complexes was applied to a glow-discharged carbon-coated copper grid. Excess sample was blotted onto filter paper, washed twice with two drops of water, stained twice with two drops 0.75% (w/v) uranyl formate, and aspirated to dry. To enhance the adsorption of Rev-RRE particles on the grid, 10 ul of sample was supplemented with 1 ul of 1 mM spermidine prior to applying to the grid. This concentration of spermidine did not alter the composition of the particles as judged by native gel electrophoresis and EM (data not shown).

Images were taken on a Tecnai T12 electron microscope (FEI Company, The Netherlands) equipped with a LaB_6 filament and operating at a 120 kV acceleration voltage. Micrographs were recorded using low-dose procedures on a Gatan 4096 x 4096 UltraScan (Gatan, Inc., Pleasanton, CA) camera at 1.7 um underfocus and 52,000x magnification where 1 pixel equals 2.21 Å.

Particles were interactively selected from micrographs using Ximdisp[52]. Individual particles were cut from micrographs into 90 pixel boxes, merged into one stack of images, and renormalized using IMAGIC[53]. Particles were classified into 40 classes after 10 rounds of reference-free alignment using the program SPIDER[54].

References:

1. Cullen, B.R. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**, 203-210 (1998).
2. Cullen, B.R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**, 419-24 (2003).
3. Pollard, V.W. & Malim, M.H. The HIV-1 Rev Protein. *Annu. Rev. Microbiol.* **52**, 491-532 (1998).
4. Cook, K.S. et al. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* **19**, 1577-83 (1991).
5. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. & Singh, M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. USA* **88**, 7366-70 (1991).
6. Huang, X. et al. Minimal Rev-response element for Type 1 human immunodeficiency virus. *J. Virol.* **65**, 2131-34 (1991).
7. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res.* **20**, 6465-72 (1992).

8. Kjems, J., Brown, M., Chang, D.D. & Sharp, P.A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* **88**, 683-7 (1991).

9. Malim, M.H., McCarn, D.F., Tiley, L.S. & Cullen, B.R. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J Virol* **65**, 4248-54 (1991).

10. Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. & Cullen, B.R. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc. Natl. Acad. Sci. USA* **89**, 758-62 (1992).

11. Battiste, J.L. et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547-1551 (1996).

12. Tan, R., Chen, L., Buettner, J.A., Hudson, D. & Frankel, A.D. RNA recognition by an isolated alpha helix. *Cell* **73**, 1031-40 (1993).

13. Tan, R. & Frankel, A.D. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry* **33**, 14579-14585 (1994).

14. Kjems, J. & Sharp, P.A. The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. *J Virol* **67**, 4769-76 (1993).

15. Mann, D.A. et al. A molecular rheostat: Co-operative rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J. Mol. Biol.* **241**, 193-207 (1994).

16. Hope, T.J., McDonald, D., Huang, X.J., Low, J. & Parslow, T.G. Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. *J. Virol.* **64**, 5360-6 (1990).

17. Malim, M.H. & Cullen, B.R. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241-48 (1991).

18. Daugherty, M.D., D'Orso, I. & Frankel, A.D. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell* **31**, 824-34 (2008).

19. Wingfield, P.T. et al. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. *Biochemistry* **30**, 7527-34 (1991).

20. Daly, T.J., Rusche, J.R., Maione, T.E. & Frankel, A.D. Circular dichroism studies of the HIV-1 Rev protein and its specific RNA binding site. *Biochemistry* **29**, 9791-5 (1990).

21. Auer, M. et al. Helix-loop-helix motif in HIV-1 Rev. *Biochemistry* **33**, 2988-96 (1994).

22. Cole, J.L., Gehman, J.D., Shafer, J.A. & Kuo, L.C. Solution oligomerization of the rev protein of HIV-1: implications for function. *Biochemistry* **32**, 11769-75 (1993).

23. Edgcomb, S.P. et al. Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. *Protein Sci* (2008).

24. Jain, C. & Belasco, J.G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7**, 603-614 (2001).

25. Ihm, Y. et al. Structural model of the Rev regulatory protein from equine infectious anemia virus. *PLoS ONE* **4**, e4178 (2009).

26. Blanco, F.J., Hess, S., Pannell, L.K., Rizzo, N.W. & Tycko, R. Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J Mol Biol* **313**, 845-59 (2001).

27. Havlin, R.H., Blanco, F.J. & Tycko, R. Constraints on protein structure in HIV-1 Rev and Rev-RNA supramolecular assemblies from two-dimensional solid state nuclear magnetic resonance. *Biochemistry* **46**, 3586-93 (2007).

28. Huth, J.R. et al. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci* **6**, 2359-64 (1997).

29. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11**, 852-61 (2002).

30. Hammarstrom, M., Woestenenk, E.A., Hellgren, N., Hard, T. & Berglund, H. Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J Struct Funct Genomics* **7**, 1-14 (2006).

31. Deshmukh, M.V. et al. mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell* **29**, 324-36 (2008).

32. Hargous, Y. et al. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *Embo J* **25**, 5126-37 (2006).

33. Christodoulou, J. et al. Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc Natl Acad Sci U S A* **101**, 10949-54 (2004).

34. Mulder, F.A. et al. Conformation and dynamics of ribosomal stalk protein L12 in solution and on the ribosome. *Biochemistry* **43**, 5930-6 (2004).

35. Koch, M.H., Vachette, P. & Svergun, D.I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* **36**, 147-227 (2003).

36. Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191-285 (2007).

37. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886 (1999).

38. Svergun, D.I., Petoukhov, M.V. & Koch, M.H.J. Determination of domain structure of proteins from X-ray solution scattering *Biophys. J.* **80**, 2946-2953 (2001).

39. Albertini, A.A. et al. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360-3 (2006).

40. Green, T.J., Zhang, X., Wertz, G.W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357-60 (2006).

41. Ye, Q., Krug, R.M. & Tao, Y.J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078-82 (2006).

42. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458**, 475-80 (2009).

43. Waugh, D.S. Making the most of affinity tags. *Trends Biotechnol* **23**, 316-20 (2005).

44. Pond, S.J., Ridgeway, W.K., Robertson, R., Wang, J. & Millar, D.P. HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc Natl Acad Sci U S A* **106**, 1404-8 (2009).

45. Milligan, J.F. & Uhlenbeck, O.C. Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.* **180**, 51-62 (1989).

46. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-93 (1995).

47. Goddard, T.D. & Kneller, D.G. SPARKY 3.112, University of California, San Francisco. (2007).

48. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277-1282 (2003).

49. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495-503 (1992).

50. Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860-864 (2003).

51. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* **6**, 23-34 (2004).

52. Smith, J.M. Ximdisp--A visualization tool to aid structure determination from electron microscope images. *J Struct Biol* **125**, 223-8 (1999).

53. van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17-24 (1996).

54. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-9 (1996).

55. Charpentier, B., Stutz, F. & Rosbash, M. A dynamic in vivo view of the HIV-I Rev-RRE interaction. *J. Mol. Biol.* **266**, 950-62 (1997).

56. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V. & Cullen, B.R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254-7 (1989).

Acknowledgments

We thank John Gross and Mark Kelly for assistance with NMR experiments, Dan Southworth for assistance with MALS experiments and Kristin Krukenburg, Stephen Floor and the staff of the ALS SIBYLS beamline 12.3.1 for assistance with the SAXS experiments. Special thanks to John Gross, Geeta Narlikar and Bhargavi Jayaraman for critical reading of the manuscript. M.D.D. is a Howard Hughes Medical Institute predoctoral fellow. D.S.B is a NIGMS-IMSD fellow. This work was supported by NIH grants P50GM82250 to A.D.F and Y.C. and P0GM56531 to A.D.F.

Figure Legends:

Figure 1: Rev is highly soluble and dependent on counterions

A) Domain structure of Rev and HG-Rev used in this study. B) Predicted secondary structure of the RRE and stem IIB constructs used in this study. Stem nomenclature is from Charpentier et al.[55], and numbering corresponds to the originally described RRE structure[56]. C) TEV cleavage time course reaction of HG-Rev. Potassium phosphate pH 8.0, sodium sulfate or tRNA were added to stabilize cleaved Rev. Numbers indicate time of incubation of the reaction. P denotes any pelleted precipitate observed in the reaction after 60 minutes. Molecular masses are indicated as are positions of protein bands.

Figure 2: Rev and Rev-RNA complexes are soluble but heterogeneous

A) Size exclusion chromatography (SEC) of HG-Rev in the absence of RNA at 500 uM (top) and 200 uM (bottom) with absorbance measured at 280 nm. B) Measured multi-angle light scattering (MALS) (dashed lines, left axis) and calculated molar masses (solid line, right axis) determined from the major peak from SEC at the concentrations indicated. C) SEC of 200 uM HG-Rev in the presence of 50 uM (top), 100 uM (middle) and 200 uM (bottom) IIB 34 RNA with absorbance measured at 260 nm. For reference, the elution profile of 10 uM IIB 34 alone is displayed on the bottom panel to indicate the retention volume of unbound RNA. Arrows indicate fractions that were analyzed by native PAGE. D) Native PAGE analysis of fractions from SEC. Input IIB 34 RNA is shown in the left lane and remaining lanes are from corresponding fractions from SEC indicated in C. E) MALS (dashed lines, left axis) and calculated molar masses (solid

line, right axis) determined from the major peak from SEC. As a reference on the right, light scattering data from 50 uM IIB 34 alone is shown, along with calculated molar mass.

Figure 3: Defined monomeric and dimeric Rev-RNA complexes

A) Size exclusion chromatography (SEC) of 500 uM HG-Rev L18Q/L60R in the absence of RNA (top, measured at 280 nm), 200 uM Rev wildtype with 240 uM IIB 34 RNA (middle, measured at 260 nm) and 200 uM Rev L18Q/L60R with 240 uM IIB 34 RNA (bottom, measured at 260 nm). Arrows indicate fractions used for native PAGE analysis or the position of unbound IIB 34 RNA. B) Native PAGE analysis of input RNA and fractions from SEC (top) and MALS and calculated molar masses from SEC (bottom, dashed line and solid line respectively). C) SEC of 500 uM HG-Rev L60R in the absence of RNA (top, measured at 280 nm), 200 uM Rev wildtype with 240 uM IIB 42 RNA (middle, measured at 260 nm) and 200 uM Rev L60R with 240 uM IIB 42 RNA (bottom, measured at 260 nm). Arrows indicate fractions used for native PAGE analysis or the position of unbound IIB 42 RNA. D) Native PAGE analysis of input RNA and fractions from SEC (top) and MALS and calculated molar masses from SEC (bottom, dashed line and solid line respectively).

Figure 4: NMR analysis of Rev-RNA reveals a disordered C-terminus and GB1 interactions

A) ^{15}N HSQC spectra of full-length and C-terminally truncated Rev-RNA complexes. The indicated length Rev L60R protein was combined with IIB 42 RNA and purified as a

2:1 protein-RNA complex. B) GB1 interaction data. ^{15}N HSQC spectra of indicated protein alone or protein-RNA complexes were measured at the indicated temperature. The right panels show an expanded view of one region of the spectra to demonstrate the differences in peak broadening across the different samples.

Figure 5: Model of the Rev dimer on RNA

A) Small-angle X-ray scattering data of the Rev dimer on IIB 42 RNA at three different concentrations as indicated. The lack of aggregation of the sample can be observed by the nearly identical shape of the curves at low scattering angles, as well as the linearity of the Guinier plot (Figure S1) B) The interatomic distance probability function calculated from the average scattering curve from all three concentrations showing a broad maximum, indicative of an extended or lobed structure. C) DAMMAVER average dummy atom model of 20 independent Gasbor calculated models. Dummy atoms are shown as spheres, with a molecular surface drawn around them. D) Fitting of the crystal structure of the Rev dimer (green cartoon) and the NMR structure of the IIB 34 RNA (blue cartoon) into the SAXS model (green surface). The NMR structure only accounts for 34 of the 42 nucleotides that would be expected from the RNA that was used in the SAXS experiments. The remainder of the IIB 42 RNA would be expected to extend toward the second Rev monomer, and fill in the additional density that is currently unaccounted for the in the model. The contribution of the disordered C-termini of the Rev protein would be expected to be small.

Figure 6: Purification and analysis of Rev-RRE complexes

A) Size exclusion chromatography (SEC) of RRE alone (top), HG-Rev-RRE (middle) and Rev-RRE (bottom) at the concentrations indicated, with arrows denoting fractions used for native PAGE analysis. Fraction A is due to the presence of a small amount of dimeric RRE obtained during RNA annealing. B) Native PAGE analysis of input RRE and fractions from SEC. C) Measured multi-angle light scattering (MALS) (dashed lines, left axis) and calculated molar masses (solid line, right axis) determined from the major peak from SEC.

Figure 7: Electron microscopy of discrete Rev-RRE particles.

A) A representative negative stain field of Rev-RRE particles (top). The scale bar indicates 50 nm. Below are representative classes of particle averages from 7801 individual particles with the scale bar representing 10 nm. B) A representative negative stain field of HG-Rev-RRE particles (top) with 50 nm scale bar. Below are representative classes of particle averages from 5333 individual particles with 10 nm scale bar.

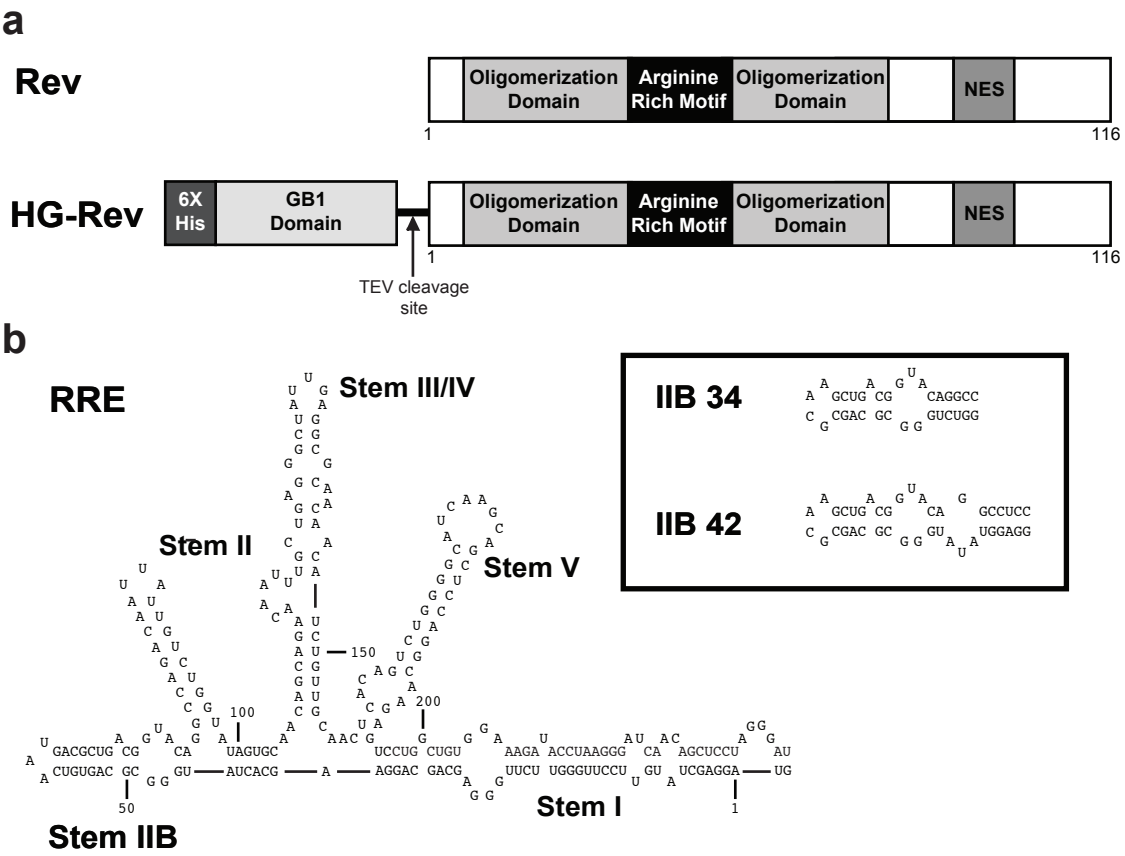
Table 1:

Complex	Predicted Mass (kDa)	Measured Mass (kDa)	# of Rev monomers
HG-Rev	21.7		
Rev	13.2		
IIB 34	11.0	11.0	
IIB 42	13.6	13.4	
RRE	90.9	86.0	
500 uM HG-Rev WT		210	9.7
200 uM HG-Rev WT		147	6.8
200 uM HG-Rev WT 200 uM IIB 34		238	10.5
200 uM HG-Rev WT 100 uM IIB 34		178, 63.6	7.7, 2.4
200 uM HG-Rev WT 200 uM IIB 34		63.5	2.4
500 uM HG-Rev L18/L60		101, 43.5	4.7, 2.0
200 uM Rev WT 240 uM IIB 34		56.9	3.5
200 uM Rev L18/L60 240 uM IIB 34		24.5	**1.0**
500 uM HG-Rev L60		98.0, 67.5	4.5, 3.1
200 uM Rev WT 110 uM IIB42		80.0, 60.9	5.0, 3.6
200 uM Rev L60 110 uM IIB 42		39.8	**2.0**
40 uM HG-Rev WT 5 uM RRE		216	**6.0**
40 uM Rev WT 5 uM RRE		161	**5.7**

Table 1: Predicted and calculated masses of Rev-RNA complexes

Indicated complexes were assembled and subjected to analytical size exclusion chromatography followed by multi-angle light scattering (MALS). Molar masses were determined from MALS and the number of Rev monomers was calculated based on

predicted masses for protein and measured masses for RNA and always assume only one RNA molecule per complex.


a
Rev
Oligomerization Domain
Arginine Rich Motif
Oligomerization Domain
NES
1
116
HG-Rev
6X His
GB1 Domain
Oligomerization Domain
Arginine Rich Motif
Oligomerization Domain
NES
1
116
TEV cleavage site
b
RRE
Stem III/IV
Stem II
Stem V
Stem I
Stem IIB
IIB 34
IIB 42

