Structural Studies of a Subunit of the Murine Cytomegalovirus Nuclear Egress Complex

Citation

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:14226065

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Structural Studies of a Subunit of the Murine Cytomegalovirus Nuclear Egress Complex

A dissertation presented

by

Kendra Elizabeth Leigh

to

The Division of Medical Sciences

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

December 2014
Abstract

The *Herpesviridae* family of viruses includes a number of human pathogens of clinical importance. Like other herpesviruses, cytomegaloviruses require a heterodimeric nuclear egress complex (NEC) consisting of a membrane-bound protein and a soluble nucleoplasmic protein, termed in murine cytomegalovirus (MCMV) M50 and M53, respectively. Genetic, electron microscopic, and immunocytochemical studies have revealed the importance of this complex for viral replication, most predominantly in facilitating egress of viral nucleocapsids across the nuclear membrane. Despite the significance of the NEC to the herpesvirus life cycle, there is a dearth of structural information regarding the components of the complex. We present here an NMR-determined solution-state structure of the conserved, structured, soluble portion of M50 (residues 1-168), which exhibits novel structural character. We mapped the binding site of a highly conserved minimal binding domain of the M53 homologue from human cytomegalovirus (HCMV; UL53) required for heterodimerization onto the structure and identified specific residues in a groove within the M50 protein fold that interact with the UL53 peptide. This site was verified biophysically and biologically: single amino acid substitutions of the corresponding residues of the homologous protein from HCMV (UL50) resulted in decreased UL53 binding *in vitro*, as measured by isothermal titration calorimetry, and substitutions that had the greatest
effect on binding affinity caused disruption of UL50-UL53 co-localization and lethal defects in the context of HCMV infection. We then compared the effect of binding UL53 peptide with binding of the larger natural binding partner, M53 (residues 103-333) via NMR, with the results suggesting that conformational changes most likely occur on a fold-wide level in the context of the full complex. We suggest that these findings combined with the clinical relevance, the virus-specific aspects of nuclear egress, and the novelty of the structure make the HCMV NEC an attractive potential drug target. To this end, we used in silico screening to identify possible small molecule inhibitors and have begun validating top screen hits biophysically and biologically.
# Table of Contents

Abstract .......................................................................................................................................... iii

Table of Contents ............................................................................................................................ v

List of Figures ............................................................................................................................... vii

List of Tables .................................................................................................................................. ix

Acknowledgements ......................................................................................................................... x

I. INTRODUCTION TO HERPESVIRAL NUCLEAR EGRESS ......................................................... 1

II. NMR METHODS .......................................................................................................................... 15

   II.i Protein Purification ............................................................................................................... 16

   II.ii Selective Isotopic Labelling ............................................................................................... 19

   II.iii Non-uniform Sampling and 4D $^{13}$C-HMQC-NOESY-HMQC ............................................ 19

   II.iv NMR Structure Calculation ............................................................................................. 21

   II.v NMR-based Titration Experiments ..................................................................................... 24

   II.vi T1, T2, and Hetero-NOE Experiments ............................................................................. 25

   II.vii Residual Dipolar Coupling Experiments ......................................................................... 26

III. SOLUTION-STATE NMR STRUCTURE OF M50 .................................................................... 27

   III.i Solution-state NMR Structure of M50 (residues 1-168) ..................................................... 29

   III.ii Novelty of the M50 (residues 1-168) Protein Fold ............................................................. 33

   III.iii Dynamics of the M50 (residues 1-168) Structure ............................................................. 37
III.iv Threading of Other Sequences onto the M50 (residues 1-168) Structure .................. 40

III.v Residual Dipolar Couplings, and M50 (residues 1-168) Structure Validation .......... 45

IV. HETERODIMERIZATION INTERFACE .......................................................................... 46

IV.i Initial Characterization of the Binding Interface ................................................... 48

IV.ii Cross-Saturation Transfer Experiments with UL53 Peptide ................................. 56

IV.iii Purification of M53 (residues 103-333) ............................................................... 58

for any NMR-based experiments that might be conducted in the future using M53 (103-333)
alone. ..................................................................................................................................... 60

IV.iv Titration Experiments with M53 (residues 103-333) ......................................... 60

V. SMALL MOLECULE INHIBITION OF HETERODIMERIZATION ........................ 63

VI. DISCUSSION .............................................................................................................. 74

References .......................................................................................................................... 83
List of Figures

Chapter I

Figure I-1: Schematic depicting herpesviral nuclear egress ........................................................... 3

Figure I-2: Key residues identified in the literature of M50 and its human herpesviral homologues ................................................................................................................................... 10

Chapter III

Figure III-1: CYANA M50 (residues 1-168) structure calculation bundle and cartoon schematic of the lowest energy structure ....................................................................................................... 30

Figure III-2: Electrostatic surface of the lowest energy CYANA-calculated structure of M50 (residues 1-168) ..................................................................................................................................... 32

Figure III-3: Schematic of the topology of M50 (residues 1-168) ................................................ 35

Figure III-4: T1, T2, and Hetero-NOE values graphed by residue .............................................. 39

Figure III-5: Sequence alignment of the human herpesvirus homologues with MCMV M50 .... 42

Figure III-6: Homologous residues from UL50 and UL34 threaded onto the M50 (residues 1-168) structure ................................................................................................................................ 44

Chapter IV

Figure IV-1: HSQC spectra of M50 (residues 1-168) with and without UL53 peptide and bar graph of intensity ratios by residue ........................................................................................................ 49

Figure IV-2: Residues affected by UL53 peptide binding mapped onto the M50 (residues 1-168) structure ...................................................................................................................................... 51

Figure IV-3: Heat map of ITC K_D values mapped onto the M50 (residues 1-168) structure ...... 53
Figure IV-4: Residues affected by UL53 peptide binding in CST experiments mapped onto the M50 (residues 1-168) structure ................................................................. 57
Figure IV-5: TROSY-selected $^{15}$N-HSQC of M53 (residues 103-333) ........................................ 59
Figure IV-6: Residues affected by titration of M53 (residues 103-333) in comparison to residues affected by titration of UL53 peptide ................................................................. 61

Chapter V

Figure V-1: Chemical structures of the purchasable top-hits from in silico screen of the ZINC database against the M50 (residues 1-168) fold ................................................................. 67
Figure V-2: HSQC spectra of M50 (residues 1-168) with and without addition of ZINC38147030 .................................................................................................................. 69
Figure V-3: HSQC spectra of M50 (residues 1-168) with and without addition of ZINC01192203 .................................................................................................................. 70
Figure V-4: M50 (residues 1-168) residues affected by addition of small molecules identified by in silico screening of the ZINC database ................................................................. 71
Figure V-5: Effect of in silico small molecule inhibitors on HCMV viral titer ............................... 72
List of Tables

Chapter I
Table I-1: Nomenclature for the nucleoplasmic and membrane-bound components of the nuclear egress complex in different herpesviruses. ................................................................. 4

Chapter III
Table III-1: NMR refinement statistics for the M50 (residues 1-168) CYANA structure calculation ..................................................................................................................................... 31
Table III-2: Top hits, scores, and secondary structure analysis from structural comparison databases ................................................................................................. 36

Chapter IV
Table IV-1: Effect of alanine mutations on binding affinity of M50 (residues 1-168) and UL50 (residues 1-169) for their heterodimeric binding partners and on HCMV replication .......... 52

Chapter V
Table V-1: Binding free energy values of the purchasable top-hits from in silico screen of the ZINC database against the M50 (residues 1-168) fold ................................................................. 68
Acknowledgements

I would like to thank first and foremost my advisor, Jim Hogle, for all the support he has given me over the course of my graduate school career. He has been much more than the principal investigator of the lab I have worked in; he has been a true mentor and inspiration in my development as a scientist. I would also like to thank all the members of the Hogle Lab, both past and present, with particular shout-outs to Hazel Levy who first taught me how to purify poliovirus and whose mentorship during my summer rotation was one of the reasons I chose to join Jim’s lab for my graduate work; to My D. Sam Mansueto who is a dear friend and whose work laid the foundation for the project that has become the subject of this dissertation; to Dave Filman who helped me with the Phyre2 calculations in this study, to Molly Coseno and Mike Strauss who have been nothing but encouraging throughout the years; and to Lise Schotte on whose birthday I will be defending this dissertation and whose sweet words of encouragement gave me a much needed boost in the months immediately before the date.