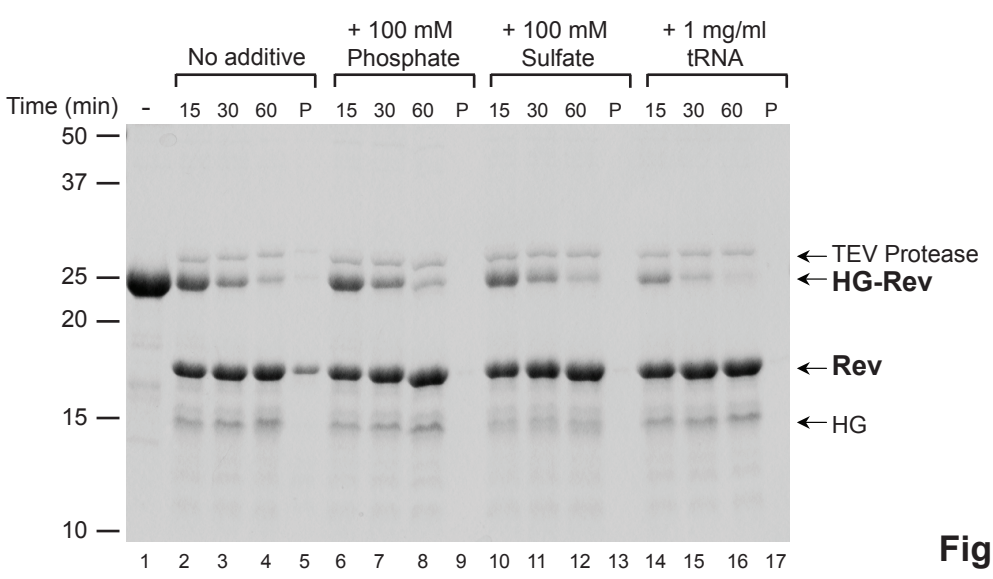

c
No additive
+ 100 mM Phosphate
+ 100 mM Sulfate
+ 1 mg/ml tRNA
Time (min)
- 15 30 60 P 15 30 60 P 15 30 60 P 15 30 60 P
50
37
25
20
15
10
TEV Protease
HG-Rev
Rev
HG
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Figure 1

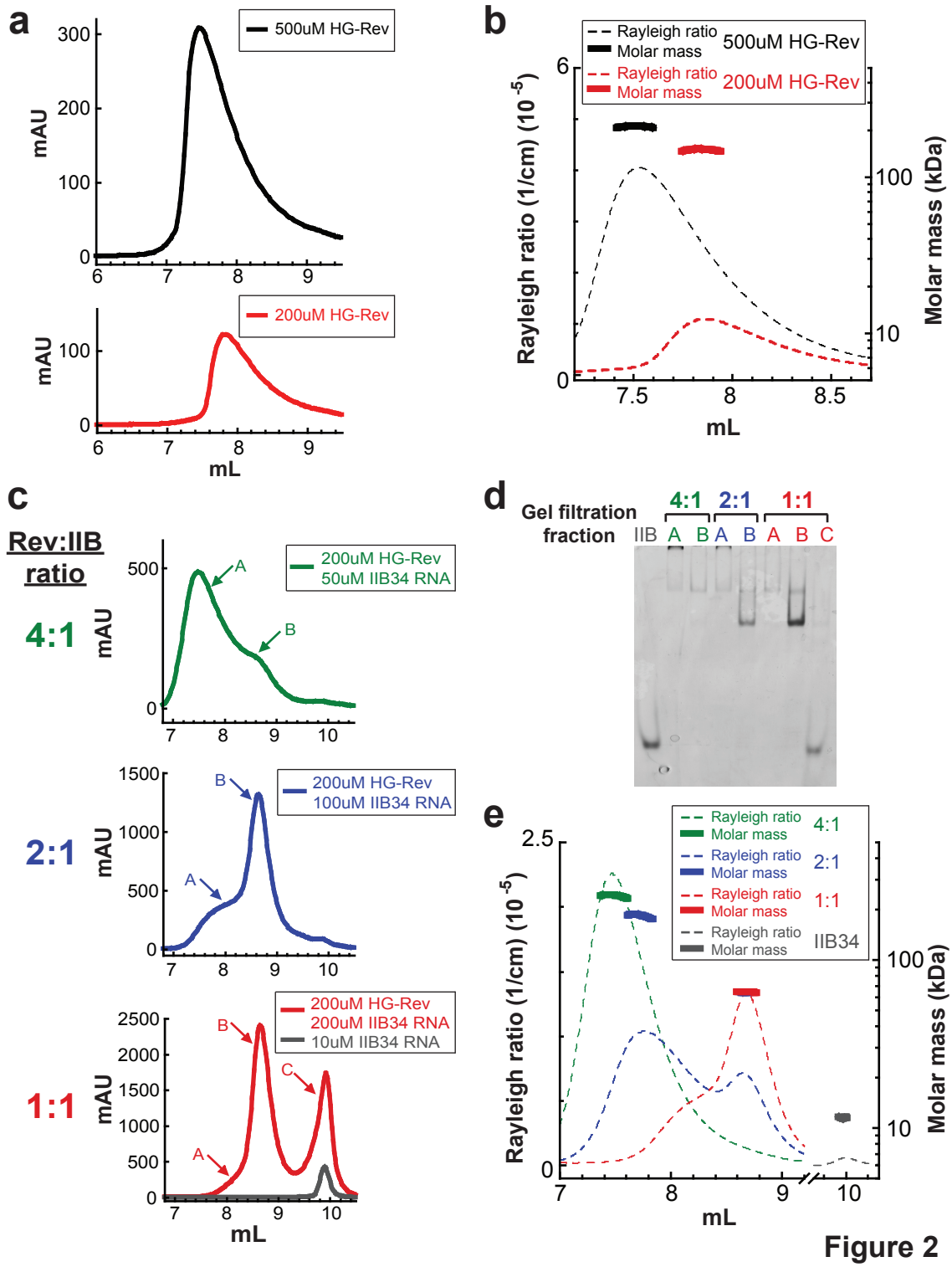

a
500uM HG-Rev
200uM HG-Rev
mAU
mL
b
Rayleigh ratio
Molar mass
500uM HG-Rev
200uM HG-Rev
Rayleigh ratio (1/cm) (10 -5)
Molar mass (kDa)
mL
c
Rev:IIB ratio
4:1
200uM HG-Rev
50uM IIB34 RNA
2:1
200uM HG-Rev
100uM IIB34 RNA
1:1
200uM HG-Rev
200uM IIB34 RNA
10uM IIB34 RNA
mAU
mL
d
Gel filtration fraction
4:1
2:1
1:1
IIB A B A B A B C
e
Rayleigh ratio
Molar mass
4:1
2:1
1:1
IIB34
Rayleigh ratio (1/cm) (10 -5)
Molar mass (kDa)
mL

Figure 2

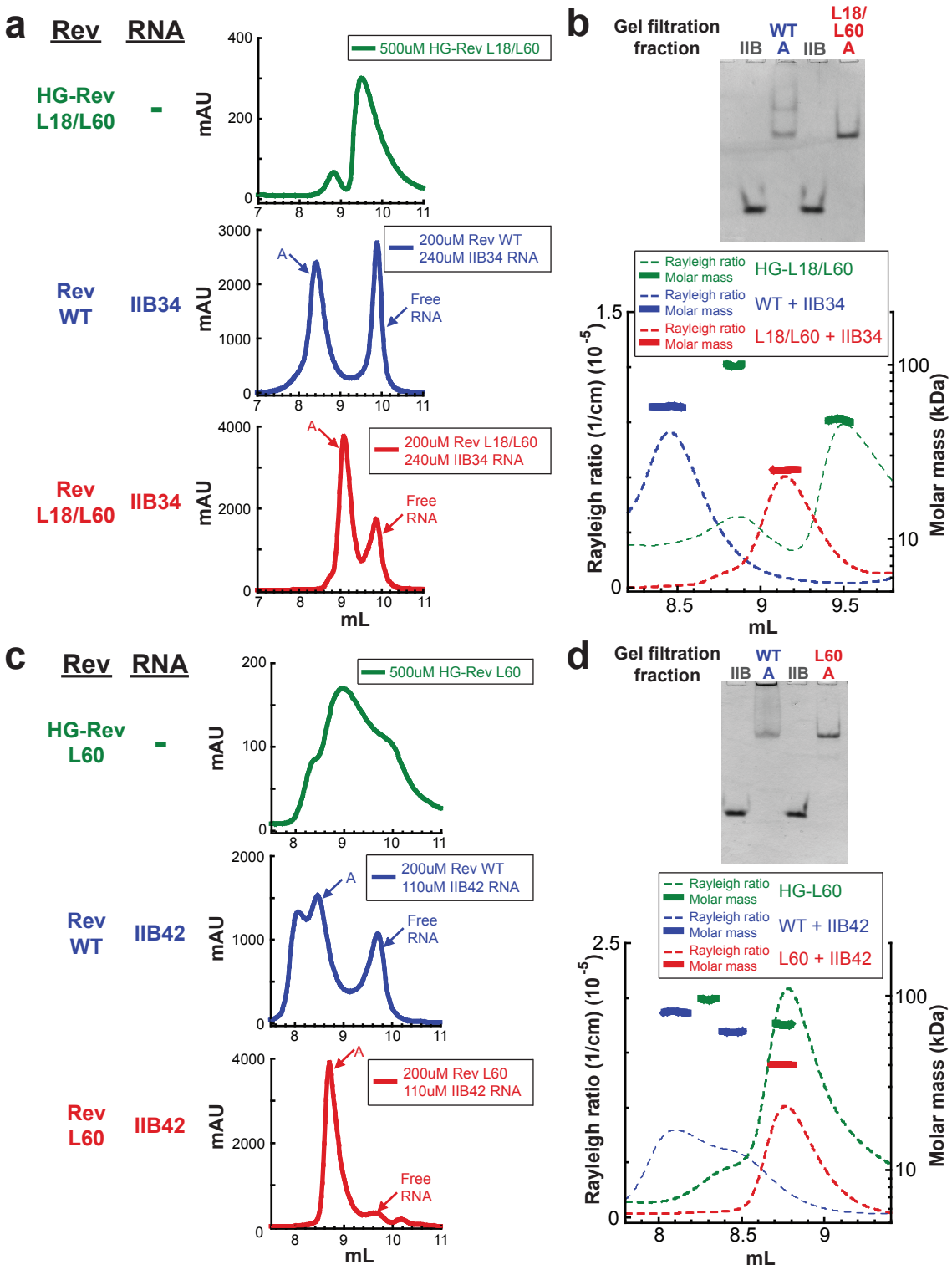

a
Rev
RNA
HG-Rev L18/L60
-
500uM HG-Rev L18/L60
mAU
Rev WT
IIB34
200uM Rev WT
240uM IIB34 RNA
A
Free RNA
Rev L18/L60
IIB34
200uM Rev L18/L60
240uM IIB34 RNA
mL
b
Gel filtration fraction
WT
L18/L60
IIB
A
Rayleigh ratio
Molar mass
HG-L18/L60
WT + IIB34
L18/L60 + IIB34
Rayleigh ratio (1/cm) (10-5)
Molar mass (kDa)
c
HG-Rev L60
500uM HG-Rev L60
Rev WT
IIB42
200uM Rev WT
110uM IIB42 RNA
Rev L60
IIB42
200uM Rev L60
110uM IIB42 RNA
d
L60
HG-L60
WT + IIB42
L60 + IIB42

Figure 3

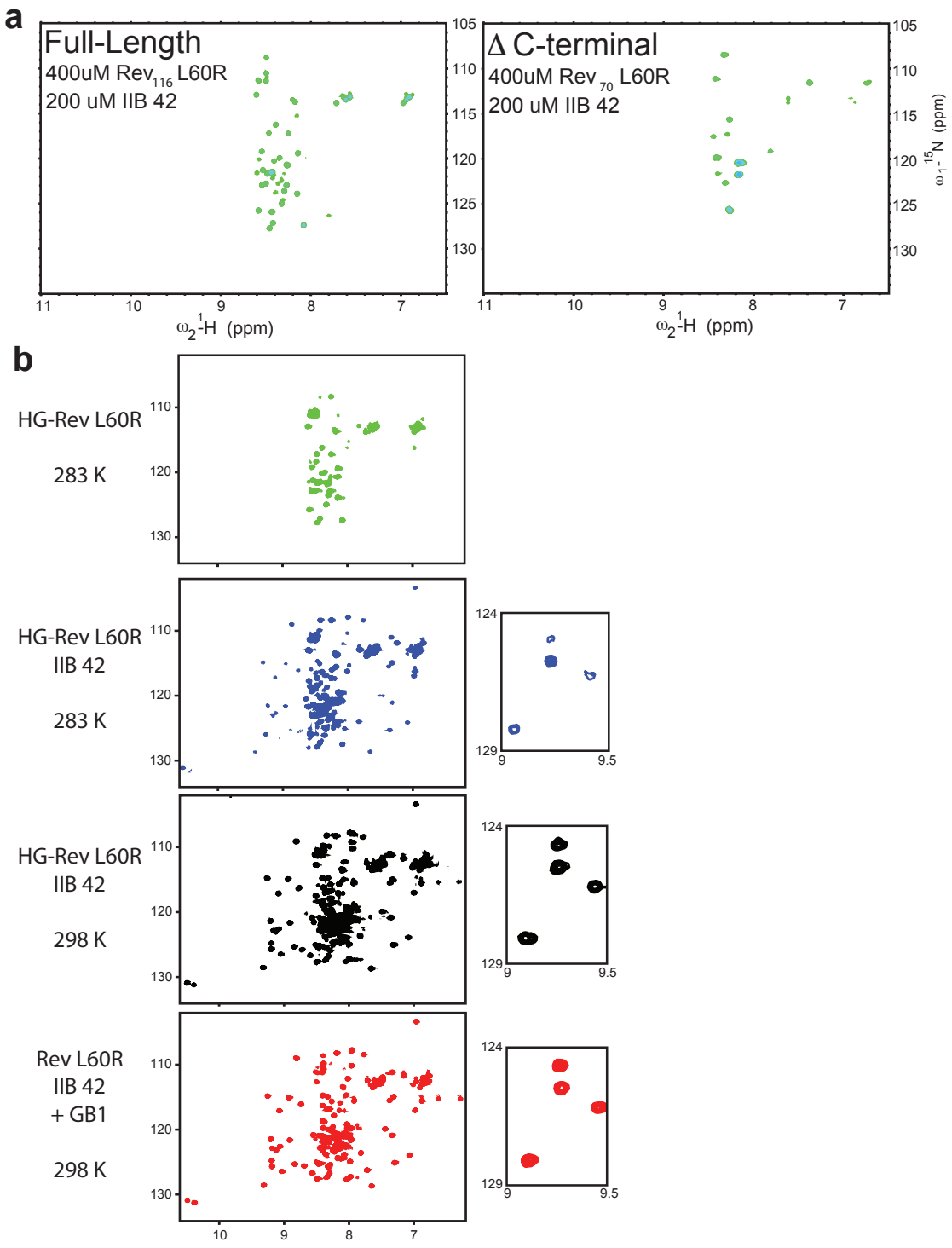

a
Full-Length
400uM Rev116 L60R
200 uM IIB 42
Δ C-terminal
400uM Rev 70 L60R
200 uM IIB 42
ω2 1H (ppm)
ω1 15N (ppm)
b
HG-Rev L60R
283 K
HG-Rev L60R
IIB 42
283 K
HG-Rev L60R
IIB 42
298 K
Rev L60R
IIB 42
+ GB1
298 K

Figure 4

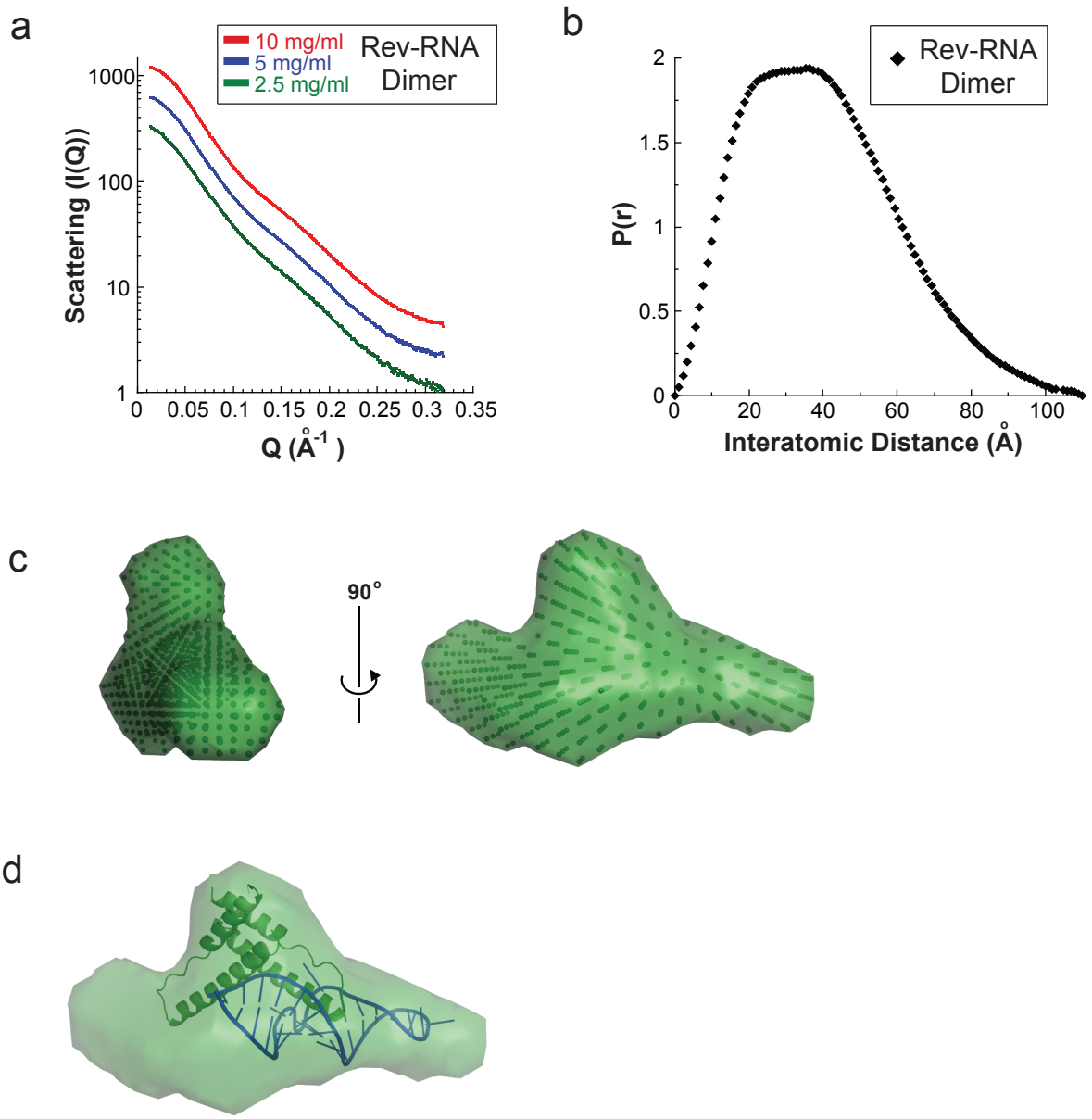

a
10 mg/ml
5 mg/ml
2.5 mg/ml
Rev-RNA Dimer
Scattering (I(Q))
1000
100
10
1
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35
Q (Å-1)
b
Rev-RNA Dimer
P(r)
2
1.5
1
0.5
0
0 20 40 60 80 100
Interatomic Distance (Å)
c
90°
d

Figure 5

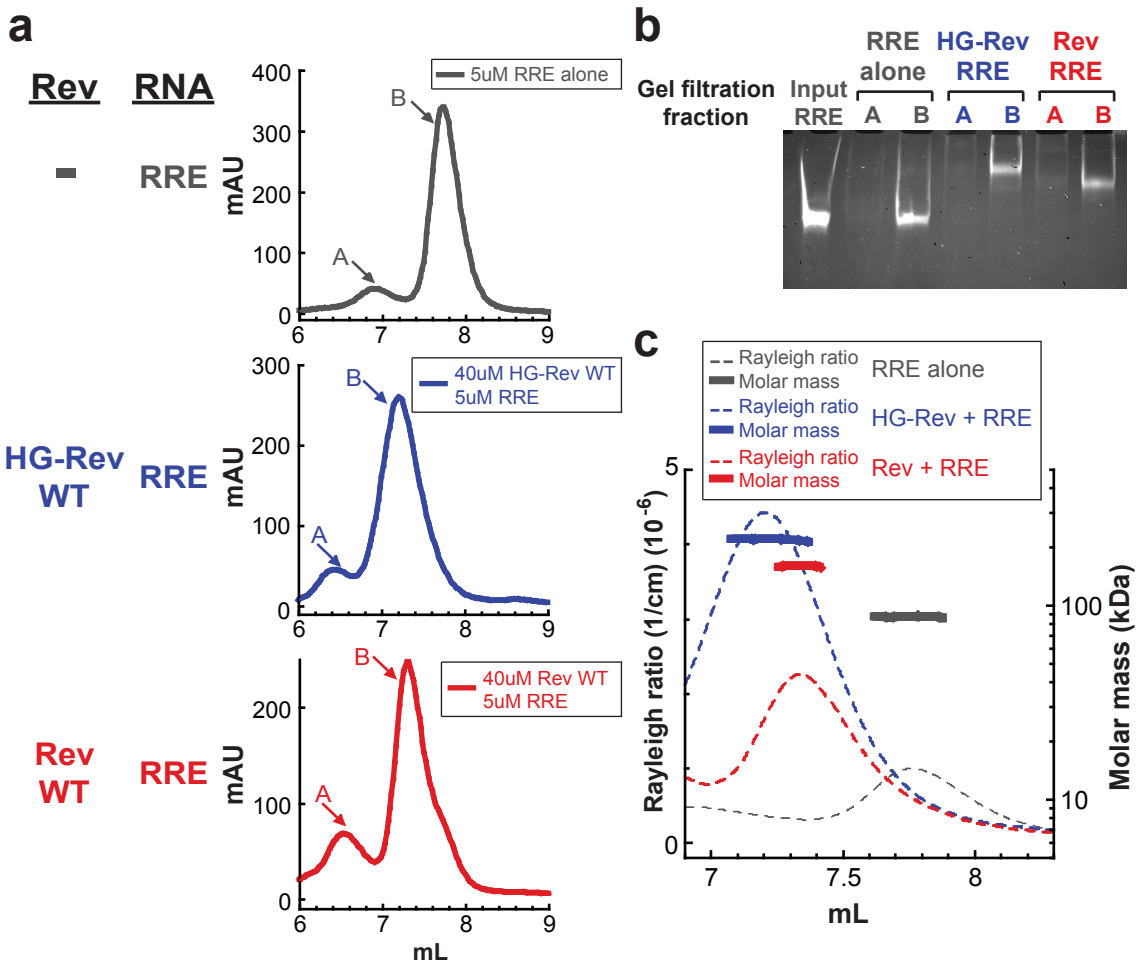

a
Rev
RNA
RRE
5uM RRE alone
HG-Rev WT
40uM HG-Rev WT
5uM RRE
Rev WT
40uM Rev WT
5uM RRE
mAU
mL
A
B
b
Gel filtration fraction
Input RRE
RRE alone
HG-Rev RRE
Rev RRE
c
Rayleigh ratio
Molar mass
RRE alone
HG-Rev + RRE
Rev + RRE
Rayleigh ratio (1/cm) (10-6)
Molar mass (kDa)
mL

Figure 6

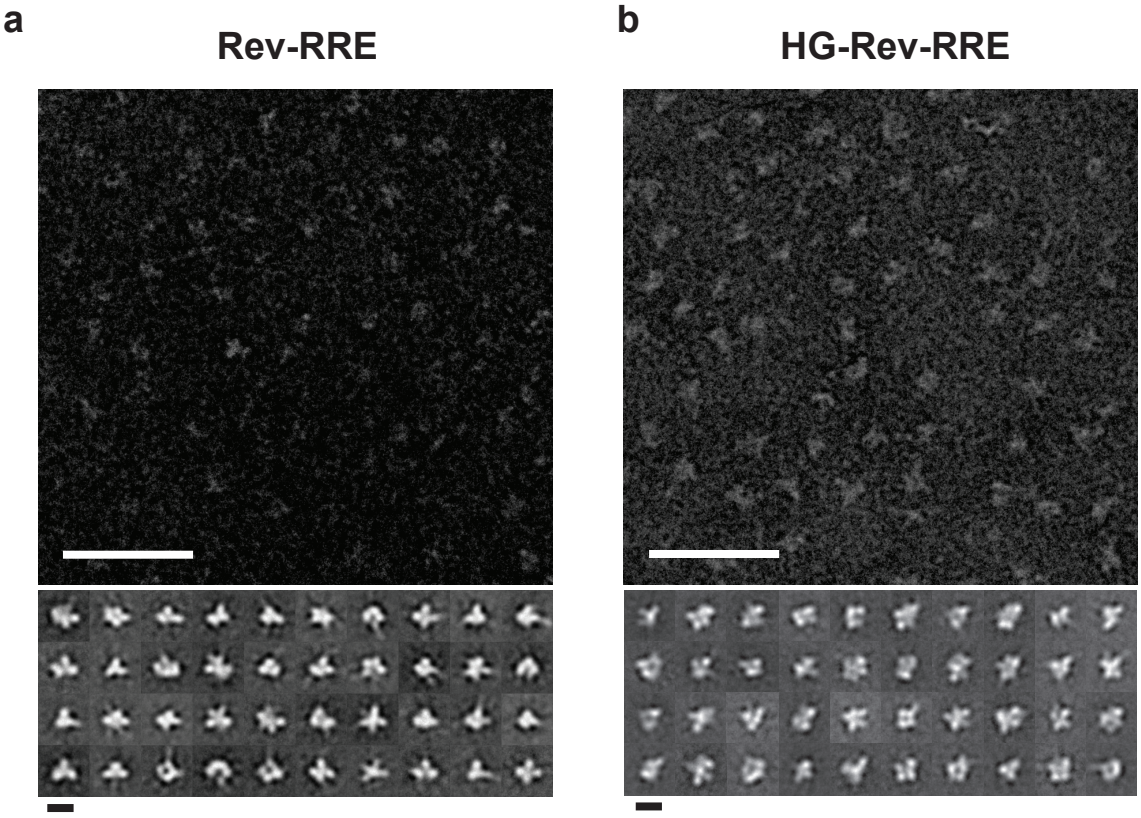


Figure 7

Supplemental Data

Assembly and structural insight into discrete HIV Rev-RNA complexes

Matthew D. Daugherty, David S. Booth, Yifan Cheng, and Alan D. Frankel*

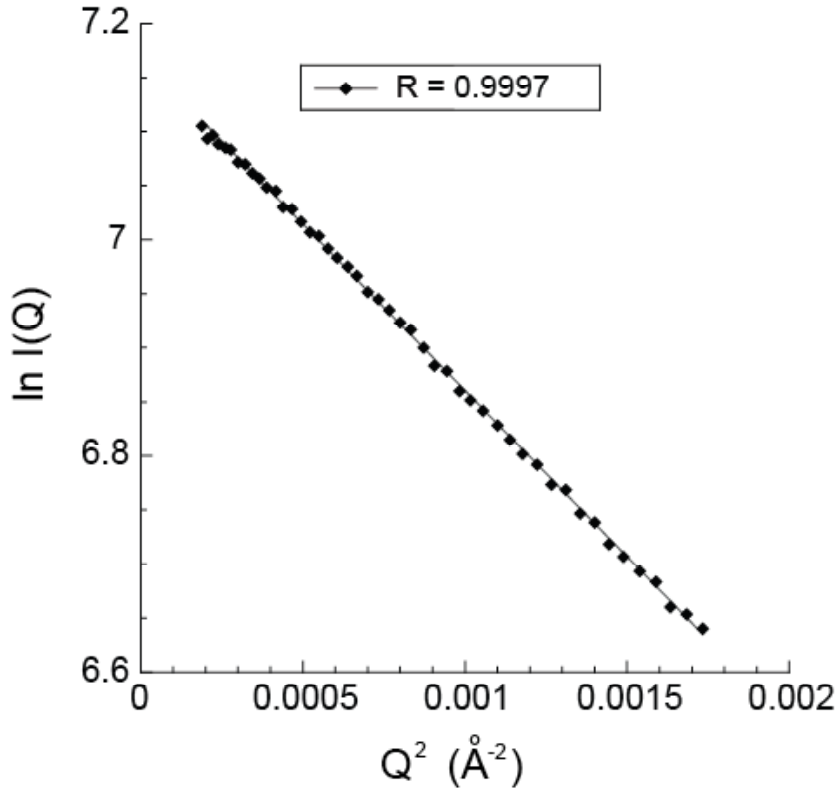


Figure S1: Guinier plot showing a lack of aggregation in SAXS samples

Initial points from the average scattering curve of the Rev dimer complex with IIB 42 show linearity in a Guinier plot, indicating monodispersity of the sample. A linear fit, with calculated R value, is shown.

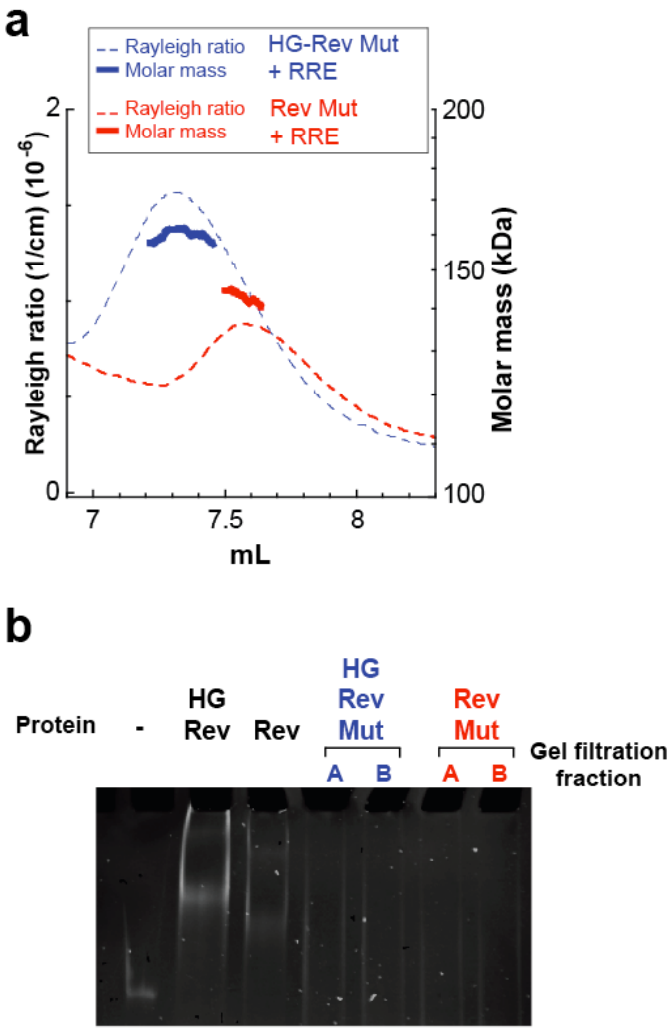


Figure S2: Discrete Rev-RRE assembly requires proper oligomerization

A) Measured multi-angle light scattering (MALS) (dashed lines, left axis) and calculated molar masses (solid line, right axis) determined from the major peak from size exclusion chromatography (SEC) of the Rev oligomerization mutant L18Q/L60R on the RRE. Complexes of 40 uM Rev L18Q/L60R or HG-Rev L18Q/L60R were assembled with 5uM RRE (red and blue, respectively). B) Native gel analysis of two fractions of the above SEC experiments, showing a lack of discrete complexes on a native gel. As a reference, a fraction from SEC with wildtype HG-Rev and Rev complexes with the RRE (as in Figure 6) are shown.

Chapter 4

Structure of the HIV-1 Rev dimer at 2.8 Å resolution

Matthew D. Daugherty[1], Bella Liu[2], and Alan D. Frankel[2,*]

[1]Chemistry and Chemical Biology Graduate Program

[2]Department of Biochemistry and Biophysics

University of California, San Francisco

San Francisco, CA 94158

[*]Contact: Alan Frankel
Department of Biochemistry and Biophysics
UCSF
600 16th St.
San Francisco, CA 94158
415-476-9994
frankel@cgl.ucsf.edu

Summary:

Replication of HIV requires the nuclear export of unspliced viral RNAs for the translation of structural proteins and viral packaging. Crucial to this process is oligomeric binding of the regulatory protein Rev to the Rev Response Element (RRE) within the viral RNA, and subsequent export through the nuclear pore complex. Despite extensive biochemical characterization of the Rev protein and its importance in viral replication, the structure of the Rev protein has remained elusive for nearly two decades. Here we present the crystal structure of the Rev protein dimer at 2.8 Å resolution. The structure reveals that the folding and stabilization of the Rev monomer is accomplished through extensive well-conserved hydrophobic and polar contacts between the helical oligomerization domains. Moreover, dimerization involves the burial of >700 $Å^2$ of surface area on each monomer and determines the orientation of the RNA-binding surfaces of Rev relative to one another. The structure of the Rev dimer reveals the structural basis of Rev oligomerization and cooperative binding to RNA, as well as providing the first glimpse of this essential HIV protein.

Complex retroviruses, such as HIV, encode the essential regulatory protein Rev to facilitate nuclear export of the viral RNA genome. Rev binds to and oligomerizes on the Rev Response Element (RRE) RNA found in the introns of partially and fully unspliced viral mRNAs, and directs their transport to the cytoplasm before splicing is completed[1-3]. The Rev-RRE complex then binds to the host cell factor, Crm1, for export through the nuclear pore complex and release of unspliced viral RNAs into the cytoplasm. These unspliced viral RNAs encode most of the structural and non-structural viral proteins, as well as the viral genome, and thus Rev mediated RNA export is essential for HIV replication.

Assembly of the Rev-RRE complex proceeds by formation of a homooligomer of the Rev protein on the RRE. Oligomerization has been shown to be essential for RNA export[4,5] as well as cooperative assembly of a high affinity complex with the RRE[6]. Circular dichroism studies have indicated that the oligomerization domains stabilize the predicted helix-turn-helix structure of Rev[7] (see Supplementary Figure 1) and extensive biochemical mapping has led to a model for monomer stability and oligomerization[8]. However further explanation for how Rev cooperatively oligomerizes on RNA has been hampered by the lack of a molecular structure of Rev.

Oligomerization mediated multipartite binding to the RRE is accomplished through interactions between the Arginine Rich Motif (ARM) of the Rev protein and multiple sites within the RRE. Initial studies defined one site within the RRE, known as stem IIB, that is required for in vivo function as well specific binding to the RRE in vitro[9-15]. An extended stem IIB RNA can cooperatively bind a dimer of the Rev protein if an additional helical disruption in RNA structure is properly oriented relative to the IIB

site [6,16], suggesting that oligomerization positions Rev monomers for RNA binding. Interestingly, only one additional specific site has been located within the RRE in stem IA, and this site binds a different surface of the Rev ARM than stem IIB[6]. Taken together, these data suggest that cooperative binding to RNA is likely determined by the organization of Rev monomers in the oligomer and may orient the Rev ARMs differently to different binding sites.

Despite this wealth of functional data, the molecular structure of the Rev protein has remained elusive for nearly two decades due to the low solubility of the protein at concentrations higher than 1-5 uM[10,17]. We have recently reported methods to increase the solubility of the Rev protein, as well as specifically control the oligomerization state of the protein in complexes with RNA (Daugherty et al., in preparation). Following removal of the disordered C-terminus and using established mutations at positions 12 and 60 in the oligomerization domains[8], we were able to assemble and purify the intact dimer of the Rev protein, termed $Rev_{70\text{-}Dimer}$ (see Supplementary Figures 1-2).