I would like to thank the Division of Medical Sciences Virology Program, including the heads and deputy-heads of the program, David Knipe, Dana Gabuzda, and Sean Whelan, and all the faculty members that make up the Harvard virology community for fostering such a vibrant scientific and educational environment and for allowing me to be a part of it for over six years. I would like to thank all the members of my cohort (Jenny Lee, Kevin McCarthy, Eileen Sun, Timothy Soh, Silvia Piccinotti, Sharon Lim, Marcus Tan, Lauren Thorpe, Laura Ellis, Alvin Tan, and Jeffery Teigler) for the fun and the friendship they have provided over the years. Thank you to my Prequalifying Exam Committee (Mike Farzan, Xinzhen Yang, and Eric Johannsen) for one of the most nerve-wracking experiences of my life before officially clearing me to be a Ph.D. candidate. Thank you to my Dissertation Advisory Committee (Max Nibert as chair, Steve
Harrison, Andres Leischziner, and Gerhard Wagner) for all the advice and guidance you have given me over the years. Last, but most certainly not least, thank you to James Chou, David Knipe, and Celia Schiffer for agreeing to serve as my dissertation defence examiners and for their never-ending patience and enthusiasm during all the stages of the defence process.

I would like to acknowledge the incredible debt that I owe to our collaborators, Don Coen, Gerhard Wagner, and their respective labs. My work would not have been possible without their help. From double checking my slides before my first international talk to inviting me to group meeting so that I could learn more about a virus quite different from the one I had initially prepared to work with, Don did everything he could to support me as a graduate student and for that I am incredibly grateful. I would also like to thank Mayuri Sharma, from the Coen Lab, for all her work with human cytomegalovirus that helped to support and build the conclusions of our structural studies. I am incredibly thankful to Gerhard and members of the Wagner Lab for all the training they have given me in the field of NMR, for all the company they have given me during the late nights in the lab, and for all the friendship they have offered to a younger scientist. Special thanks go to Jeffrey Hoch for his initial help with the 4D spectrum and Sven Hyberts for all his help with the final hmsIST reconstruction. I also owe thanks to Remy Sounier, formally of the Chou Lab, for the help he gave us with the RDC calculations. Learning NMR from Gerhard and other members of the Harvard NMR community by auditing BCMP228 for the first of three times back as a third-year grad student changed my life and from there I never looked back. In no order of importance, I would like to give particular thanks to Dharma Tsyry-Yan Yu, his wife Mafina Wu, and their son Shanti; Ricard Rodriguez-Mias and his wife, Judit Villen; Patrick Hagner; Mélissa Léger Abraham and her husband, Jonathan Abraham; Scott Robson; Rafael Luna; Paul Coote; Naotaka Sekiyama and his wife, Mariko Satou Sekiyama;
Andras Boeszoermenyi; Thomas Raschle; Elizabeth O’Day; Maura Kilcommons; and Gregory Heffron. Finally, I would like to give my deepest and most heartfelt thanks to Haribabu Arthanari. He has been my teacher, my collaborator, my friend, and my inspiration. I literally could not have done it without him.

I would like to thank all my former mentors who have helped me get to this point and all the friends who have helped to remind me that there is more to life than lab. Thank you to my high school biology teacher, Angelique Bosse, and my first scientific mentor, Maribeth Eiden. Thank you to my mentors at the University of Cambridge: Sarah Lummis, my Director of Studies, Michael Powell, my Part II project mentor, and T. David K. Brown, the man who always encouraged my curiosity. Thank you to T.M. Krishna, with whom I have had many spirited debates. His questions always challenged me to think deeply, to explain simply, and to explore widely. Finally, thank you to John Chai, Randy Li, Michael Forbes, Patrick Detzner, and Doug Adams, gentlemen I have known since I was eleven years old, and to Joann Lin, our honorary Blazer. They helped welcome me to back to America, dragged me out of lab when needed, always provided a friendly ear, and have been about the best friends a girl could have.

I have wanted to be a virologist since I was twelve years old and since that declaration my mother and father have done everything they can to support me in that dream. It was my father who first read Principles of Virology with me from cover to cover. When following my heart meant applying to university in an entirely different country for my undergraduate education, it was my parents who did everything to support my unusual choice, and when I was accepted to Cambridge, it was my parents who let their only child move three thousand miles away with their blessing. While a graduate student at Harvard, it was my father who sent me any book I could ever need to build a strong background in my fields of interest and it was my mother who
made sure I didn’t starve years before I ever got to the point of defending a dissertation. For all their love and all their help, to them I simply say “thank you.”
I dedicate this dissertation to

Shanti

in the hope that he too will grow to follow his dreams.
I. INTRODUCTION TO HERPESVIRAL NUCLEAR EGRESS
The family *Herpesviridae* of the order *Herpesvirales* (1) contains three subfamilies: alpha-, beta- and gammaherpesviruses, and included amongst these subfamilies are a number of clinically significant human pathogens and veterinary infectious agents (2). Human cytomegalovirus (HCMV) and its related murine cytomegalovirus (MCMV) are members of the *Betaherpesvirinae* subfamily (1). While most of the population has been exposed to HCMV and the majority become seropositive by the age of 40, few will experience any symptoms upon infection; however, in the case of immunocompromised individuals and newborns, infection can mean much more serious outcomes (3). Despite its medical relevance, there are few therapeutic options available for HCMV infection and no licensed drug for the treatment of congenital HCMV (3,4).

During the lytic part of the herpesvirus lifecycle, the viral genome is replicated in the nucleus and packaged into intact protein shells in an ATP-dependent manner, at which point the resulting nucleocapsids (also known as C-type capsids) are too large to be accommodated by the nuclear pore complex (5,6) and must exit the nucleus through a process known as nuclear egress (Figure I-1). A combination of viral and cellular factors breaks down the inner nuclear lamina, allowing the capsid access to the inner nuclear membrane (INM) and resulting in membrane invagination. This envelopment of the nucleocapsid progresses and results in the formation of an intact vesicle in the perinuclear space, which then subsequently fuses with the outer nuclear membrane, thus releasing the immature particle for further maturation in the cytoplasm and trans-Golgi network (2,6-12).

This mechanism of envelopment-de-envelopment is orchestrated by a heterodimer, conserved across herpesviruses, known as the nuclear egress complex (NEC) (Table I-1). This complex has been implicated in the recruitment of one or more protein kinases to disrupt the nuclear lamina,
(A) Capsids are assembled and loaded with viral genome in an ATP-dependent manner (red hexagons).