The $Rev_{70\text{-}Dimer}$ structure was solved to 2.8 Å using multi-wavelength anomalous diffraction (MAD) with sodium selenate soaked crystals. The asymmetric unit contains two copies of the Rev dimer (Figure 1A). The four copies of the Rev monomer adopt the predicted antiparallel helix-turn-helix motif, with the oligomerization domains forming the core of the protein monomer. All copies of Rev in the asymmetric unit superimpose well (root mean squared deviation (r.m.s.d.) <1 Å for all pairwise alignments) (Figure 1B). In all monomers, residues 10-63 are well ordered, with residues 9-69 being ordered in at least one monomer. The two Rev dimers also superimpose well, with an r.m.s.d. of 0.6 Å between them (Figure 1C). The primary difference between the two dimers is the

angle at which the two monomers cross each other, which is 95 degrees in one dimer and 105 degrees in the other.

Folding and stability of the Rev monomer is mediated by hydrophobic and polar residues that contact one oligomerization domain from the other. The four branched hydrophobic residues (Ile 19, Leu 22, Ile 52 and Ile 59) pack well against each other (Figure 2A) and have previously been shown to be important for monomer stability[8]. An extensive hydrogen bonding network is also observed, with the sidechain carbonyl of Asn 26 hydrogen bonding to the guanidinium group of Arg 48 and the sidechain amino group of Asn 26 hydrogen bonded to the carbonyl of the Gln 49 sidechain (Figure 2A). The result of sandwiching Asn 26 between Arg 48 and Gln 49 assures that the end of helix 1 is appropriately placed near the end of the ARM on helix 2. In three of the four monomers, located one helical turn away is another hydrogen bonding interaction between the hydroxyl group on Tyr 23 and the side chain of His 53. The residues involved in these hydrophobic and polar interactions are highly conserved among all HIV-1 isolates (Figure 2B and Supplementary Figure 1), indicating the importance of placement and register of the oligomerization domains relative to each other.

Packing of the Rev dimer structure is accomplished primarily by hydrophobic residues that lie at the interface between two monomers. An extensive interface buries over 700 $Å^2$ of each monomer, totally more than 15% of the total surface area of each monomer (Figure 3A). Essential to dimerization are Leu 18 and Ile 55, which form symmetric contacts between monomers and Phe 21, which packs into the groove between the helices of the opposing monomer (Figure 3B). Notably, residues that are important for monomer structure, especially Leu 22 and Ile 59, are also important for dimer

formation. These residues, which would be partially solvent exposed in the monomer, become buried in the dimer and contribute the complete dimerization interface (Figure 3C and 3D). In this way, dimerization of the protein stabilizes each monomer, providing a structural explanation for the cooperativity of dimeric binding to RNA.

The orientation of the RNA-binding ARMs that results from dimerization provides additional structural evidence for cooperative RNA binding. An electrostatic potential surface representation of the Rev dimer reveals the positive overall charge, primarily due to ARMs of the protein (Figure 4A). The angle and position of these positively charged RNA binding surfaces is determined by the oligomerization surfaces, and must be properly oriented for cooperative RNA binding. In this way, the two ARMs are able to reach out from the body of the protein and grasp the RNA, in the same manner the arms of the human body are positioned relative to each other to grip objects. By modeling in the known structure of the IIB element bound to the Rev ARM, we are able to infer how the next RNA site in the RRE might be bound. As can be seen in Figure 4B, the end of this short IIB element projects towards the next ARM. Although the structure of the longer RNA that cooperatively binds the Rev dimer is unknown[6], we predict that the additional 8 nucleotide that make up that RNA will be sufficient to contact the second Rev ARM as it is configured in the crystal structure. Even more interesting is the face of the second Rev ARM that is presented to this modeled RNA. The binding mode employed by the first ARM utilizes Asn 40 and the surface of the ARM that faces the other monomer, or the 'inner face' of the ARM[18]. However, the end of the IIB RNA is projecting toward the 'outer face' of the second Rev ARM (Figure 4B), consistent with data suggesting that multiple binding modes are important for Rev-RRE assembly[6].

Modeling the C-terminus of the protein allows us position the binding site for the Crm1 export receptor. Though the residues in this region are disordered, they originate at the end of helix 2 in the structure, suggesting that they would project away from the RNA binding surfaces of the Rev dimer (Figure 4C). In this way, Rev oligomerization and binding to RNA is a modular process that spatially separates Crm1 recognition, similar to the way that some transcription factors dimerize on DNA while their activation domains are spatially removed from dimerization and DNA binding. The full molecular details of Rev binding to Crm1 will be revealed when a cocrystal structure is solved, a process that will be greatly facilitated by the recent crystal structures of Crm1 in complex with other cargoes[19,20].

The crystal structure of the Rev dimer serves as the end of a two decade pursuit of the Rev structure. Previous attempts to propose a structural model for Rev turned to molecular genetic screens that were critical in defining the oligomerization domains and providing an initial molecular model[8]. Our structure provides the physical detail to synthesize decades of such functional and biochemical information. The picture that emerges is one in which the Rev oligomerization domains form a hydrophobic core stabilized by intra- and intermolecular contacts that construct a V-shaped topology for the dimer. This quaternary fold is ideally suited to present the ARMs for cooperative RNA binding to individual sites positioned at defined points in the RRE scaffold. Thus, RNA sites within the RRE should have specificity at the level of not only sequence recognition but also spatial recognition. This structure thus also serves as the beginning for understanding how the entire Rev-RRE ribonucleoprotein forms to export the viral RNA. We are encouraged by recent EM data that indicates that the Rev-RRE complex is

structurally discrete and has intriguing lobes that may represent individual Rev dimers assembled on the RRE (Daugherty et al., in preparation). We look forward to future structures of Rev assembled with the RRE and host export factors to more fully understand the molecular basis of Rev mediated export.

Methods:

Protein expression and purification

To eliminate disordered regions of Rev, the C-terminal 46 residues were removed, which are experimentally and computationally predicted to be disordered[7] (Daugherty et al., in preparation). In order to prevent oligomerization beyond a dimer, two mutations in the oligomerization domains, L12S and L60R, were introduced guided by the seminal work of Jain and Belasco[8]. The resulting $Rev_{70\text{-}Dimer}$ protein was expressed in *E. coli* strain BL21/DE3 from pHGB1-derived vectors as an N-terminal fusion with a His_6 tag, GB1 domain and a TEV-protease cleavage site as have been previously described[6]. Cells were grown with 100 ug/ml ampicillin to O.D.600 = 0.8 at 37°C in "SBMX", a phosphate-based minimal medium supplemented with minerals and vitamins. The full description of the medium can be found in Supplementary Materials and Methods. Isopropyl-b-D-thiogalactopyranoside was added to 1 mM to induce expression by shaking 4-5 h at 37°C. Cell pellets were frozen in liquid nitrogen and resuspended in lysis buffer [25 mM HEPES pH 7.5, 200 mM NaCl, 100 mM Na_2SO_4, 0.1% Tween-20, 2 mM β-mercaptoethanol (β-ME) supplemented with 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor cocktail (Roche)]. The suspension was incubated on ice for 20 min with 1 mg/ml lysozyme, sonicated, and centrifuged at 14,000 rpm for 30 min to remove cell debris. RNase A (50 mg/ml) and T1 (50 U/ml) (Roche) and NaCl to 2 M were added to the cleared lysate to remove endogenous *E. coli* RNA.

Ni-NTA affinity chromatography was performed using standard procedures. Briefly, supernatant from the cell lysate was applied to Ni-NTA superflow resin (Qiagen)

that had been equilibrated with buffer A+SO_4 (50 mM Tris 8.0, 2 M NaCl, 100 mM Na_2SO_4, 0.1% Tween-20, 2 mM β-ME, 10 mM imidazole). The resin was rinsed thoroughly with buffer A+SO_4 then buffer A-SO_4 (same as buffer A+ but with 250 mM NaCl and no Tween-20). A stepwise elution was performed using buffer A-SO_4 with increasing concentrations of imidazole. Fractions were analyzed by SDS-PAGE, pooled, and dialyzed against buffer B [40 mM Tris pH 8.0, 200 mM NaCl, 100 mM Na_2SO_4, 400 mM $(NH_4)_2SO_4$, 2 mM β-ME] at 4°C.

To remove the GB1 domain, TEV protease was added and incubated at room temperature for 1-2 h. The reaction was loaded on Ni-NTA resin equilibrated with buffer B containing 20 mM imidazole to remove the His-tagged TEV protease, the free His-GB1 tag, and any uncleaved protein. $Rev_{70\text{-}Dimer}$ was collected in the flow through and concentrated. During concentration, the protein was diluted with crystallization buffer [40 mM Tris pH 8.0, 200 mM NaCl, 100 mM Na_2SO_4, 400 mM $(NH_4)_2SO_4$, 2 mM DTT] to remove imidazole and β-ME, and brought to a final concentration of 14 mg/ml.

Crystallization

Crystals of the Rev protein were obtained after 2-3 days by vapour diffusion by mixing 1ul protein at 10-14 mg/ml in crystallization buffer and 1ul reservoir solution containing 100 mM Tris pH 8.0, 50-100 mM NaCl, 1.45-1.55 M $(NH_4)_2SO_4$, and 3% PEG1000. After 10-14 days, a cryoprotectant solution containing 1.8 M sodium malonate, 100 mM Tris pH 8.0, 500 mM NaBr and 250 mM Na_2SO_4 was added incrementally to a total volume of 5 ul. Following slow addition of cryoprotectant, crystals were transferred to a drop containing only cryoprotectant, allowed to equilibrate

for 15-60 minutes and flash frozen in liquid nitrogen. Crystals used for MAD phasing were treated identically utilizing a cryoprotectant with Na_2SO_4 replacing Na_2SeO_4.

Data collection and structure determination

Data sets were collected at the Advanced Light Source beamline 8.3.1 on crystals at 100K. All crystals were cryoprotected with sodium malonate containing NaBr to reduce the c-axis of the unit cell from as long as 330 Å to 81 Å and increase the symmetry of the crystals from $P6_4$ to $P6_422$ (Daugherty and Frankel, unpublished). Cryoprotected in this way, crystals were of space group $P6_422$ (a,b = 115.8 Å, c= 81.2 Å). Native data were collected at 1.1159 Å, two-wavelength MAD data was collected on SeO_4 soaked crystals at 0.9792 Å (edge) and 0.9715 Å (remote) and SAD data was collected at 0.9202 Å (edge) to locate Br atoms. Anomalous data sets were collected using an inverse beam collection strategy to minimize differences between Friedel pairs. Diffraction data were processed and scaled with the HKL-2000 package[21]. Attempts to phase with the soaked Br atoms were unsuccessful, including using 3-wavelenth MAD data collection, likely due to the large number of Br atoms in the asymmetric unit as was determined after the structure was solved (see below). To phase the structure, Na_2SO_4 in the cryoprotectant was replaced with Na_2SeO_4 and two wavelength MAD data was collected at 0.9792 Å (edge) and 0.9715 Å (remote). Two Se atoms were located using SOLVE and initial models were built with RESOLVE[22].

Initially, loose NCS restraints were applied using REFMAC5[23] on the 4 monomers in the asymmetric unit. Subsequent iterative model building and refinement were carried out using COOT[24] and PHENIX[25] without NCS restraints. The refined

structure from the experimentally phased SeO_4 data was used as a molecular replacement model for the native data set using PHASER[26]. Ten Br atoms were located with PHASER using molecular replacement and SAD data from the Br edge. Excellent electron density can be seen for residues 10-63 in all four monomers, with variable ends of the electron density for each chain. Refinement statistics can be found in Supplementary Table 1. Solvent accessible areas were calculated using AREAIMOL in the CCP4 suite[27], and electrostatic potentials were calculated using Delphi[28]. All figures were generated using PyMOL[29].

A structural model of the Rev dimer with stem IIB was generated using the known stem IIB-ARM structure[18] (PDB code 1ETF) solved using NMR. The Rev ARM in the NMR structure was aligned with the ARM of Chain A (r.m.s.d. = 1.2 Å), unambiguously placing the RNA relative to the Rev dimer.

References:

1. Cullen, B.R. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* **249**, 203-210 (1998).
2. Cullen, B.R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**, 419-24 (2003).
3. Pollard, V.W. & Malim, M.H. The HIV-1 Rev Protein. *Annu. Rev. Microbiol.* **52**, 491-532 (1998).
4. Hope, T.J., McDonald, D., Huang, X.J., Low, J. & Parslow, T.G. Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. *J. Virol.* **64**, 5360-6 (1990).
5. Malim, M.H. & Cullen, B.R. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241-48 (1991).
6. Daugherty, M.D., D'Orso, I. & Frankel, A.D. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell* **31**, 824-34 (2008).
7. Auer, M. et al. Helix-loop-helix motif in HIV-1 Rev. *Biochemistry* **33**, 2988-96 (1994).
8. Jain, C. & Belasco, J.G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7**, 603-614 (2001).

9. Cook, K.S. et al. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res*. **19**, 1577-83 (1991).

10. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. & Singh, M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. USA* **88**, 7366-70 (1991).

11. Huang, X. et al. Minimal Rev-response element for Type 1 human immunodeficiency virus. *J. Virol*. **65**, 2131-34 (1991).

12. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res*. **20**, 6465-72 (1992).

13. Kjems, J., Brown, M., Chang, D.D. & Sharp, P.A. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc. Natl. Acad. Sci. USA* **88**, 683-7 (1991).

14. Malim, M.H., McCarn, D.F., Tiley, L.S. & Cullen, B.R. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J Virol* **65**, 4248-54 (1991).

15. Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. & Cullen, B.R. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc. Natl. Acad. Sci. USA* **89**, 758-62 (1992).

16. Zemmel, R.W., Kelley, A.C., Karn, J. & Butler, P.J. Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. *J. Mol. Biol.* **258**, 763-77 (1996).

17. Wingfield, P.T. et al. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. *Biochemistry* **30**, 7527-34 (1991).

18. Battiste, J.L. et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547-1551 (1996).

19. Dong, X. et al. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* **458**, 1136-41 (2009).

20. Monecke, T. et al. Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* **324**, 1087-91 (2009).

21. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 307-326 (1997).

22. Terwilliger, T.C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**, 849-61 (1999).

23. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).

24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

25. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).

26. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

27. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).

28. Rocchia, W., Alexov, E. & Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **105**, 6507-6514 (2001).

29. Delano, W.L. *The PyMOL molecular graphics system. v.0.99*. (Delano Scientific, 2006).

Acknowledgements:

We are extremely grateful for crystallography and structure determination assistance from Zach Newby, Pascal Egea, Janet Finer-Moore, Robert Stroud, Miles Pufall, JJ Miranda, Dong Young Kim, John Gross, Lesa Beamer, James Holton and George Meigs. Special thanks to David Booth, John Gross and Geeta Narlikar for helpful discussions and critical reading of the manuscript. M.D.D. is a Howard Hughes Medical Institute predoctoral fellow. This work was supported by NIH grants P50GM82250 and P0GM56531 to A.D.F.

Figure Legends:

Figure 1: Overall structure of Rev dimer and Rev monomer

A) The asymmetric unit of the $Rev_{70\text{-}Dimer}$ crystals contains 2 dimers. Dimer 1 is shown as a blue surface with two ribbon diagrams for the two monomers (Chain A in dark blue, Chain B in light blue). Dimer 2 is shown similarly (Chain C in dark green, Chain D in light green). Interactions between the dimers are minimal, although the surface between them is the one that would be expected to form higher order oligomers of Rev. B) Overlay of the four Rev monomers represented in the asymmetric unit showing the high similarity of all monomers. Indicated are the structural and functional regions of the Rev protein. Shown in lines are the sidechains of Chain A from Dimer 1. C) Overlay of the two Rev dimers in the asymmetric unit. A difference of approximately 10 degrees is observed between the crossing angles of Dimer 1 (blue) and Dimer 2 (green).

Figure 2: Stabilization and conservation of the Rev monomer.

A) An expanded view of the oligomerization domains of Chain A. Shown as spheres are the hydrophobic residues observed to mediate folding and stabilization of all Rev monomers. Indicated as lines is the hydrogen bonding network between Asn 26 on helix 1 with Arg 48 and Gln 49 on helix 2. B) Conservation among all 1201 HIV-1 isolates represented in the Los Alamos HIV sequence database (http://www.hiv.lanl.gov/) is indicated on the ribbon diagram of Rev ranging from green (lowest conservation) to red (highest conservation). Residues involved in stabilization of the monomer structure are displayed as above but colored according to sequence conservation.

Figure 3: Rev dimerization interface

A) Cutaway of the Rev dimerization interface showing the surface representation of Chain A. Shown in blue is the solvent accessible surface not contacted by Chain B. Shown in orange is the predicted surface area buried upon dimerization. B) Expanded view of the three crucial hydrophobic residues (shown as orange lines) that mediate dimerization between Chain A (dark blue) and Chain B (light blue). C) Residues that form the core of the monomer structure (Leu 22 and Ile 55) are stabilized upon dimerization providing a structural basis for cooperative RNA binding. Spheres colored as blue are involved in hydrophobic packing of the monomer while spheres colored as orange are involved in hydrophobic packing of the dimer. D) View of the overall dimerization interface, showing the surface of Chain A with the dimerization residues from Chain B projecting into the surface.

Figure 4: Model for cooperative RNA binding and Crm1 binding

A) Electrostatic potential surface representation of the Rev dimer. Shown is the "inner face" of the ARM of Chain A (containing Asn 40) and the "outer face" of the ARM of Chain B. B) Two views of a model of the Rev dimer aligned with the known stem IIB-ARM structure[18], showing the projection of the end of the IIB RNA toward the second Rev ARM of the dimer. The right panel clearly shows that the "outer face" of the second ARM would be expected to contact the RNA, even though the "inner face" is used to anchor Chain A to stem IIB using Asn 40 (yellow). C) Expected location of the

disordered C-terminal residues involved in Crm1 binding. Shown in yellow are the modeled residues of the nuclear export sequence (residues 73-83).