(B) Nucleocapsids are too large to exit via the nuclear pore. Instead they interact with the nuclear egress complex (NEC) at an area cleared of nuclear lamins (green), resulting in an invagination of the inner nuclear membrane. In the case of MCMV, the NEC is composed of two proteins: the membrane-bound M50 (blue) and the nucleoplasmic M53 (orange).

(C) The membrane invagination continues until membrane scission occurs, resulting in an intact vesicle in the intermembrane space.

(D) The vesicle fuses with the outer nuclear membrane, releasing the nucleocapsid into the cytoplasm where it goes on to acquire tegument proteins and undergo secondary envelopment in the trans-Golgi network.

Figure I-1: Schematic depicting herpesviral nuclear egress
Table I-1: Nomenclature for the nucleoplasmic and membrane-bound components of the nuclear egress complex in different herpesviruses.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Nucleoplasmic Component</th>
<th>Inner Nuclear Membrane-bound Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human herpesvirus 1 (Herpes simplex virus 1; HSV-1) (13)</td>
<td>UL31</td>
<td>UL34</td>
</tr>
<tr>
<td>Human herpesvirus 4 (Epstein-Barr virus; EBV) (13,14)</td>
<td>BFLF2</td>
<td>BFRF1</td>
</tr>
<tr>
<td>Human herpesvirus 5 (Human cytomegalovirus; HCMV) (13,15)</td>
<td>UL53</td>
<td>UL50</td>
</tr>
<tr>
<td>Murine cytomegalovirus (MCMV) (13,16)</td>
<td>M53</td>
<td>M50</td>
</tr>
<tr>
<td>Suid herpesvirus 1 (Pseudorabies virus; PrV) (13,17,18)</td>
<td>UL31</td>
<td>UL34</td>
</tr>
<tr>
<td>Human herpesvirus 8 (Kaposi’s sarcoma virus; KSHV) (14)</td>
<td>ORF69</td>
<td>ORF67</td>
</tr>
</tbody>
</table>
permitting access of nucleocapsids to the inner nuclear membrane (8,13,16,19-25), although immunoprecipitation assays have suggested that direct binding of lamin A/C by the NEC, at least by the nucleoplasmic component, is not likely (26). Furthermore it has been shown in several cases that expression of the NEC components, even in the absence of viral infection, can result in invaginations of the inner nuclear membrane (14,27-29) resembling those seen in the nucleus during active infection. In one recent study, purified HSV-1 NEC alone resulted in vesicle formation in an in vitro membrane context suggesting the complex is independently capable of membrane scission (27), although other studies in Epstein-Barr virus suggest that ESCRT machinery may be necessary for successful egress (30) and mass spec analysis of proteins that co-precipitated with M50/M53 identified proteins involved in vesicular trafficking such as endophilin-A2 (31).

Similarly to the homologous heterodimers found in other herpesviruses (13), in HCMV, the NEC is comprised of a type II transmembrane protein that localizes to the inner nuclear membrane and a nucleoplasmic phosphoprotein binding partner, termed UL50 and UL53 respectively (15,23,32,33). These proteins are homologous to MCMV proteins M50 and M53 (13,16,34). Both the UL50/UL53 and M50/M53 heterodimers have been shown to be essential for HCMV/MCMV replication and efficient nuclear egress of assembled nucleocapsids into the cytoplasm (8,16,19,33-38). Experiments in pseudorabies virus using immunogold staining and transmission electron microscopy show the presence of the NEC components UL31 and UL34 along the nuclear membrane and in the membrane of virions in the intermembrane space, but not in virions found in the cytoplasm or extracellularly (17,18), suggesting these proteins are synthesized de novo during viral infection and not brought into the cell via the viral envelope or tegument. This is confirmed by the absence of UL31 and UL34 in cytosolic HSV-1 particles in
other studies (39). In contrast, UL53 has been detected in mature HCMV virions, again by immunoelectron microscopy (40), suggesting some differences in the global role of NEC components between the subfamilies. It should also be noted that while important, in alpha- and gammaherpesviruses, the need for the NEC is not absolute. In HSV-1 and pseudorabies virus, for example, evidence of very low levels of infectious particle production may still be detected in the absence of UL34 (M50 homologue) or UL31 (M53 homologue), although transmission electron microscopy shows a distinct lack of enveloped particles in the cytoplasm or on the cell surface (17,18,41,42). Similar results are seen in Epstein-Barr virus (43).

There is evidence to suggest that the NEC may be involved in more during nuclear egress than just orchestration of the primary envelopment stages at the inner nuclear membrane. Early studies showed that deletion of UL31 from HSV-1 resulted in decreased detection of total viral DNA, decreased cleavage of concatameric DNA into monomers, lack of enveloped virions in the cytoplasm or extracellular space, and decreased viral yield (42) leading to the hypothesis that UL31 is involved in viral DNA processing and packaging, leading to overall egress defects.

Immunoprecipitation studies for GST-tagged UL34 in cell lysates have demonstrated an interaction between UL34, UL31, and ICP5, also known as VP5 or major capsid protein, in HSV-1 (44). Budding at the inner nuclear membrane is known to be biased towards capsids containing viral DNA (C-type capsids) (45), but VP5 is found in A, B, and C-type capsid forms (46) and therefore cannot be the exclusive determinant of this selection preference. Immunofluorescence shows co-localization of UL47, a major tegument protein, with UL31, UL34, and US3 and deletion of UL47 shows capsid accumulation in the nucleus as seen by transmission electron microscopy (47), suggesting that the viral protein may play a role in capsid recruitment and initial stages of primary envelopment. Other immunoprecipitation studies show
an interaction between UL31 of HSV-1 and two proteins enriched on, but not exclusive to, C-type capsids: UL17 and UL25. This interaction was shown to require UL25, was improved by, but did not require UL17, and immunogold electron microscopy demonstrated a UL31 association at multiple sites on the capsid (26,45). This data in combination with previous work showing that deletion of UL17 or UL25 result in viral DNA cleavage and packaging defects (48,49), led the authors to hypothesize a natural progression during viral assembly and egress from UL17-UL25 complex enrichment on B-type capsids facilitating DNA encapsidation and formation of C-type capsids, followed by recruitment to the NEC via an interaction with UL31. Yet these conclusions are further confounded by results from pesudorabies virus where in the absence of UL34, UL31 can be co-purified with capsid from cells infected with pseudorabies virus, and this interaction was shown to be independent of UL25, UL6 and UL33 (50). Clarifying the capsid-NEC interaction whether it is general to all herpesviruses or virus specific will require more detailed molecular determinations.

While invagination of the membrane and formation of the vesicle in the intermembrane space, also known as primary envelopment of the virion, may at a minimum need nothing more than the NEC itself (14,27-29), studies have implicated more than one protein in facilitating the subsequent fusion of this temporary envelope to the outer nuclear membrane (11). Studies using immunoelectron microscopy have shown that deletion of gB and gH from HSV-1, both viral glycoproteins also involved in entry, results in accumulation of enveloped particles in the intermembrane space; however, deletion of only one of the glycoproteins (gB or gH) showed little to no effect on nuclear egress of particles suggesting redundancy of their role (51). Studies have also implicated viral kinase US3 phosphorylation of gB in fusion with the outer nuclear membrane (52), and disrupting phosphorylation of UL31 by US3 also results in accumulation of
particles in the intermembrane space. On the other hand, pseudophosphorylation of UL31 results in a defect in envelopment at the inner nuclear membrane rather than fusion at the outer nuclear membrane (53). Taken together these data suggest that phosphorylation events may serve as some sort of trigger for fusion of the viral vesicles with the outer nuclear membrane. Studies have also implicated US3 in phosphorylation of UL34 (54-56); however, other evidence seems to suggest that the US3-UL34 interaction is not necessary for correct viral morphogenesis (57,58). Confusingly, other studies in an HSV-1-related alphaherpesvirus, pseudorabies virus, found no effect on nuclear egress from deletion of viral glycoproteins gB and gH (59) and gB is not detected in primary enveloped virions by immunoelectron microscopy (57). Transmission electron microscopy of cells infected with a HSV-1 VP16 null mutant also shows accumulation of capsid in the intermembrane space. Although there are a number of other global defects, including overall lower capsid numbers, the data does argue for at least a minor role of VP16 in facilitating de-envelopment at the outer nuclear membrane (60). There is also some evidence for negative regulation of outer nuclear membrane fusion. For example, overexpression of gK during HSV-1 infection, results in accumulation of capsids in the intermembrane space (61). It should further be noted that betaherpesviruses such as HCMV do not code for US3. If phosphorylation does serve as some sort of trigger for outer nuclear membrane fusion, then another kinase will have to substitute functionality. Possible substitutes are UL97 or protein kinase C, which have been implicated in other steps of nuclear egress (15,16,22,32,62,63). These differing results, even among herpesviruses of the same subfamily, indicate that the mechanisms of this stage of nuclear egress may be less well conserved than those centered on the NEC and the inner nuclear membrane.
The proteins composing the NEC have also been implicated in stages of the herpesviral lifecycle beyond that of egress from the nucleus. Immunoprecipitation studies have demonstrated an interaction (possibly indirect) between UL34 of HSV-1 and the neuronal isoform of the intermediate chain of cytoplasmic dynein. This data led Ye et al. to speculate that UL34 plays a role in retrograde transport of the capsid-tegument complex to the nucleus for the introduction of the viral genome through the nuclear pore (44). A Y68A mutation in HSV-1 UL34 has shown effects on cell-to-cell spread and trafficking of viral glycoprotein gE in addition to those on nuclear egress (64), and another study implicated the N-terminal domain of M50 in interaction with and downregulation of IRE1, which is part of the cellular unfolded protein response (65). Viral proteins are often multifunctional and these alternative roles for the NEC components during the viral lifecycle are in keeping with that principle.

M50, which is the primary focus of the structural studies described here, is an approximately 35 kDa protein (66,67) that shows strong conservation amongst herpesvirus homologues in the N-terminal half of the protein (34,68) and has a predicted C-terminal transmembrane domain (16). Its closest human herpesviral homologue is UL50 from HCMV (Table I-1, Figure I-2). Several studies in different herpesviruses have already shown the N-terminal domain of the membrane-bound NEC component to be responsible for the interaction with its nucleoplasmic partner (18,34,37,68,69). While loss of or mutational insertions into the putative transmembrane domain has been shown to destroy functionality (29,34) and targeting to the nuclear rim (15), it has further been shown in other studies that replacement of the transmembrane domain with a similar transmembrane domain from a cellular inner nuclear membrane protein does not interfere with NEC component function (70,71).
Figure I-2: Key residues identified in the literature of M50 and its human herpesviral homologues

The alignment of the human herpesvirus homologues with the MCMV M50 sequence (bold) was generated using Clustal Omega (72). * (asterisk) indicates positions that have an identical residue; : (colon) indicates conservation between groups of strongly similar properties – scoring $>0.5$ in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties – scoring $\leq 0.5$ in the Gonnet PAM 250 matrix. Residues appearing in the literature as having significance in nuclear egress are highlighted in red (20,24,27,29,31,34,37,56,58,64,68,69,73,74). The underlined portion is a minimal fragment identified as capable of pulling down UL31 in an immunoprecipitation assay (75). The predicted transmembrane (TM) domain of M50 is highlighted by the grey box.
Figure I-2 (continued).
In vitro work with HSV-1 UL34 suggests that the tail anchor of the protein is inserted into the membrane post-translationally. In the absence of UL31, UL34 shows a perinuclear ER-like distribution as well as a nuclear rim presence in immunofluorescence studies; however, co-expression of UL31 shifts the distribution almost exclusively to the nuclear rim (71), suggesting that somehow UL31 is able to bias targeting of its membrane-bound partner. PSort identifies a putative nuclear localization signal (NLS) in HSV-1 UL34 (76), and NLStradamus predicts motifs in the MCMV, Human herpesvirus-7, and Epstein-Barr virus homologues of UL34 (77), but such a motif is lacking in the pseudorabies virus homologue (76), nor has such a motif been identified in UL50 (77). NLSs have been identified in MCMV M53 in the region from residue 16 to 106 (36), in HCMV UL53 in the region from residue 18 to 27 (77), and in pseudorabies virus UL31 in the region from residue 5 to 20 (78). In the case of pseudorabies virus, mutation of the putative NLS resulted in localization of the mutant UL31 exclusively to the cytoplasm, suggesting the existence of a nuclear export signal (NES). An NES was identified computationally from residues 246 to 254 and when combined with mutation of the NLS, resulted in a diffuse cellular distribution of UL31. Interestingly, no NES motif could be identified computationally in the homologues from HSV, HCMV, or MCMV (78). Early in infection, MCMV UL50 shows a punctate cytoplasmic distribution, which rapidly transitions to a nuclear rim distribution as infection progresses. In contrast, the binding partner, UL53 builds a nucleoplasmic presence before being recruited to the nuclear rim by UL50. Experiments in which the molecular weight of UL50 is artificially increased show a negative correlation between size and nuclear recruitment. This result combined with lack of computational identification of any canonical NLS or inner nuclear membrane sorting motifs led Schmeiser et al. to speculate that UL50 traffics to the nucleus via membrane-bound diffusion and that binding
to UL53, which has more canonical targeting motifs, helps with retention of the membrane protein on the nuclear envelope (77). Roughly similar conclusions were drawn by Yamauchi et al. based on their observations of the cellular distributions of homologues HSV-2 UL31 and UL34 by immunofluorescence (79).

Most of the work in identification of key regions and residues in M50 and its homologues has been through the use of transposon mutagenesis or the generation of point or truncation mutants. The generated mutants are then often evaluated for one or more of the following: effect on viral growth through titering of supernatants, heterodimer formation through co-immunoprecipitation, cellular localization using immunofluorescence or transmission electron microscopy, and complementation through exogenous expression of viruses lacking the protein in question. Residues or regions identified as having an impact on nuclear egress are identified in Figure I-2. Of the studies referenced, the one done by Bubeck et al. is the most comprehensive and made use of random insertion of a five residue transposon in order to identify important areas of M50, primarily through a complementation assay. As is evident from the figure, many of the residues and regions identified as important fall in the first half of the protein, which also shows the highest sequence conservation across the herpesviruses. The other regions appearing in the literature are the transmembrane domain, whose importance has already been discussed, and the region from approximately residue 175 to residue 250 in M50. This region contains a polyproline-rich sequence in the betaherpesviruses, which has been shown to be important for viral replication (34) and possible binding to other cellular factors (31), as well as possible phosphorylation sites (56,62). A minimal binding domain was identified in HSV-1 (underlined in Figure I-2), which is sufficient to immunoprecipitate UL31 (75); however, studies in other
homologues such as HCMV UL50 have shown that the “minimal” domain encompasses as much as the entire N-terminal half of the protein (68,76).