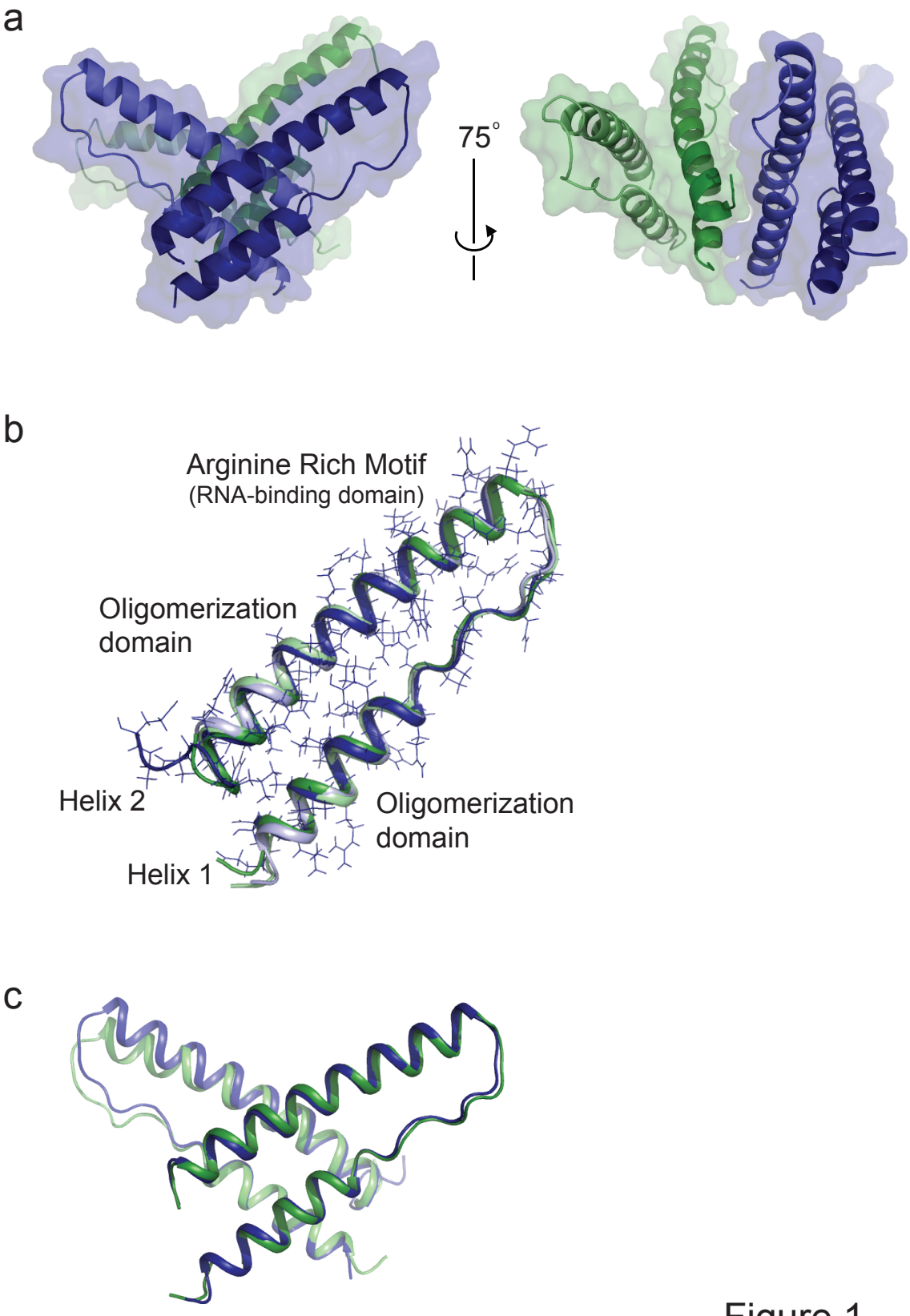
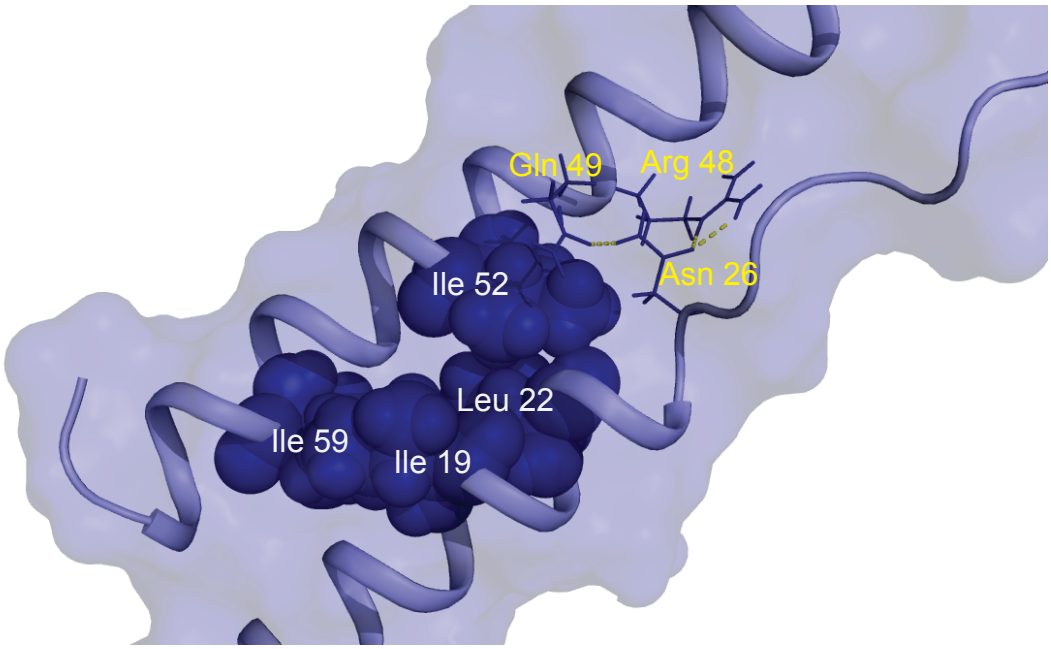

a
75°
b
Arginine Rich Motif
(RNA-binding domain)
Oligomerization
domain
Helix 2
Oligomerization
domain
Helix 1
c

Figure 1

a



b

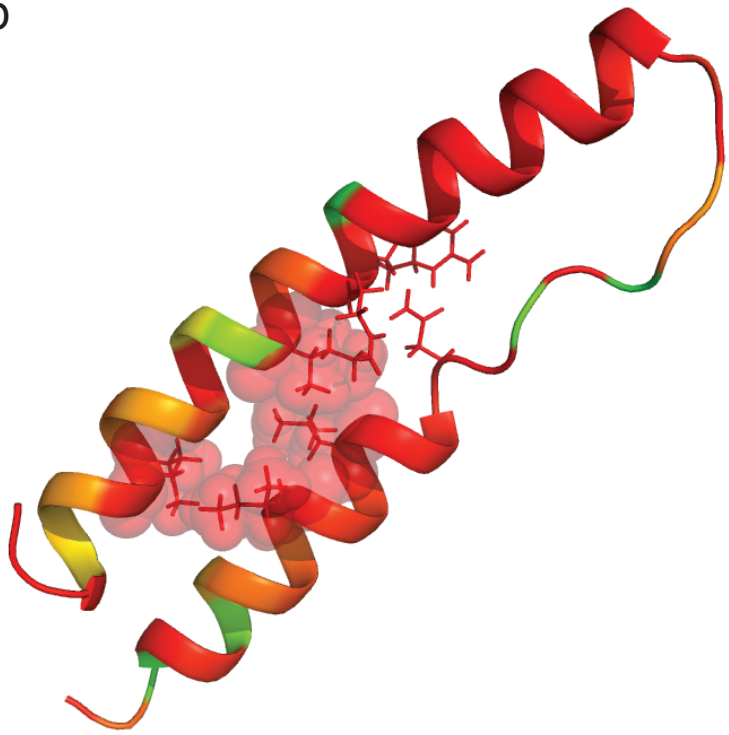


Figure 2

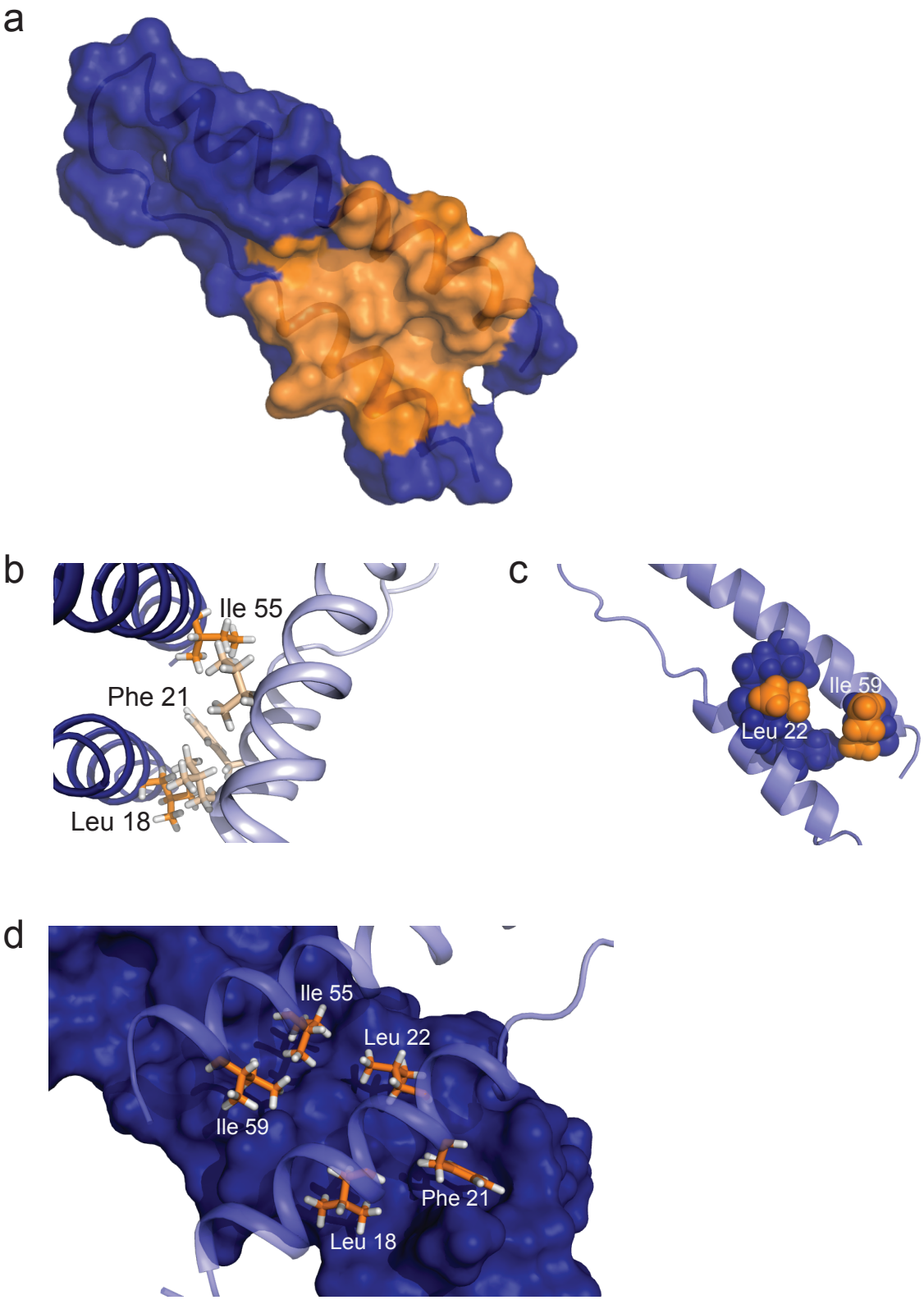

a
b
Ile 55
Phe 21
Leu 18
c
Ile 59
Leu 22
d
Ile 55
Leu 22
Ile 59
Phe 21
Leu 18

Figure 3

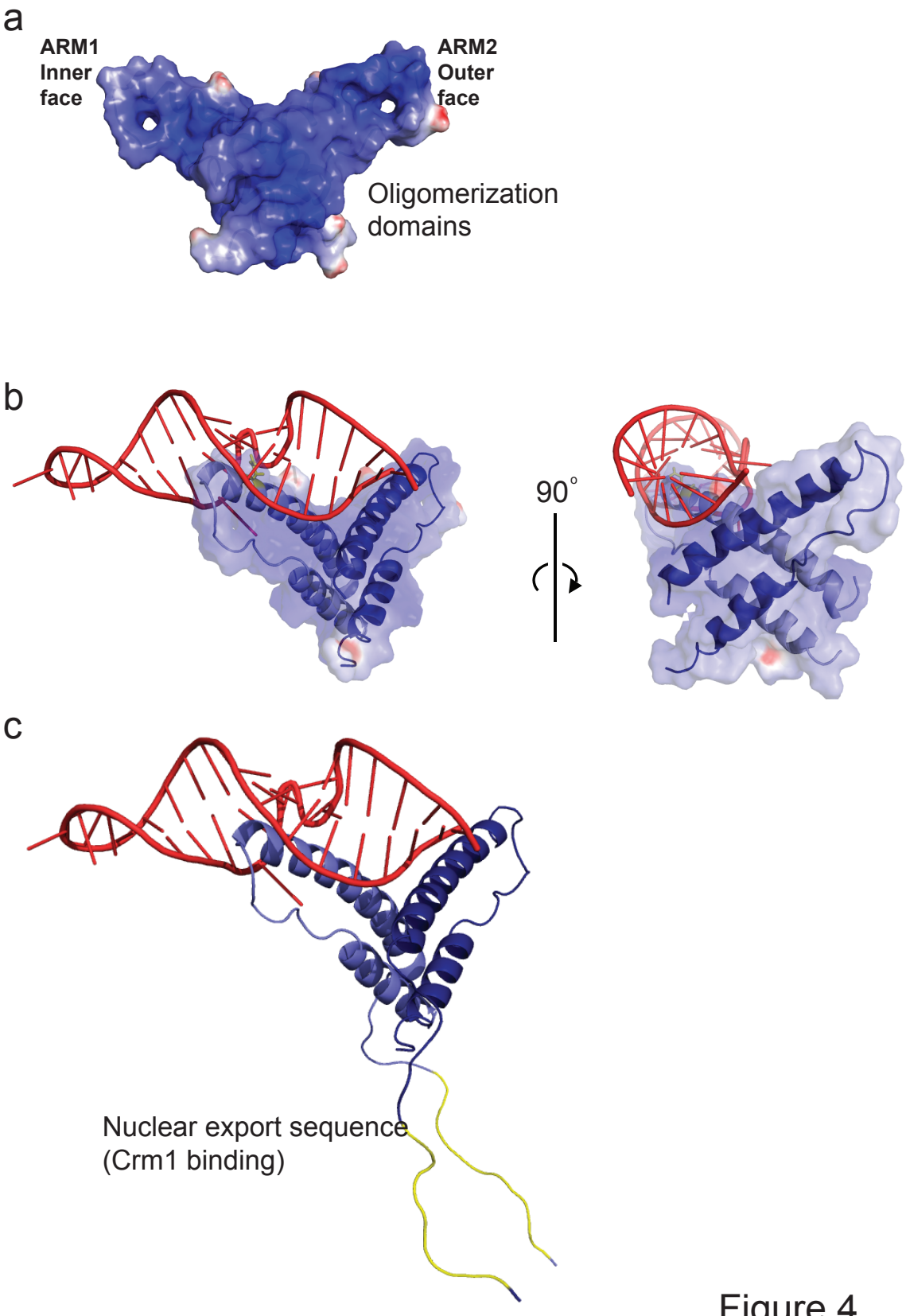

a
ARM1
Inner
face
ARM2
Outer
face
Oligomerization
domains
b
90°
c
Nuclear export sequence
(Crm1 binding)

Figure 4

Supplementary Materials and Methods:

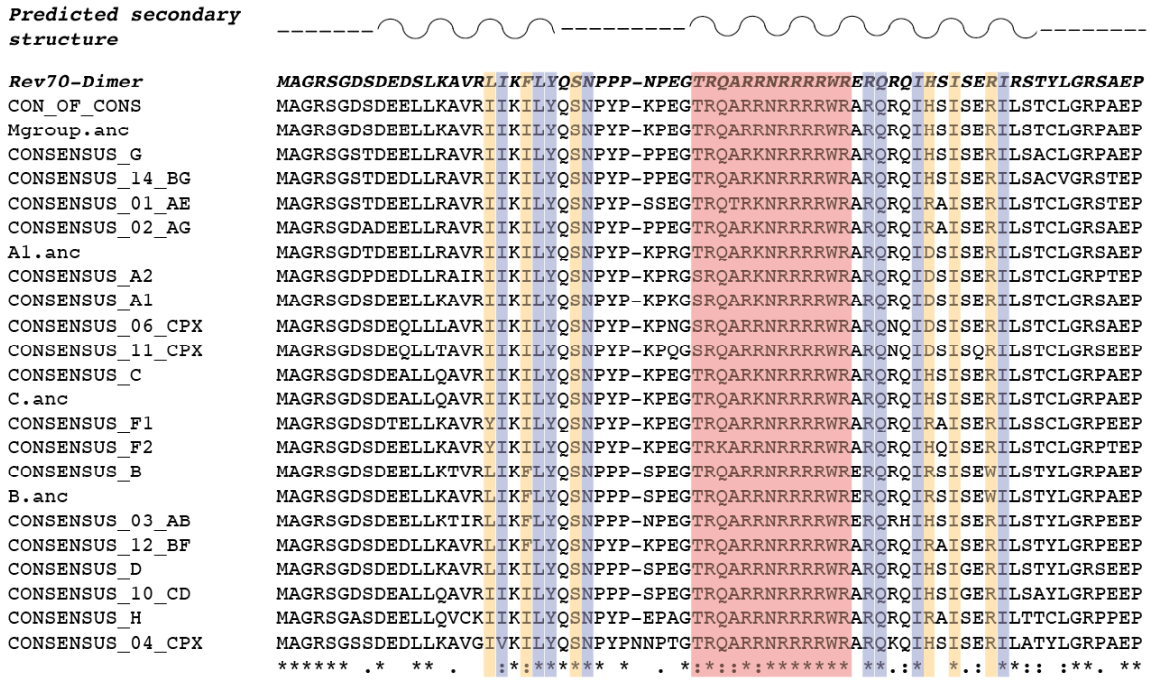
Phosphate-based minimal media

It was observed that the solubility of Rev proteins shorter than 96 amino acids was greatly enhanced by being expressed in a phosphate-based medium such as M9 rather than LB (data not shown). As a result, $Rev_{70\text{-}Dimer}$ was expressed in "SBMX", a phosphate-based minimal medium supplemented with metals and vitamins derived from one previously described[1]. Each liter of SBMX media contains 40 mL "25X Salts" (32.3 g KH_2PO_4, 175 g K_2HPO_4, 36.5 g NaCl, 1.2 g K_2SO_4 in 1 L H_2O), 4 g glucose, 1 g NH_4Cl, 1 mL 0.1% Thiamine-HCl, 2 ml "O solution" (500 mL total solution containing 26.8 g of $MgCl_2$•6 H_2O + 10 ml of a solution containing 92 ml H_2O, 8 ml concentrated HCl, 5 g $FeC1_2$•4 H_2O, 184 mg $CaC1_2$•2 H_2O, 64 mg H_3BO_3, 40 mg $MnCl_2$•4 H_2O, 18 mg $CoCl_2$•6 H_2O, 4 mg $CuC1_2$•2 H_2O, 340 mg $ZnCl_2$, and 605 mg Na_2NoO_4•2 H_2O), 1 mL 'Vitamin Solution' (2.2 mg biotin, 2.2 mg folic acid, 220 mg p-aminobenzoic acid, 220 mg riboflavin, 440 mg pantothenic acid, 440 mg niacinamide, 440 mg pyridoxine-HCl in 1 L 50% EtOH) and ampicillin to 100 ug/ml.

Size exclusion chromatography (SEC) and multi-angle light scattering (MALS)

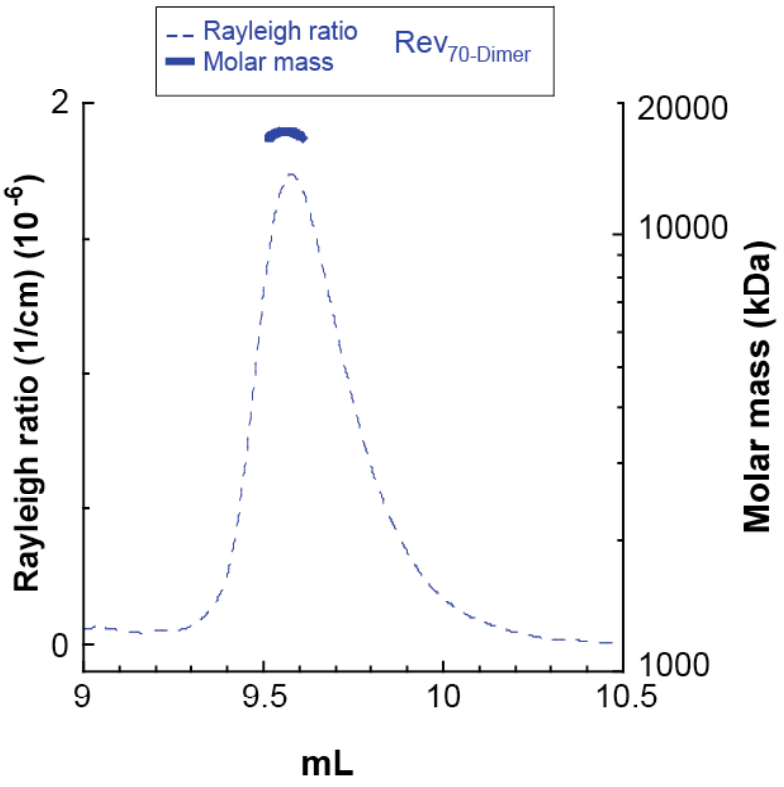
Analytical size exclusion chromatography was performed using a Ettan LC system (GE Life Sciences) with a silica gel KW803 column (Shodex) equilibrated in crystallization buffer at a flow rate of 0.35 ml/min. The system was coupled on-line to an 18-angle MALS detector (DAWN HELEOS II, Wyatt Technology) and a differential refractometer (Optilab rEX, Wyatt Technology). Molar mass determination was performed by the ASTRA 5.3.1.5 software.

Supplementary Figures:



Supplementary Figure 1

Consensus alignment of the first 70 residues of the Rev protein from reference subtypes (Los Alamos HIV sequence database (http://www.hiv.lanl.gov/)) represented using the ClustalW program[2]. The top sequence is the one used for crystallization and above that is the predicted secondary structure of this region of the protein[3]. Regions predicted to be helical are indicated by wavy lines, whereas coiled regions are indicated by dashed lines. Highlighted are residues shown in the Rev dimer structure to be important for Rev monomer stability and folding (blue) and Rev dimer formation (orange). Residues in red are the highly conserved residues of the ARM.



Supplementary Figure 2

Isolation of the Rev dimer used for crystallization. Initially, crystallization screens were performed with the $Rev_{70\text{-}Dimer}$ protein in complex with RNA and in the absence of sulfate salts. Crystals grew only in conditions containing ammonium sulfate as a precipitant and were later shown to contain no RNA. Although the protein had previously been shown to be unstable without RNA, at such elevated concentrations of ammonium sulfate (0.5 M total SO_4), the protein purifies well as a dimer. Shown is the measured multi-angle light scattering (MALS) (dashed lines, left axis) and calculated molar masses (solid line, right axis) determined from the major peak from size exclusion chromatography (SEC) of the $Rev_{70\text{-}Dimer}$ protein in crystallization buffer, which contains 0.5 M SO_4. The measured mass of 17 kDa corresponds to two Rev monomers (8.5 kDa each).

Supplementary Table:

	Native	SeO_4 – MAD		Br – SAD
Data collection				
Space group	$P6_422$	$P6_422$		$P6_422$
Cell dimensions				
a, b, c (Å)	115.8, 115.8, 81.2	115.9, 115.9, 81.4		116.1, 116.1, 81.3
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0		90.0, 90.0, 120.0
Wavelength	1.1159	0.9792	0.9715	0.9202
Resolution (Å)	50.0 – 2.5	50.0 – 2.8	50.0 – 2.8	50.0 – 3.0
R_{merge}	6.4 (56.0)	6.6 (40.0)	6.2 (33.5)	10.0 (59.8)
$I / \sigma I$	33.1 (4.2)	24.5 (3.5)	25.1 (5.1)	24.4 (3.8)
Completeness (%)	99.9 (100.0)	99.9 (99.9)	99.9 (99.9)	100.0 (100.0)
Redundancy	10.8 (9.9)	14.4 (12.4)	14.4 (12.4)	18.1 (18.0)
Refinement				
Resolution (Å)	47.2 – 2.80			
No. reflections	7933			
R_{work} / R_{free}	0.281/0.289			
No. atoms				
Protein	1766			
Ions/Water	48			
B-factors				
Protein	50.9			
Ions/Water	58.5			
R.m.s. deviations				
Bond lengths (Å)	0.002			
Bond angles (°)	0.514			

Supplementary Table 1: Crystallographic data collection and refinement statistics

Supplementary References:

1. Weber, D.J. et al. NMR docking of a substrate into the X-ray structure of staphylococcal nuclease. *Proteins* 13, 275-87 (1992).
2. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-8 (2007).
3. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292, 195-202 (1999).

Chapter 5

Implications and Future Directions

HIV is one of the most well-studied organisms in existence. In twenty-five years of research, we have learned an enormous amount about the essential systems of the virus as well as the systems in the host cell that it exploits. However, there remains a great deal that we do not understand about the virus.

The HIV Rev protein is an excellent example of all of these points. Aspects of the protein and its cognate RNA element, the Rev Response Element (RRE), have been well characterized at the cellular, molecular and structural level[1]. Moreover, studies of Rev-mediated RNA export were paramount in identifying an entire cellular system for nuclear export that employs the human protein Crm1[2]. But our knowledge of the Rev protein remains incomplete.

Much of the research in the last two decades has focused on minimal components of the Rev-RRE system, specifically the binding of the Rev arginine-rich motif (ARM) peptide to the stem IIB element in the RRE. As a result, we understand this interaction at a functional and structural level[1,3]. Importantly, investigation of induced fit binding by ARM peptides such as those from HIV Tat and Rev has helped provide the basis for understanding the role of binding mode adaptability in molecular recognition[4-6].

The mechanism and structure of Rev oligomerization has been less well explored. In large part, this has been due to two issues. First, initial observations indicated that oligomerization was not an essential step in specific recognition of the RRE[7-13]. As a result, studies aimed at revealing the role of oligomerization in Rev-RRE recognition were supplanted by more in-depth analysis of the interaction between the Rev ARM and stem IIB. Oligomerization was clearly required for RNA export, but one of the only proposed models for its role in vivo was that multiple Rev nuclear export sequences were

required for efficient binding to Crm1[14]. Structurally, the models that emerged were ones in which Rev coats the RNA[8,10,15,16], similar to the way proteins coat genomic RNA during viral packaging[17]. With the exception of the work by Jain and Belasco[16,18], little headway had been made since the early 1990's in the mechanism of oligomerization and cooperative RNA binding.

The second barrier to studying Rev oligomerization was the poorly behaved nature of the Rev protein. The solubility limit of the protein was observed to be in the low micromolar range[8,19], which has hampered many biochemical studies, and has made most structural work impossible. Until the studies described in this thesis, the highest resolution structural information on the Rev protein came from studying filaments that precipitate out of a Rev solution under specific conditions[8,19-21], but even the authors of those studies mention that these filaments are not likely to be physiologically relevant[15,21]. Removal of the oligomerization domains abrogates the solubility problem, so this was assumed to be a problem inherent to working with the full protein. In hindsight, the results described in Chapter 3 indicate that electrostatic considerations are more important in solubilizing the Rev protein than oligomerization. Nevertheless, this solubility limit had created a reputation for Rev that meant that biochemical studies were only pursued by the brave and structural work was only pursued by the desperate or foolhardy.

Although I did not originally intend to solve the structure of Rev, it was my goal to provide a more comprehensive understanding of Rev oligomerization at the mechanistic and structural level. The in vitro selection studies on the Rev ARM[22-24] indicated that there might be additional specific sites in the RRE for Rev binding, so that

is where my research began. What rapidly became obvious was that recognition of additional sites in the RRE requires cooperativity in binding that is mediated by protein-protein interactions. Thus, the model that emerges from the results described in Chapter 2 is one in which oligomerization is crucial to highly specific recognition of the RRE. Though this model is of very low resolution (with proteins being ovals), it strongly suggests that there is a defined structural basis to the complex that forms between Rev and the RRE, rather than 'runaway' oligomerization that is nucleated by binding to stem IIB.

In retrospect, there had been previous indications that the Rev-RRE complex is structurally defined, most obviously in work by Jonathan Karn's group[15,25] and the aforementioned Jain and Belasco study[16]. However, the work in Chapter 3 allows us to finally define discrete complexes between Rev and RNA at a biophysical and structural level. Most importantly, the EM data firmly establish that the Rev-RRE complex is a distinct ribonucleoprotein (RNP) that is amenable to structural studies in the way that the ribosome and spliceosome are.

Also crucial to the work described in Chapter 3 is the role of electrostatics in Rev solubility. The importance of the serendipitous choice of GB1 as a fusion domain for the Rev protein (with enormous thanks to Dr. John Gross for providing it) can not be underappreciated. Following the poor binding of a 6X His-tagged version of Rev to a Ni-NTA affinity column, I decided to try avoiding His-tag occlusion by distancing the purification tag from the N-terminus of the protein. I chose to do this by placing a solubility enhancing domain between them. Though a Glutathione S-transferase (GST) domain worked well, the GB1 domain worked better, so it was pursued further. This was

probably one of the best choices I made in graduate school. Chapter 3 describes the effect that GB1 has on Rev solubility by acting as an RNA surrogate. Importantly, this allows for Rev expression and purification in a highly soluble and well folded form. Once RNA, or an oxyanionic ion such as sulfate, is presented to Rev, GB1 can be removed and Rev retains its highly soluble and well folded form. It is possible that other routes to a well folded Rev protein are available, but this was the route that we have taken. Significantly, this makes structural studies of Rev no longer for the foolhardy or the brave, but the obvious next step.

Though the Rev protein continued to be not as straightforward as hoped (described in Appendix 1), the structure of the Rev dimer was finally solved almost exactly one month before the writing of this thesis. That crystal structure, described in Chapter 4, ends a two decade pursuit of the structure of the Rev protein. It also reveals the structural basis for cooperative RNA binding, and clearly visualizes the dimerization surface of Rev. Since dimerization and cooperative RNA binding have been shown to be essential for Rev-RRE assembly and RNA export[10,16] (Chapter 2), this represents another potential site for small-molecule intervention. With the continued spread of HIV despite increased access to antiretrovirals targeting the other structurally characterized HIV proteins, the physical details of the Rev dimer may be of use in development of novel therapeutics.

In spite of the structural characterization presented in Chapters 3 and 4, the work in this thesis is really just the beginning of studies of the Rev-RRE export complex. The methods described have already enabled our NMR, SAXS, EM and X-ray

crystallography studies, but they also open the door to future characterization of Rev with the RRE and even with host cell factors. With the demonstration that the Rev protein and Rev-RNA complexes are soluble, can be discretely defined, and are amenable to structural characterization, I believe our structural and functional understanding of the Rev-RRE export complex is on the verge of exponential expansion. What follows is my expectation for the directions of this expansion.

The first avenue that needs to be pursued is to complete the crystal structure of the Rev dimer. As mentioned, the phasing data that allowed initial determination of the structure was acquired a little over a month ago. With the inclusion of higher resolution native data sets (Chapter 4, Supplementary Table 1) and proper refinement, I expect that we will gain more insight into the details of the Rev dimer structure.

Immediate additional crystallographic goals include generating a high resolution structural model of the higher order Rev oligomer and of the Rev oligomer with RNA. In Appendix 1, I describe initial unsuccessful crystallization screens that were employed to capture the higher order oligomer with wildtype Rev protein (rather than an oligomerization mutant), and capture the Rev-RNA complex with the Rev dimer on various RNAs. However, these attempts were by no means exhaustive, and with the structural insight generated from the Rev dimer structure, I have little doubt that these structures will be solved.

Determination of all of these sub-structures is obviously an attempt to build up to a structure of the full Rev-RRE complex. Beginning with the EM data from Chapter 3, I believe the molecular details of the functional Rev-RRE RNP will become clearer. An X-ray crystal structure may be a ways off in the future, due to the need for the

biochemical definition of a stable, homogenous and crystallizable complex, but low resolution EM models should be immediately attainable. Experiences with the ribosome, spliceosome and viral nucleocapsid complexes suggest that the EM model of the Rev-RRE complex will precede a crystal structure by a number of years. However, in the above three cases, X-ray crystal structures were eventually solved[26-30] suggesting that the Rev-RRE crystal structure will eventually be revealed.

Building up from there will involve forming defined complexes with host cell factors. Primary on that list will be a Rev-Crm1-RanGTP complex, and ideally a Rev-RRE-Crm1-RanGTP complex. The recent crystal structures of Crm1 and Ran with other molecular cargoes[31,32] will no doubt facilitate assembly and characterization of these essential host factors with Rev-RRE. Numerous other factors have been implicated in Rev-mediated export, but biochemical characterization of those interactions will need to precede structural determination.

It is my hope that the description of a biochemically well-behaved Rev protein will also promote more mechanistic and functional work with the Rev protein. A number outstanding questions about the role of Rev in viral replication could be resolved with careful studies employing this folded Rev protein. One study already underway is the kinetic dissection of Rev binding to the RRE. All of our work has employed equilibrium measurements of Rev binding to the RRE, but the Millar group has been actively investigating the order and cooperativity of Rev-RRE assembly using single-molecule kinetic measurements[33]. We are quite curious to see how the Rev protein that cooperatively assembles on the RRE with high affinity is able to kinetically recognize and bind to the RRE.

At a larger scale, the role of Rev outside of the nucleus is still unknown. Certainly one open question is how the Rev-RRE complex is disassembled. With an in vitro binding affinity of ~100 pM in our assays, it seems unlikely that this complex will passively fall apart. Implicated in this disassembly function are the DEAD-box helicases DDX3[34] and DDX1[35], but this process is not fully understood. In fact, this may be yet another way that the Rev-RRE system may illuminate aspects of human cell biology; disassembly and release of exported cargo from the nuclear pore is poorly characterized and use of the viral Rev-RRE system may be instrumental in future studies. Moreover, the role, if any, that Rev plays in cytoplasmic shuttling of the RRE, translation of viral proteins, or packaging of viral RNA, remains unknown and may be more accurately studied at a biochemical level with a well folded Rev protein.

Finally, I believe the Rev-RRE complex is a fascinating model system to understand how molecular recognition evolves. As discussed in Chapter 2, HIV is like many viruses in its requirement for a compact genome that is highly resistant to the deleterious mutations that can result from low-fidelity replication of the genome. The use of the adaptable Rev ARM in the context of a homooligomeric protein seems ideally suited to meet these needs. In this way, any single interaction between the protein and the RNA can be disrupted without the whole complex falling apart, and the coding capacity of the single 116 amino acid protein is quite small. Even more interesting is that both Rev and the RRE overlap with coding regions of the *Env* gene, suggesting that there are multiple selective pressures on Rev and the RRE. This results, I believe, in a system that has evolved to be flexible as possible at both the structural and the genomic level.

What this tells us about the nature and evolvability of molecular recognition remains to be seen but is something in which I am quite interested.