While originally thought to be a mechanism exclusive to herpesviruses, more recently a similar nuclear egress process was described for export of large ribonucleoprotein particles from the nucleus into the cytoplasm in *Drosophila* (80). However, so far, no orchestrating host cell homologue of the NEC heterodimer has been identified, although TorsinA, the AAA⁺-ATPase, has been implicated as a major player in the cellular process (81). Interesting overexpression of TorsinA during HSV-1 infection leads to the accumulation of membrane-bound capsid-like structures that co-localize with UL34 and impaired virus production (82), whereas HSV-1 infection in *Tor1a-/-* mouse embryonic fibroblasts (MEFs) also shows reduced growth characteristics and total nuclear envelope breakdown (83), suggesting that HSV-1 relies on some cellular process that at least indirectly involves TorsinA at endogenous levels. Despite the new cellular parallels, sequence analysis of NEC homologues has not yielded any regions predicted to be homologous to protein-protein interaction domains (73) and lack of any structural information for any component of the NEC across the herpesvirus family means that molecular understanding of nuclear egress mechanism is still limited.
II. NMR METHODS
II.i Protein Purification

Purified protein used in the structural studies described in this dissertation was obtained by recombinant expression in BL21-CodonPlus(DE3)-RP *Escherichia coli* (Stratagene). The plasmids used for this expression were created similarly to the methods detailed in Sam *et al.* (84). Platinum Pfx DNA Polymerase (Invitrogen) was used to amplify the open reading frames (ORFs) coding for UL50 residues 1 to 349 and UL53 residues 1 to 376 from HCMV strain AD169 and for M50 residues 1 to 316 and M53 residues 1 to 333 from MCMV strain Smith by PCR. The reaction product was cloned into an IMPACT-CN pTYB12 expression vector (New England Biolabs, Inc.). For the purposes of affinity purification, an open reading frame that fused the N-terminus of each target protein to the C-terminus of the intein expression tag was constructed using NdeI and EcoRI restriction enzyme cleavage sites. The QuikChange protocol (Stratagene) was used to introduce internal stop codons in the ORFs of UL50, M50 to create C-terminal truncations at residues 169 and 168 respectively. Similar methods were used to create the truncated M53 construct encompassing residues 103-333. A construct with a pGEX-6P-1 vector backbone expressing a truncated version of UL53 (residues 50 to 292) was engineered with PCR product amplified from HCMV strain AD169 using the BamHI and EcoRI restriction enzyme cut sites. The QuikChange protocol (Stratagene) was also used to introduce the changes necessary for expression of single amino acid alanine mutants. Nucleic acid sequences of all the constructs were verified using the Dana Farber Cancer Institute, Molecular Biology Core Facility.

The protein expression constructs were introduced to the expression *E. coli* via heat-shock-induced transformation. After recovery following heat-shock, the initial transformation culture was used to inoculate a 30-50 mL lysogeny broth (LB) (85,86) starter culture, which was then
used to inoculate a larger culture for bulk expression. For expression of unlabelled proteins (proteins with natural isotopic abundance), this larger culture was 4 to 6 litres of LB medium. For expression of isotopically-enriched, uniformly-labelled proteins, this larger culture was 2 litres of M9 (minimal) medium containing 1 g $^{15}$N-NH$_4$Cl and 2 g $^{13}$C-glucose. The M9 was made up in D$_2$O and 2 g $^2$H,$^{13}$C-glucose was substituted for expression of perdeuterated proteins. All expression cultures were grown with 50 µg/mL carbenicillin as a selection agent. The culture was allowed to grow at 37°C, at a shake rate of approximately 200 RPM, until reaching an optical density at 600 nm of approximately 0.8. At this point, protein overexpression was induced with addition of 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the culture was allowed to incubate for a further 16 to 24 hours at 16°C.

The protocol for purification of the protein from culture is similar to that described in Sam et al. (84). Cells were spun down and lysed at 4°C using a sonicator (total pulse time approximately five minutes) in cold lysis buffer consisting of 0.5 M NaCl, 50 mM Tris, pH 8, 0.005% NaN$_3$, and protease inhibitor cocktail (Roche). From this point on, protein purification was done entirely at 4°C using pre-chilled buffers. 20 mL of chitin resin (New England Biolabs, Inc.) in a gravity column was prepared by washing with approximately 20 column volumes (400 mL) of ddH$_2$O followed by equilibration with approximately 20 column volumes (400 mL) wash buffer (1 M NaCl and 25 mM Tris, pH 8), before being allowed to chill to 4°C. Lysate was clarified by centrifugation at approximately 35000 x g for 50 minutes and the resulting supernatant is loaded onto the gravity column. The resin was mixed well with the supernatant and allowed to settle before allowing the supernatant to flow through. Unbound protein was washed from the column by allowing flow through of 20 column volumes (400 mL) or more of wash buffer. After washing, 3 column volumes (60 mL) of fresh cleavage buffer (0.5 M NaCl, 25 mM Tris, pH 8,
and 50 mM dithiothreitol (DTT)) was mixed well with the resin and allowed to incubate for 50 to 60 hours at 4°C to allow the self-cleavage reaction of the intein tag to occur. An optional additional 3 column volumes (60 mL) of cleavage buffer could be added half-way through the incubation period to refresh the reducing potential of the buffer. After the incubation period, the cleaved protein was allowed to elute and an additional 3 column volumes of cleavage buffer lacking DTT was used to wash the resin. The pooled eluate was then concentrated using Amicon® Ultra-15 Centrifugal Filter Units (EMD Millipore) with a 10,000 nominal molecular weight limit (NMWL) to a total volume of 10 mL. The concentrated eluate was loaded onto a Superdex 200 10/300 GL size-exclusion column (GE Healthcare Life Sciences) for final purification and exchange into the final NMR- sample buffer (25 mM NaPO₄, pH 6.5, 150 mM NaCl, 1 mM DTT) or isothermal titration calorimetry (ITC)-sample buffer (200 mM NaCl, 25 mM HEPES, pH 7.0, 1 mM TCEP (Tris [2-carboxyethyl] phosphine)). Fractions containing the protein were pooled and concentrated. Final protein concentration was calculated from the optical density absorbance at 280 nm measured by NanoDrop UV/Vis spectrophotometer (ThermoFisher Scientific, Inc.) and the extinction coefficient using the Beer-Lambert law:

\[ A_{280} = \varepsilon c \ell \]

where \( A_{280} \) is the absorbance as measured at 280 nm, \( \varepsilon \) is the extinction coefficient of the protein being measured (units M⁻¹·cm⁻¹), \( c \) is the molar concentration of the protein, and \( \ell \) is the path length of the sample in cm. Extinction coefficients for M50 (1-168) and M53 (103-333) were calculated from the amino acid sequence using ProtParam (66), taking into account the additional Ala-Gly-His upstream of the N-terminal methionine residue, resulting from cleavage of the intein expression tag.
II.ii Selective Isotopic Labelling

The addition of $^{15}$N-NH$_4$Cl and $^{13}$C-glucose to the bacterial growth medium results in isotopic enrichment of the resulting sample with nuclei of spin-1/2 (preferred for NMR experiments) and production of protein that is “uniformly labelled.” In cases of large proteins or poor spectral quality due to peak overlap, using amino acid precursors that have been isotopically enriched with spin-active nuclei allows the selective enrichment of specific amino acids or even specific moieties (such as terminal methyl groups) within those amino acids in the final protein. This effectively reduces the number of resonances observed and improves relaxation properties.