References:

1. Pollard, V.W. & Malim, M.H. The HIV-1 Rev Protein. *Annu. Rev. Microbiol.* **52**, 491-532 (1998).
2. Cullen, B.R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**, 419-24 (2003).
3. Battiste, J.L. et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547-1551 (1996).
4. Frankel, A.D. & Smith, C.A. Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell* **92**, 149-51 (1998).
5. Williamson, J.R. Induced fit in RNA-protein recognition. *Nat Struct Biol* **7**, 834-7 (2000).
6. Leulliot, N. & Varani, G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **40**, 7947-56 (2001).
7. Cook, K.S. et al. Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* **19**, 1577-83 (1991).
8. Heaphy, S., Finch, J.T., Gait, M.J., Karn, J. & Singh, M. Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. USA* **88**, 7366-70 (1991).

9. Huang, X. et al. Minimal Rev-response element for Type 1 human immunodeficiency virus. *J. Virol.* **65**, 2131-34 (1991).

10. Malim, M.H. & Cullen, B.R. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241-48 (1991).

11. Iwai, S., Pritchard, C., Mann, D.A., Karn, J. & Gait, M.J. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res.* **20**, 6465-72 (1992).

12. Kjems, J., Calnan, B.J., Frankel, A.D. & Sharp, P.A. Specific binding of a basic peptide from HIV-1 Rev. *EMBO J* **11**, 1119-29 (1992).

13. Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. & Cullen, B.R. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc. Natl. Acad. Sci. USA* **89**, 758-62 (1992).

14. Askjaer, P. et al. RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. *Mol. Cell. Biol.* **19**, 6276-6285 (1999).

15. Mann, D.A. et al. A molecular rheostat: Co-operative rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J. Mol. Biol.* **241**, 193-207 (1994).

16. Jain, C. & Belasco, J.G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7**, 603-614 (2001).

17. Harrison, S.C., Skehel, J.J. & Wiley, D.C. Virus structure. in *Virology*, Vol. 1 (ed. Fields, B.N.) 59-99 (Lippincott-Raven, Philadelphia, 1996).

18. Jain, C. & Belasco, J.G. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell* **87**, 115-125 (1996).

19. Wingfield, P.T. et al. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. *Biochemistry* **30**, 7527-34 (1991).

20. Blanco, F.J., Hess, S., Pannell, L.K., Rizzo, N.W. & Tycko, R. Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J Mol Biol* **313**, 845-59 (2001).

21. Havlin, R.H., Blanco, F.J. & Tycko, R. Constraints on protein structure in HIV-1 Rev and Rev-RNA supramolecular assemblies from two-dimensional solid state nuclear magnetic resonance. *Biochemistry* **46**, 3586-93 (2007).

22. Xu, W. & Ellington, A.D. Anti-peptide aptamers recognize amino acid sequence and bind a protein epitope. *Proc. Natl. Acad. Sci. USA* **93**, 7475-80 (1996).

23. Bayer, T.S., Booth, L.N., Knudsen, S.M. & Ellington, A.D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA* **11**, 1848-57 (2005).

24. Landt, S.G., Ramirez, A., Daugherty, M.D. & Frankel, A.D. A simple motif for protein recognition in DNA secondary structures. *J. Mol. Biol.* **351**, 982-94 (2005).

25. Zemmel, R.W., Kelley, A.C., Karn, J. & Butler, P.J. Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. *J. Mol. Biol.* **258**, 763-77 (1996).

26. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-20 (2000).

27. Albertini, A.A. et al. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360-3 (2006).

28. Green, T.J., Zhang, X., Wertz, G.W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357-60 (2006).

29. Ye, Q., Krug, R.M. & Tao, Y.J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078-82 (2006).

30. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458**, 475-80 (2009).

31. Dong, X. et al. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* **458**, 1136-41 (2009).

32. Monecke, T. et al. Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* **324**, 1087-91 (2009).

33. Pond, S.J., Ridgeway, W.K., Robertson, R., Wang, J. & Millar, D.P. HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc Natl Acad Sci U S A* **106**, 1404-8 (2009).

34. Yedavalli, V.S., Neuveut, C., Chi, Y.H., Kleiman, L. & Jeang, K.T. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **119**, 381-92 (2004).

35. Fang, J. et al. A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. *Virology* **330**, 471-80 (2004).

Appendix 1

Crystallization and preliminary X-ray analysis of HIV-1 Rev: the role of cryoprotectants in phase transitions of crystal symmetry

The molecular structure of Rev has been unknown for nearly two decades. We previously described methods to assemble discrete complexes of Rev with RNA that are highly soluble and have defined stoichiometry (Daugherty et al., in preparation). One tangible goal of those studies was to use these complexes as starting points for crystallization and structure determination of the Rev protein or Rev-RNA complexes.

Initial crystallization trials

Two complexes were selected initially for crystallization screening of 500-700 conditions. The first is the full length, wildtype Rev protein with the solubility enhancing GB1 tag still appended to it. Though this complex is known to be large in size and heterogeneous in its oligomerization state (Daugherty et al., in preparation), screening was initiated in the hope that crystallization may trap a defined oligomer. The second complex that was screened was a discrete complex between the Rev dimer and a version of the stem IIB RNA which contains a second Rev binding site[1]. In order to prevent oligomerization beyond a dimer, this protein contains two known mutations in the oligomerization domains, L12S and L60R[2]. No conditions yielded crystals even after many months.

To eliminate disordered regions of Rev, the C-terminal 46 residues were removed, which are experimentally and computationally predicted to be unstructured[3] (Daugherty et al., in preparation). This construct, named $Rev_{70\text{-}Dimer}$, was purified and assembled with the same IIB RNA as the full-length Rev protein described above and screened for suitable crystallization conditions. Multiple conditions produced promising crystals

within two weeks. In all cases, crystals appeared in conditions with 1.5-1.8 M ammonium sulfate as a precipitant and displayed a morphology of hexagonal rods with maximum dimensions of approximately 30 um x 200 um. The results were reproducible on a microliter scale with crystals appearing within 2-3 days. Initial optimization indicated that the concentration of ammonium sulfate was the primary determinant of crystal number and size. Promising initial diffraction was obtained from these crystals (see below).

Extensive washing and subsequent dissolving of these crystals demonstrated that the only component in the crystal was protein, based on native gel analysis and UV absorbance (data not shown). To support the conclusion that the crystal only contained protein and to attempt to find an RNA construct compatible with protein-RNA complex crystallization, crystallization trials were set up with different RNA constructs (34mer, 36mer and 42mer treated with alkaline phosphatase to remove the 5' triphosphate) (Table 1). In all three cases, very similar crystals were obtained from identical ammonium sulfate conditions and no additional hits were seen.

To confirm that only protein was being crystallized, we asked whether identical crystals could grow from a solution lacking any RNA. In buffer conditions that the Rev-RNA complex is soluble (40 mM Tris, 200 mM NaCl, 2 mM DTT), the protein alone is insoluble at high concentrations. However, we inferred from the lack of RNA in the high ammonium sulfate containing crystals that the protein must be soluble in intermediate concentrations of ammonium sulfate. Therefore, we added 400 mM ammonium sulfate to the protein and observed that the protein is soluble without RNA to >10 mg/ml. The purified apo protein was used for crystallization screening. Due to the fact that the final

protein buffer (40 mM Tris pH 8.0, 200 mM NaCl, 100 mM Na_2SO_4, 400 mM $(NH_4)_2SO_4$, 2 mM DTT) had a higher osmolarity than many of the crystallization buffers, many conditions produced drops of increased volume. However, in all other conditions, only well solutions containing 1.5-1.8 M ammonium sulfate produced crystals. Crystals on the microliter scale appeared under similar conditions and revealed identical morphology and unit cell dimensions to the crystals set with Rev-RNA complexes. Crystallization screening under identical conditions with the Rev_{70} protein with no oligomerization mutations were unsuccessful at producing crystals (Table 1).

Extensive optimization with the apo $Rev_{70\text{-}Dimer}$ protein crystals demonstrated that identity of the buffer (HEPES, Tris and sodium cacodylate) had little influence on crystallization, and pH's ranging from 6.0 to 9.0 appeared able to support crystallization, albeit with slightly different optimal ammonium sulfate concentrations. Diffraction of the crystals appeared to be limited by the narrow width of the crystals, which at most grew to 60 um while the lengths were at times as long as 800 um (Figure 1A). Only at high concentrations of protein or high concentrations of ammonium sulfate were crystals able to grow in width:length ratios of greater than 0.2:1, but crystals were very small due to extensive nucleation. Initial crystallization hits, as well as additive screens indicated that low concentrations of NaCl, PEG 400 and PEG 1000 were able to reduce nucleation and might therefore be used to increase the width:length ratio. Indeed, optimization revealed that NaCl in concentrations of 50-100 mM and PEG 1000 at 3% reduced nucleation to a point at which high protein and precipitant concentrations could be used and the resulting width:length ratio was as high as 2:1. In addition to reducing nucleation, low concentrations of NaCl increase the width:length ratio without adjusting

protein or precipitant concentrations, making NaCl a doubly important additive to the crystallization buffer. Under optimal crystallization conditions, in which 1 ul Rev protein solution (1.3-1.5 mM $Rev_{70\text{-}Dimer}$) is mixed with 1 ul crystallization buffer (100 mM Tris pH 8.0, 1.45-1.55 M $(NH_4)_2SO_4$, 50-100 mM NaCl and 3% PEG 1000), crystals appears within 2-3 days and grow to a maximal size of 150 um x 200-300 um within 10-14 days (Figure 1B).

Cryoprotectant screening

Cryoprotectant screening was initiated using the work of McFerrin and Snell as a starting point[4] due to the fact that the final Rev crystallization condition was close to conditions tested in that work. Glycerol, PEG 400 and ethylene glycol were screened for use as a cryoprotectant for Rev crystals. Preliminary trials employed both rapid transfer and slow incremental transfer to each cryoprotectant. The most promising initial candidate for a cryoprotectant was a short exposure to PEG 400 based on diffraction quality. Encouragingly, short exposures to PEG 400 yielded diffraction to better than 4 Å with unit cell of approximate dimensions 118 Å x 118 Å x 165 Å and belonging to space group $P6_2$ or $P6_4$ (called 'Cell 1', see Table 2). Interestingly, all other cryoprotectant conditions yielded diffraction to worse than 7 Å with unit cell dimensions that were doubled in the c-axis but with increased point group symmetry (118 Å x 118 Å x 330 Å, $P6_222$ or $P6_422$, called 'Cell 2', see Table 2). Further trials including short exposures to glucose, PEG 1000 and sodium malonate revealed that cryoprotecting with glucose or PEG 1000 also resulted in diffraction with 'Cell 2' dimensions while cryoprotecting with

sodium malonate resulted in weaker diffraction with 'Cell 1' dimensions. Attempts at collecting diffraction images from unfrozen crystals were hampered by the very high air sensitivity of ammonium sulfate grown crystals as has been observed previously[5].

Optimization with PEG 400 as a cryoprotectant yielded excellent data sets (Table 2) but with inconsistent quality. In particular, exposure of the crystal to cryoprotectant for more than 3-5 minutes often resulted in diffraction consistent with 'Cell 2'. More troubling was the occasional increase in the estimated twin fraction, as assessed by phenix.xtriage[6] and other twinning estimation programs. In the most extreme cases, the diffraction data could be merged and scaled in space group $P6_222$ or $P6_422$, but with an estimated twin fraction of 0.5.

Initial data sets

Even with data sets that had an estimated twin fraction of <0.03, structure determination from 'Cell 1' diffraction data was unsuccessful due to a lack of phasing information. Due to the fact that no diffraction quality crystals of SeMet constructs could be found, three challenges made this problem particularly difficult to solve. First, the crystals were grown in high concentrations of ammonium sulfate, which is incompatible with many potential heavy-metal compounds that can be used for experimental phasing[7]. Second, the propensity of the unit cell to undergo phase transitions to a larger unit cell or one with an increased twin fraction hampered experimental phasing with many other heavy metal compounds. With only brief exposures to cryoprotectant yielding high-quality diffraction, long soaks were impossible. Even with brief exposures to halides

such as sodium bromide[8], data sets showed reasonable anomalous signal but always diffracted poorly with 'Cell 2' dimensions. The final challenge was with the large, and undetermined, number of Rev monomers in the asymmetric unit. Estimating a range of solvent contents from 40%-75% and assuming a dimer as the basic oligomeric state, the total number of Rev monomers in the asymmetric unit could be from 8 to 18. This makes molecular replacement with poor models and experimental phasing methods significantly more difficult.

Using selenate in place of sulfate in the crystallization buffer, we were able to obtain MAD data sets to 4 Å of crystals with 'Cell 1' dimensions but heavy metal sites could never be conclusively found and the structure was never solved. This could be a result of the large volume of the asymmetric unit, low occupancy of selenate sites, low resolution of the data or any number of other problems. With the recent solution of the structure with 'Cell 4' unit cell dimensions (see below), it will be interesting to see if this solution will be useful as a molecular replacement model for data from 'Cell 1'.

Malonate as a cryoprotectant

To try to circumvent the challenges of experimental phasing with crystals grown from ammonium sulfate conditions, we returned to sodium malonate as a cryoprotectant. Numerous studies have found sodium malonate to be a suitable stabilizing solution and cryoprotectant for crystals grown in high salts, especially ammonium sulfate[5,9]. Though initially not promising, we expected that diffraction from crystals cryoprotected in sodium malonate could be optimized. It was soon found that rapid transfer to sodium

malonate causes many crystals to crack, but that a slow transition from the crystallization buffer to sodium malonate gives very consistent diffraction data. Importantly, this not only resolves the problem of ammonium sulfate in the cryoprotectant, but also reveals that crystals transferred to sodium malonate are stable for up to 24 hours before freezing with no obvious reduction in the quality of the diffraction.

The unexpected positive effect of sodium malonate was that it also causes changes in the unit cell. Rapid transfer and exposure to malonate (<10 seconds) gives 'Cell 1' dimensions, but slow transfer and/or long exposure causes a significant change in the diffraction pattern. Most obviously, at crystal angles that directly visualize the c-axis dimensions, a pattern emerges of two intense spots separated by 3 weak spots (Figure 2A). Indexing the unit cell using all 5 spots reveals a c-axis dimension of ~160 Å, similar to 'Cell 1'. Merging and scaling reveals an increase in the point group symmetry (from P6 to P622 with no observed twinning). More interestingly, the native Patterson map reveals a strong peak (fractional coordinates x, y, z = 0.000, 0.000, 0.250) with a height 80-90% of the origin height, indicating substantial pseudo translational symmetry or a reduced unit cell size (Figure 2B). Predictably, indexing with only the most intense spots in the diffraction pattern yields a unit cell size of 118 Å x 118 Å x 40 Å and point group symmetry of P622 (called 'Cell 3'). In this way, changing only the cryoprotectant, the volume of the asymmetric unit is reduced by a factor of 8, from 330,000 $Å^3$ in 'Cell 1' to 42,000 $Å^3$. Using a estimated V_m of 2.4 $Å^3$/Da, corresponding to a solvent content of 48%, this asymmetric unit is predicted to contain a single 17 kDa Rev dimer. Importantly, that also predicts that the contents of the 'Cell 1' asymmetric unit contains 8 Rev dimers.

Troubling to us was the fact that the less intense spots that corresponds to the larger unit cell are present in all diffraction data measured. Under no conditions were those spots absent, including changes in pH, malonate concentration, SO_4 additives or exposure time to cryoprotectant. This suggested to us that the most basic unit cell might be larger than 'Cell 3'. Moreover, the diffraction of these crystals seemed to be limited to ~3 Å regardless of the above changes in conditions.

Sodium bromide causes an additional phase transition

Our experience with PEG 400 cryoprotected crystals ('Cells 1 and 2') described above indicated that sodium bromide may be useful as a heavy metal derivative based on the anomalous signal we observed with NaBr soaked crystals. Unfortunately, with that particular cryoprotectant condition, many added salts including Na_sSO_4, NaCl, NaI and NaBr cause an apparent increase in unit cell size from 'Cell 1' to 'Cell 2' dimensions and drastically decrease resolution limits. Crystals cryoprotected in sodium malonate ('Cell 3') are less sensitive to changes in cryoprotectant concentrations, exposure times and additives, so we wished to test sodium bromide as a heavy metal derivative using sodium malonate as a cryoprotectant. Remarkably, when NaBr was included at a concentration of 0.5 M in the cryoprotectant, we observed yet another unit cell transition, to a unit cell size of 118 Å x 118 Å x 81 Å and point group symmetry of P622 (called 'Cell 4'). Although larger in volume than 'Cell 3', we were pleased to see no evidence of less intense spots corresponding to a unit cell with larger dimensions. In fact, the diffraction limit was also much improved, with data sets being measured to 2.5 Å, similar to our best

'Cell 1' data sets. A significant peak in the native Patterson map corresponding to a translational symmetry operator of ~40 Å along the c-axis is still present (ranging from 30-50% of the origin peak height). Incredibly, this effect appears to be specific to high concentrations of sodium bromide. Reducing the concentration to 200 mM or replacing NaBr with NaCl, NaI, RbCl or Na_2SO_4 produces diffraction with 'Cell 3' dimensions.

Unfortunately, although strong anomalous signal is seen in MAD data sets collected using the bromide edge, no consistent set of heavy metal sites could be located even using data from 3 wavelengths.

Structure determination using sodium selenate

Due to the fact that data collected on sodium bromide soaked crystals was unable to solve the Rev structure, we returned to testing other heavy metal derivatives. The first attempt we made was using sodium selenate, since it had also yielded strong anomalous signal in 'Cell 1' conditions although those data had not allowed us to determine the structure. We were hopeful that in an asymmetric unit predicted to only contain 2 Rev dimers (83,000 $Å^3$), sodium selenate would be a useful heavy metal derivative. Using a cryoprotectant condition containing sodium malonate, 0.5 M NaBr and 0.25 M Na_2SeO_4, we observed consistent diffraction better than 3 Å with 'Cell 4' dimensions. As reported recently (Daugherty et al., in preparation) MAD data sets were collected and used to locate 2 selenate sites, which allowed automatic building of parts of all 8 expected alpha-helices. Further details of structure determination and refinement are described elsewhere (Daugherty et al., in preparation).