ILV(Ile-Leu-Val)-methyl-labeled samples require the introduction of $\alpha$-ketobutyrate and $\alpha$-ketoisovalerate precursors into D$_2$O growth media as described in Gardner et al. (88) and Rosen et al. (89). This results in a final purified protein product that is protonated wherever there are labile protons (the backbone amides) and as well as in the terminal methyl groups of only the Ile, Leu, and Val residues. The remainder of the sidechains become “invisible” in the spectrometer due to the presence of deuterons ($^2$H) in place of the protons ($^1$H). For the purposes of the experiments described here, $^{15}$N,$^{13}$C-perdeuterated-ILV-labelled M50 (residues 1-168) samples were prepared and used for the acquisition of a 3D $^{13}$C,$^{15}$N time-shared NOESY (90) and a 4D $^{13}$C-HMQC-NOESY-HMQC (91).

II.iii Non-uniform Sampling and 4D $^{13}$C-HMQC-NOESY-HMQC

Non-uniform sampling is an NMR technique in which data in the time-domain is sampled in a non-linear fashion, i.e. the intervals between sampled points are not all the same length (92). Since the data is no longer linearly acquired, the fast Fourier transform (FFT) can no longer be
used to convert the data into the frequency domain and other computational methods are applied to reconstruct a spectrum consistent with the points sampled. The advantage of this technique is that sparsity of sampling in the indirect dimension means that spectra of similar resolution can be acquired in shorter periods of time or experiments of similar length can achieve higher resolution, which is particularly important when taking advantage of the resolution potential of higher field spectrometers. It also means that higher multidimensional spectra, such as 4D NOESYs, which once required prohibitively long acquisition times, are now possible.

The 4D $^{13}$C-HMQC-NOESY-HMQC used in these structural studies was recorded in a non-uniform fashion with 20% grid point sampling at 291 K on a $^{15}$N, $^{13}$C-perdeuterated-ILV-labelled M50 (residues 1-168) sample. With this sampling schedule, the experiment time was approximately twelve days. The full spectrum was reconstructed with the help of Sven Hyberts (Harvard Medical School) using an iterative soft thresholding algorithm (93) developed in the Wagner Lab. The resulting spectrum was loaded into CARA (94,95) where the $^1$H-$^{13}$C planes for each of the terminal methyl protons of each Ile, Leu, and Val in the primary sequence were examined for crosspeaks. The $^1$H and $^{13}$C frequencies of each crosspeak were recorded and used to identify the residue and proton it belonged to by cross-referencing an assigned $^{13}$C-HSQC of M50 (residues 1-168). One disadvantage of a reconstructed NUS spectrum at this point is that the background noise is not constant across the plane. This means that sometimes areas of high noise occur, mimicking the appearance of a true crosspeak. In an ideal setting, repeated acquisition of the spectrum would allow definitive identification of these spurious peaks as they are generated by noise, which is not consistent between experiments; however in reality, these artifactual crosspeaks are easily eliminated as they virtually never fall on an assigned resonance in the $^{13}$C-HSQC of M50 (residues 1-168). The resulting list of identified constraints were used
II.iv  NMR Structure Calculation

Calculation of an NMR structure requires a series of experiments that are designed to allow the assignment of resonance frequencies to specific protons in the primary sequence and the calculation of distances between those atoms. The resulting distance constraints provide restrictions on the possible conformational space available to the known polypeptide chain and when combined with other energy based optimizations, results in a set of structural models that are consistent with the limitations on spatial geometry placed by the experimental results (96). CYANA (97,98), the software package used to do the structure calculations and solve the M50 (residues 1-168) structure, uses a combination of simulated annealing and molecular dynamics simulations of torsion angle dynamics to avoid local minima and settle the structure into a stable potential well. An NMR structure calculation usually returns a set of structures (unless parameters are set otherwise) and areas of divergence between these structures represent areas with fewer distance constraints and therefore more conformational freedom during the modelling.

The set of spectra necessary for the structure calculation were all acquired at 291 K with M50 (residues 1-168) protein samples at a concentration of ~150-300 µM in 25 mM NaPO₄, pH 6.5,
150 mM NaCl, 1 mM DTT buffer at a variety of spectrometer field strengths. The traditional TROSY-based backbone triple resonance experiments (HNCA/HNCOCA, HNCO/HNCA CO, HNCACB) were acquired using an $^{15}$N,$^{13}$C-perdeuterated M50 (residues 1-168) sample and used to assign the backbone chemical shifts. CCONH, HCCONH, and HCCH-TOCSY experiments carried out on $^{15}$N,$^{13}$C-60% deuterated M50 (residues 1-168) with a TOCSY transfer mixing time of 18 ms and 3D $^{15}$N- and $^{13}$C-dispersed NOESY experiments on uniformly labelled $^{15}$N,$^{13}$C-labelled M50 (residues 1-168) with a mixing time of 90 ms were used to aid the assignment of distal sidechain resonances. Distance constraints were obtained using the 3D $^{15}$N- and $^{13}$C-dispersed NOESY spectra in addition to 3D $^{13}$C,$^{15}$N time-shared NOESY (90) and 4D $^{13}$C-HMQC-NOESY-HMQC experiments (91) acquired on a $^{15}$N,$^{13}$C-perdeuterated-ILV-labelled M50 (residues 1-168) sample. Experiments making use of the ILV-labelled sample required a longer mixing time of 200ms.

All NMR data for the structure calculation were processed using NMRPipe (99) and analyzed using CARA (94,95). The NOESY crosspeaks from all four 3D NOESY experiments were integrated using peakint (N. Schäfer, diploma thesis, ETH), and the resulting volumes were converted into distance constraints using CALIBA (97). In contrast, distance constraints generated from the 4D $^{13}$C-HMQC-NOESY-HMQC were manually identified and then set as upper limit constraints of 5 Å. TALOS+ (100) was used to generate dihedral angle constraints from the backbone chemical shift data and these constraints were combined with the NOESY distance constraints to run a CYANA (97,98) structure calculation.

The final result of the calculation is a PDB file with the coordinates of the twenty lowest energy fold models and an overview file that lists constraint violations. These violations are then examined individually, and if necessary deleted from the constraint file. This process is then
iterated with a new structure calculation. Following each iteration, the results are examined for lower target function values, convergence of the overlaid twenty structures into a tight bundle with RMSD below 1.5 Å, and minor constraint violations if any. The structure calculation is usually initiated with a list of high-confidence constraints. Once an initial fold has been generated, the structure in the current iteration can be used to assign previously ambiguous cross-peaks, allowing addition of more constraints. One must be careful not to do this too early in the refinement process as it is an easy way for bias to be introduced into the structure calculation.

Two common causes of a distance constraint violation is peak overlap and spin diffusion. Peak overlap can make two smaller peaks appear to be one large peak, causing the peak integration program to overestimate the strength of the NOE and thus underestimate the distance between the two linked atoms. In this case, manually loosening the constraint can eliminate this type of violation. Spin diffusion is when magnetization is transferred from one nucleus to another via an intermediate nucleus through the nuclear Overhauser effect (NOE) (96). This can occur if the mixing time used is slightly longer than optimal for the protein in question, and spin diffusion can make two nuclei which are distal to each other appear to be proximal. If not common, spin diffusion can often be identified when a violating distance constraint contains two atoms that are far apart on the current iteration structure and NOE crosspeaks are present from each atom to an intervening nucleus in space. Once such a constraint is identified, it can be loosened or eliminated from the constraint list as appropriate.
The majority of the NMR-based titration experiments were performed using a peptide representing the minimal binding domain of the heterodimeric binding partner as characterized in Sam et al. (84). To this end, a peptide of sequence

RLTLHDLHDFREHPELELKYLNMMKMA (UL53 residues 58-85)

synthesized at the Tufts University Core Facility (Boston, MA) was purchased for use in these experiments.