In conclusion, we have found that the manner in which Rev crystals are cryoprotected has an enormous effect on the resulting unit cell and point group symmetry. Similar, although less drastic effects, in unit cell dimensions have been reported with other crystals[10,11], including one case of sodium malonate playing a role in the doubling of the c-axis of a unit cell with point group symmetry of P6[12]. The effects that we observe with the Rev crystals are likely the result of defined binding of malonate and bromide ions to Rev monomers near the interface of non-crystallographic symmetry axes in the large unit cell. This binding presumably results in slight structural changes in the orientation of Rev monomers or dimers, which results in a transition in crystallographic unit cell dimensions and space group symmetry. Indeed, we observe a small peak in the native Patterson map of data from the 'Cell 1' crystal form at the same fractional coordinates as 'Cell 3' (fractional coordinates x, y, z = 0.000, 0.000, 0.250) but with a much reduced height of only 5-7% of the origin height. This suggests that binding to malonate rearranges the packing of the Rev proteins enough to allow for increased pseudo translational symmetry, and that additional binding to bromide allows for the translational NCS to again readjust and result in a reduction in the crystallographic unit cell. A similar phenomenon is observed between a NCS symmetry operator in the κ=180° of a self-rotation function in the 'Cell 1' data and the two-fold crystallographic axis in forms with P622 point symmetry. Again, it would appear that binding to specific ions such as malonate and bromide facilitates the reorientation of individual Rev monomers or dimers in the unit cell to allow for increased crystallographic symmetry. We are interested to determine the molecular details of these rearrangements, and

establish whether the alternate arrangements of the Rev protein in the various crystal forms can provide insight into the function of this essential HIV protein.

Methods:

Proteins and RNAs

HIV-1 Rev from isolate HXB3[13] was inserted into the pHGB1 as previously described[1]. The sequence of HXB3 Rev is:

MAGRSGDSDEDLLKAVRLIKFLYQSNPPPNPEGTRQARRNRRRRWRERQRQIHSISERI

LSTYLGRSAEPVPLQLPPLERLTLDCNEDCGTSGTQGVGSPQILVESPTILESGAKE

Full length (Rev_{116}) and only the first 70 amino acids (Rev_{70}) were introduced into pHGB1. Mutations in the oligomerization domains (L12S and L60R) or for SeMet introduction were made using standard site-directed mutagenesis protocols (Stratagene).

RNA transcription and purification was performed as previously described for stem IIB derived RNAs[1]. Templates used for RNA transcription are as follows:

42mer

GGAGGCCTGTACCGTCAGCTTGCGCTGCGCCCATATACCTCCCTATAGTGAGTCGTATTA

34mer

GGCCTGTACCGTCAGTTGCCTGCGCCCATATACCCTATAGTGAGTCGTATTA

36mer

GGGCCTGTACCGTCAGTTGCCTGCGCCCATATACCCCTATAGTGAGTCGTATTA

Protein expression and purification

Rev protein was expressed in *E. coli* strain BL21/DE3 from pHGB1-derived vectors as N-terminal fusions with a His_6 tag, GB1 domain and a TEV-protease cleavage

site. Cells were grown with 100 ug/ml ampicillin to $O.D._{600}$ = 0.8 at 37°C in either LB medium or SBMX, a phosphate-based minimal medium[14] (Daugherty et al., in preparation). It was observed that using phosphate-based M9 or SBMX medium, as well as including 100 mM Na_2SO_4 in the lysis buffer, drastically increased the soluble yield of Rev proteins that are shorter than 96 amino acids. Isopropyl-β-D-thiogalactopyranoside was added to 1 mM to induce expression by shaking 4-5 h at 37°C. Cell pellets were frozen in liquid nitrogen and resuspended in lysis buffer [25 mM HEPES pH 7.5, 200 mM NaCl, 100 mM Na_2SO_4, 0.1% Tween-20, 2 mM β-mercaptoethanol (β-ME) supplemented with 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor cocktail (Roche)]. The suspension was incubated on ice for 20 min with 1 mg/ml lysozyme, sonicated, and centrifuged at 14,000 rpm for 30 min to remove cell debris. RNase A (50 mg/ml) and T1 (50 U/ml) (Roche) and NaCl to 2 M were added to the cleared lysate to remove endogenous *E. coli* RNA.

Ni-NTA affinity chromatography was performed using standard procedures. Briefly, supernatant from the cell lysate was applied to Ni-NTA superflow resin (Qiagen) that had been equilibrated with buffer A+ (50 mM Tris 8.0, 2 M NaCl, 0.1% Tween-20, 2 mM β-ME, 10 mM imidazole). The resin was rinsed thoroughly with buffer A+ then buffer A (same as buffer A+ but with 250 mM NaCl and no Tween-20). A stepwise elution was performed using buffer A with increasing concentrations of imidazole. Fractions were analyzed by SDS-PAGE, pooled, and dialyzed against buffer B [40 mM Tris pH 8.0, 200 mM NaCl, 2 mM β-ME] or B+SO_4 [Buffer B with the addition of 100 mM Na_2SO_4 and 400 mM $(NH_4)_2SO_4$] at 4°C.

To remove the GB1 domain, TEV protease was added and incubated at room temperature for 1-2 h. For preparative reactions, either specific RNA or 100 mM Na_2SO and 400 mM $(NH_4)_2SO_4$ were added prior to TEV proteolysis to prevent Rev aggregation. The reaction was loaded on Ni-NTA resin equilibrated with buffer B containing 20 mM imidazole to remove the His-tagged TEV protease, the free His-GB1 tag, and any uncleaved protein. Full length Rev or Rev-RNA complexes were collected in the flow through, supplemented with 2 mM DTT, and stored at 4 degrees. All resulting proteins contain a GA dipeptide appended to the N-terminus of the Rev sequence.

Rev-RNA complex assembly and purification

Methods for purifying defined Rev complexes with RNA have been described previously (Daugherty et al., in preparation). Briefly, Rev protein and IB RNA are combined with a stoichiometry of 2:1.2 (protein:RNA) and separated on an AKTApurifier system (GE Life Sciences) with a Superdex 200 10/300 GL column equilibrated in buffer B at a flow rate of 0.5 ml/min. Only the peak corresponding to 2:1 Rev-RNA complexes was collected and concentrated for analysis.

Concentration for crystallization

Rev-RNA and apo Rev complexes were concentrated in Amicon Ultra-4 concentrators with a molecular weight cutoff $\leq$ 5000 Da at 3500 x g. Spin times were limited to no longer than 20 minutes with resuspension of the concentrated solution

before additional concentration. Failure to resuspend the solution off of the concentrator membrane, or longer spin times or use of Amicon Ultra-15 concentrators often resulted in precipitation of the protein at the surface of the membrane. For apo Rev, the imidazole and β-ME in the solution from the second Ni-NTA purification step was diluted with a greater than 4-fold dilution with crystallization buffer [40 mM Tris pH 8.0, 200 mM NaCl, 100 mM Na_2SO_4, 400 mM $(NH_4)_2SO_4$, 2 mM DTT]. Final concentrations were all ~10 mg/ml.

Crystallization

Crystallization screening was performed with the Mosquito Crystal nanoliter liquid handler (TTP Labtech) by dispensing 100 nl of concentrated Rev or Rev-RNA complex + 100 nl of a given screening condition onto a protein crystallization cover for the 96-well plate. The following 96-condition screens were utilized: Classics Suite, JCSG+ Suite, PEGs Suite, PEGs II Suite, Nucleix Suite, ProComplex Suite, and $AmSO_4$ Suite (Qiagen). Plates were stored in a temperature controlled environment at 25° C or 4° C. Optimization was performed in 24 well trays (Qiagen) with 1 ul protein (10-14 mg/ml) mixed with 1 ul well solution and allowed to equilibrate with 1 ml of well solution at 25° C. The optimal conditions for $Rev_{70\text{-}Dimer}$ crystallization were determined to be 100 mM Tris pH 8.0, 50-100 mM NaCl, 1.45-1.55 M $(NH_4)_2SO_4$, and 3% PEG 1000. Crystals generally appeared within 2-3 days and grew to maximal size by 10-14 days.

Cryoprotectant conditions

Initial cryoprotecting conditions were determined from McFerrin and Snell[4]. It was determined that well solution mixed with 100% PEG 400 to a final concentration of 18-21% was the cryoprotectant which resulted in the most consistent data sets belonging to 'Cell 1'. To avoid exposure of ammonium sulfate crystals to air, 5 ul of PEG cryoprotectant was added to the crystal drop. The resulting solution does not mix well, and thus crystals could be dragged from the original drop to bordering PEG cryoprotectant solution and rapidly looped out without being exposed to air. Crystals were frozen by plunging into liquid nitrogen.

Cryoprotection with sodium malonate was shown to require a slow transition from ammonium sulfate to sodium malonate to avoid crystal cracking. To accomplish this, a cryoprotectant solution containing 1.8 M sodium malonate and 100 mM Tris pH 8.0 (also including 500 mM NaBr and/or 250 mM Na_2SO_4 or Na_2SeO_4) was added to the crystal drop incrementally by 0.5 ul a total of 10 times. Following slow addition of cryoprotectant, crystals were transferred to a drop containing only cryoprotectant, allowed to equilibrate for 15-60 minutes and flash frozen in liquid nitrogen.

Data collection and processing

Data sets were collected at 1.1159 Å at the Advanced Light Source beamline 8.3.1 on crystals at 100 K. Diffraction data were processed and scaled with the HKL-2000 package[15]. Data quality was assessed for twinning and translational NCS using

phenix.xtriage[6]. Patterson maps and self-rotation functions were performed using the CCP4 suite[16].

References:

1. Daugherty, M.D., D'Orso, I. & Frankel, A.D. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell* **31**, 824-34 (2008).
2. Jain, C. & Belasco, J.G. Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7**, 603-614 (2001).
3. Auer, M. et al. Helix-loop-helix motif in HIV-1 Rev. *Biochemistry* **33**, 2988-96 (1994).
4. McFerrin, M.B. & Snell, E.H. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J. Appl. Cryst.* **35**, 538-545 (2002).
5. Xing, Y. & Xu, W. Crystallization of the PX domain of cytokine-independent survival kinase (CISK): improvement of crystal quality for X-ray diffraction with sodium malonate. *Acta Cryst.* **D59**, 1816-1818 (2003).
6. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).
7. Petsko, G.A. Preparation of isomorphous heavy-atom derivatives. *Methods Enzymol* **114**, 147-56 (1985).
8. Dauter, M. & Dauter, Z. Phase determination using halide ions. *Methods Mol Biol* **364**, 149-58 (2007).

9. Holyoak, T. et al. Malonate: a versatile cryoprotectant and stabilizing solution for salt-grown macromolecular crystals. *Acta Crystallogr D Biol Crystallogr* **59**, 2356-8 (2003).

10. Weiss, M.S. & Hilgenfeld, R. Dehydration leads to a phase transition in monoclinic factor XIII crystals. *Acta Crystallogr D Biol Crystallogr* **55**, 1858-62 (1999).

11. Sayer, C., Isupov, M.N. & Littlechild, J.A. Crystallization and preliminary X-ray diffraction analysis of omega-amino acid:pyruvate transaminase from Chromobacterium violaceum. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 117-9 (2007).

12. Gorynia, S. et al. Expression, purification, crystallization and preliminary X-ray analysis of the human RuvB-like protein RuvBL1. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 61-6 (2006).

13. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V. & Cullen, B.R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254-7 (1989).

14. Weber, D.J. et al. NMR docking of a substrate into the X-ray structure of staphylococcal nuclease. *Proteins* **13**, 275-87 (1992).

15. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol*. **276**, 307-326 (1997).

16. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).

Figure Legends:

Figure 1: $Rev_{70\text{-}Dimer}$ crystals. A) Early crystals grew as long thin hexagonal rods. B) Crystallization optimization resulted in more even width:length ratios.

Figure 2: Sodium malonate results in an altered unit cell. A) Part of a diffraction image clearly showing the result of using sodium malonate as a cryoprotectant on the c-axis of the unit cell. Every 4^{th} spot is very intense, suggesting either very strong psuedo translational symmetry or a unit cell with 1/4 the length of the c-axis of that predicted by including all of the spots. B) x=0 section of the native Patterson map calculated in CCP4[16] contoured every 5 sigma units between 10 and 50.

Protein	Protein Mutations	RNA Length
HG-Rev_{116}	-	-
Rev_{116}	L12S/L60R	42
Rev_{70}	L12S/L60R	42
Rev_{70}	L12S/L60R	34
Rev_{70}	L12S/L60R	36
Rev_{70}	L12S/L60R	42 (5'-OH)
Rev_{70}	L12S/L60R	-
Rev_{70}	-	-

Table 1: Protein and RNA constructs used for crystallization screening. Crystals only appeared in conditions in which the Rev_{70} protein with L12S/L60R mutations was used.

	Cell 1	**Cell 2**	**Cell 3**	**Cell 4**
Cryoprotectant	PEG 400	Many*	Sodium malonate	Sodium malonate + 0.5 M NaBr
Point group symmetry	P6	P622	P622	P622
Cell dimensions				
a, b, c (Å)	117.7, 166.4		116.8, 40.0	115.8, 81.2
α=β, γ (°)	90.0, 120.0	90.0, 120.0	90.0, 120.0	90.0, 120.0
Resolution (Å)	50.0 – 2.5	50.0 – 2.8	50.0 – 4.0	50.0 – 2.5
R_{merge}	5.7 (45.6)	6.6 (40.0)	3.2 (16.3)	6.4 (56.0)
$I / \sigma I$	20.0 (2.3)	24.5 (3.5)	68.3 (18.7)	33.1 (4.2)
Completeness (%)	100.0 (100.0)	99.9 (99.9)	98.3 (100.0)	99.9 (100.0)
Redundancy	4.7 (4.7)	14.4 (12.4)	9.5 (10.2)	10.8 (9.9)

Table 2: X-ray diffraction data and processing statistics showing unit cell and point group symmetry dependence on cryoprotectant. * Multiple conditions produce Cell 2, including glycerol, long exposures to PEG 400 and PEG 400 with 0.5 M NaBr.

A

B

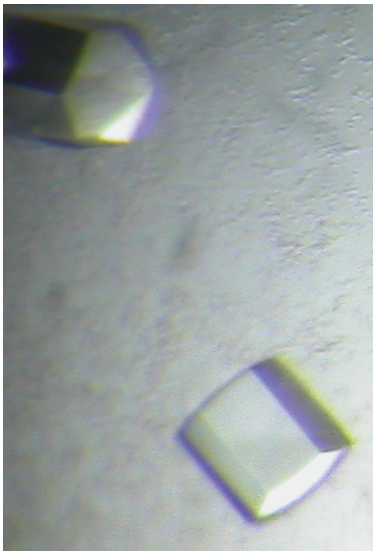
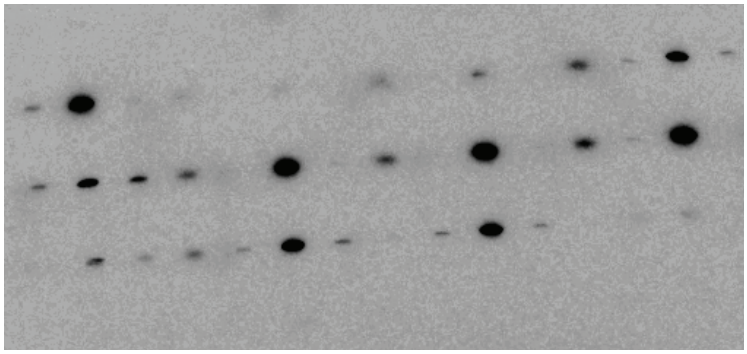


Figure 1

A



B

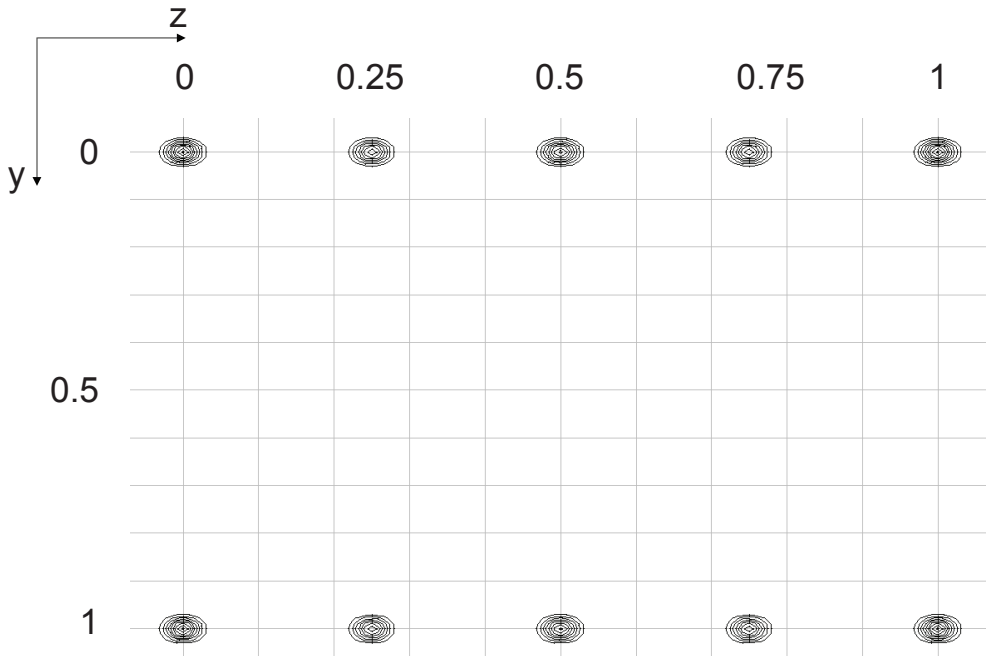


Figure 2

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

September 3 2009

Date