These experiments consist of the acquisition of a series of $^{15}$N-HSQC spectra acquired at 291 K in which an increasing amount of unlabeled ligand is added to the sample of $^{15}$N-labelled M50 (residues 1-168) protein. Peak intensities and positions are then compared across the spectra. Residues affected by the presence of the ligand, the assumption being as a result of a binding event, will show perturbation in their resonance frequencies and or their peak intensity.

In the case of a variation known as cross-saturation transfer experiments, the protein sample must be perdeuterated and each titration point requires the acquisition of two spectra (often in an interleaved fashion): one which is “on-resonance” meaning the methyl region of the spectrum has been selectively excited and one which is “off-resonance” meaning a region of the spectrum in which no resonances are expected is selectively excited. This dual spectrum acquisition allows one to control for effects on resonance intensity and position merely derived from sample heating due to exposure to the radiofrequency pulses used in the experiment. In the case of these experiments, rather than comparing raw resonance parameters, a ratio of intensities from the “on-resonance” spectrum and the “off-resonance” spectrum is used to determine whether that residue has been affected at that titration point.
II.vi  T1, T2, and Hetero-NOE Experiments

T1 relaxation, also known as longitudinal or spin-lattice relaxation, relates to the re-establishment of spin-state equilibrium, whereas T2 relaxation, also known as transverse or spin-spin relaxation, relates to the loss of coherence or dephasing of the transverse magnetization over time. Both of these occur because of oscillations in the local magnetic field experienced as result of molecular motion (101). The magnitude of the $^1\text{H}-^{15}\text{N}$-heteronuclear NOE is proportional to the correlation time, which relates to the size and tumbling rate of the molecule. While this would seem to suggest there should be one value for all residues, in reality the local motion around each residue has an effect on the hetNOE value for that residue (102). T1, T2, and hetero-NOE experiments are used to look at the internal dynamics of the protein on the ps to ns timescale. This timescale is different than that of larger conformational or domain movements, which generally occur on the $\mu$s to s timescale. For each parameter, a series of experiments is acquired in which a particular delay is increased with each experiment. For each residue, by plotting peak intensity versus delay time, the resulting data can be fit with a decaying exponential of the form

$$\text{Intensity} = A \cdot e^{-R \cdot \text{delay}}$$

where $1/R$ is the time constant (T1 or T2) for that residue. This can be done by hand or using a macro such as the rh command in SPARKY (103). The time constants are then plotted versus residue as seen in Figure III-3. HetNOE values are expressed as a ratio of peak intensities from two experiments, one with proton saturation and one without (102). Areas of greater motion are seen by dips in the T1 data, spikes in the T2 data, and dips again in the hetero-NOE ratios.
II.vii Residual Dipolar Coupling Experiments

A molecule tumbling isotropically in solution results in averaging of dipolar couplings to zero. Introduction of biomolecules to the sample solution capable of aligning to the magnetic field results in the formation of a dilute aqueous liquid crystalline “alignment medium.” This results in momentary stuttering of the tumbling molecule, and this anisotropy results in the dipolar couplings averaging to a non-zero value known as a residual dipolar coupling (RDC). These numerical values can be used to provide orientational constraints or provide validation of an existing model as they relate directly to relative bond orientation in the magnetic field (104). In the case of the experiments used in these structural studies, 10 mg/mL of Pf1 filamentous bacteriophage was used for the alignment medium after optimization of sample conditions. The RDC is not measured directly, but rather calculated as a difference in J-couplings between the isotropic and anisotropic sample. In the case of these studies, the RDC was calculated from the difference between the TROSY and semi-TROSY peak positions of the $^1$H-$^{15}$N backbone amide resonances. With the help of Remy Sounier (formerly of the Chou Lab, Harvard Medical School), these RDCs were fit to, i.e. checked for consistency with, the solution-state NMR structure of M50 (residues 1-168).
III. SOLUTION-STATE NMR STRUCTURE OF M50
[NMR structure calculations were done by Kendra Leigh with the help of Haribabu Arthanari based on NMR spectra acquired by Haribabu Arthanari and assignments done by My Sam Mansueto and Kendra Leigh. Database searches for structural homology were done and analysed by Kendra Leigh. Phyre2 threading runs were performed by Dave Filman. Spectra for the M50 dynamics experiments were acquired and analysed by Kendra Leigh with the help of Haribabu Arthanari. Spectra for the RDC calculations were acquired and analysed by Kendra Leigh with the help of Haribabu Arthanari and Remy Sounier. The unpublished crystal structure of the NEC referenced for comparison in section III.v is the work of Ming Lye.]
III.i  Solution-state NMR Structure of M50 (residues 1-168)

The solution-state NMR structure was calculated using CYANA from approximately 1400 NOE-derived distance constraints, of which about 400 were long-range distance constraints (Table III-1). Constraints are more populated in the core of the protein fold leading to good convergence of the structure in these areas and much more variation in the position of the terminal helices (Figure III-1A). These terminal helices are internally well defined by characteristic α-helix NOEs, but have few long-distance constraints tying them to the body of the protein. It should be noted that two amino acids in strand β6 (Figure III-1B) could not be assigned and therefore were unable to contribute distance constraints to the structure calculation.

The final fold has six α-helices and nine β-strands which are formed into two opposing β-sheets. The overall shape may be considered to form something of a spreading β-sandwich decorated on one side by a number of interspersed α-helices. Three α-helices are ten amino acids or more in length and include the N- and C-terminal helices. The remaining short helices are interspersed between the intervening β-strands. The overall effect of the fold has led it to be affectionately dubbed by Ekaterina Heldwein (Tufts University) as “β-taco.” The fold begins with a spanning N-terminal helix consistent with bioinformatics predictions (68), before going into strand β1 which forms the edge of Face A. The strand continues into a small intervening helix (α2) before forming strand β2, still a part of Face A. From there, the edge of Face B is formed by strand β3 followed by a long bridging helix α3. Crossing back over to Face A, strands β4 and β5 form the middle of the beta-sheet, before again traversing to form the far edge of Face B with strand β6 and of Face A with strand β7. A short helix α4 interrupts before forming the middle of Face B with strands β8 and β9, which are also interrupted by an additional short helix α5. The fold finishes C-terminally with the third long helix, α6.
Figure III-1: CYANA M50 (residues 1-168) structure calculation bundle and cartoon schematic of the lowest energy structure

(A) Side-by-side stereoview of the fifteen lowest energy structures of M50 (residues 1-168) overlayed in MOLMOL (105) by fitting to the lowest energy structure using the residues corresponding to secondary structure elements.

(B) Cartoon schematic of the lowest energy structure from the CYANA structure calculation shown in three orientations. The leftmost orientation shows the two distinct β-sheets labelled as face A and face B. The secondary structural elements are labelled and numbered α1 to α6 and β1 to β9 for the α-helices and β-strands respectively.
Table III-1: NMR refinement statistics for the M50 (residues 1-168) CYANA structure calculation

Pairwise root mean square deviation (RMSD) was calculated from twenty refined lowest energy structures. Secondary structure RMSD was calculated using MOLMOL (105) by fitting to the lowest energy of twenty structures using only the residues corresponding to secondary structure (residues 8-22, 30-32, 34-39, 46-50, 53-69, 74-80, 94-98, 104-108, 112-116, 118-124, 132-140, 144-153, 158-170) as defined in the lowest energy structure.

<table>
<thead>
<tr>
<th>NMR Distance and Dihedral Constraints</th>
<th>M50 (1-168)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distance constraints</strong></td>
<td></td>
</tr>
<tr>
<td>Total NOE</td>
<td>1384</td>
</tr>
<tr>
<td>Intra-residue</td>
<td>261</td>
</tr>
<tr>
<td>Short &amp; Medium Range (1 ≤</td>
<td>i – j</td>
</tr>
<tr>
<td>Long Range (</td>
<td>i – j</td>
</tr>
<tr>
<td><strong>Hydrogen bonds</strong></td>
<td>88</td>
</tr>
<tr>
<td><strong>Total dihedral angle restraints</strong></td>
<td>282</td>
</tr>
<tr>
<td>Φ</td>
<td>141</td>
</tr>
<tr>
<td>Ψ</td>
<td>141</td>
</tr>
<tr>
<td><strong>Structure statistics</strong></td>
<td></td>
</tr>
<tr>
<td>Violations (mean and s.d.)</td>
<td></td>
</tr>
<tr>
<td>Distance constraints (Å)</td>
<td>0.27 ± 0.12</td>
</tr>
<tr>
<td>Dihedral angle constraints (º)</td>
<td>8.76 ± 10.95</td>
</tr>
<tr>
<td>Max dihedral angle violation (º)</td>
<td>60.81</td>
</tr>
<tr>
<td>Max distance constraint violation (Å)</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>Ramachandran Plot Summary from PROCHECK</strong></td>
<td></td>
</tr>
<tr>
<td>Most favoured regions</td>
<td>78.2%</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
<td>19.9%</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>1.9%</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Average pairwise RMSD (Å)</strong></td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>1.87 ± 0.22</td>
</tr>
<tr>
<td>Backbone</td>
<td>1.21 ± 0.21</td>
</tr>
<tr>
<td>Secondary Structure (105)</td>
<td>1.126</td>
</tr>
</tbody>
</table>
Figure III-2: Electrostatic surface of the lowest energy CYANA-calculated structure of M50 (residues 1-168)

The electrostatic surface of the lowest energy structure from the CYANA structure calculation as calculated by PyMol Adaptive Poisson-Boltzmann Solver (APBS) plug-in is presented in two orientations. The units of the visualized surface potential is in $kT/e$ where $k$ is the Boltzmann constant, $T$ is the temperature of the calculation in Kelvin (310 K), and $e$ is the charge of an electron.
The electrostatic surface of the lowest energy fold from the CYANA structure calculation was visualized using the Adaptive Poisson-Boltzmann (APBS) PyMol plug-in written by Michael Lerner (Figure III-2). The electrostatics show a large patch of negative potential over the helices decorating the top of the “β-taco;” however, the presence of primarily hydrophobic residues in the helix identified to be the binding domain of UL53 (the human homologue of M53, the binding partner of M50) (84) suggests that the binding interaction is more the result of the hydrophobic effect than the result of charge-charge interactions.

III.ii Novelty of the M50 (residues 1-168) Protein Fold

The topology of the M50 (residues 1-168) was much more convoluted than the structure initially suggested upon first glance (Figure III-3). Despite the presence of two distinct β-sheets, the strands composing these sheets are mostly non-sequential and result from an interlocking pattern of strands. This hidden complexity coupled with the overall lack of structural and mechanistic understanding of nuclear egress led to us search a number of structure and structural domain databases for possible similar matches. The best results are listed in Table III-2. Strong positive hits would possibly give further hints as to the mechanism of function for M50 and its homologues.

From the CATH Database (107), the top hit was alignment with 71 residues of YIGZ, a conserved hypothetical protein from *E. coli* k12 (CATH: 1VI7A02). The aligned secondary structures showed M50 (residues 1-168) to have additional elements lacking from YIGZ in the region of structural similarity. The most obvious difference is that the CATH domain has a single beta sheet. Due to the intertwined nature of the two beta sheets in the M50 (residues 1-168) fold, it is not possible to form that fold by simple duplication of the CATH-identified
domain. This can be seen in the fact strand β3 in M50 (residues 1-168) is part of the opposing sheet and does not align with the sheet found in the CATH domain as well as the inconsistent alignment of strand β5 from the middle of Face A of M50 (residues 1-168). There is also an additional helix, α2, found in the M50 (residues 1-168) fold that is not present in the CATH domain.

The Dali Structural Comparison Database (108) gave a best hit from a structure of *A. fulgidus* alanyl-tRNA synthetase in complex with wild-type tRNA (Ala) (3WQY:A). The M50 (residues 1-168) fold aligned with residues 536-906 (371 residues) from 3WQY resulting in the aligned secondary structures showing two additional elements (α3, β4) in the region of structural similarity that are absent in the M50 (residues 1-168) fold. Furthermore, some of the aligned strands, such as strands β2 and β5 of 3WQY, run in the opposite direction from the aligned strands (β5 and β7 respectively) of M50 (residues 1-168). α1 of 3WQY is actually part of a long, connecting helix to another domain and thus aligns poorly with α1 of M50 (residues 1-168) and α4 of M50 (1-168) intersects with α3 of 3WQY, but does not align.

FATCAT (109,110) returned an alignment with 77 residues of an Apo acyl carrier protein from *E. coli* (d1T8k:A); however this hit is entirely alpha helical in nature and therefore bears little resemblance to the M50 (residues 1-168) fold which consists of both α-helices and β-strands.

VAST (111) and PDBFold (112) returned the same top hit of Toluene-4-monooxygenase System Effector Protein from *P. mendocina* (3DHI:E). Aligned secondary structural elements show additional α-helices in the M50 (1-168) fold (α2) as well as in chain E of 3DHI (α1, α4) that are absent in the other. Comparison of the topologies show differences with strands 6 and 7 of 3DHI:E aligning with strands 9 and 8 of M50 (residues 1-168) respectively. There are also
Figure III-3: Schematic of the topology of M50 (residues 1-168)

The schematic depicts the topology of the lowest energy structure from the CYANA structure calculation. α-helices are depicted by blue cylinders and β-strands are depicted by red arrows, with residue numbers for the beginning and end of the secondary structure elements included in black.
Table III-2: Top hits, scores, and secondary structure analysis from structural comparison databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Top Hit</th>
<th>Score</th>
<th>RMSD (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATH Database (107)</td>
<td>YIGZ, a conserved hypothetical protein from <em>E. coli</em> k12 (CATH: 1VI7A02) [71 residues]</td>
<td>sequential structure alignment program (SSAP) 78.3</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>CATH Domain: 1VI7A02 M50 (1-168)</td>
<td>N:β1-α1-β2-β3-α2-β4:C</td>
<td></td>
</tr>
<tr>
<td>Dali Structural Comparison</td>
<td><em>A. fulgidus</em> alanyl-tRNA synthetase in complex with wild-type tRNA (Ala) (3WQY:A) [Alignment with residues 536-906: 371 residues]</td>
<td>Z-score: 4.2 4.1</td>
<td>4.1</td>
</tr>
<tr>
<td>Database (108)</td>
<td>3WQY Chain A M50 (1-168)</td>
<td>N:α1-0.5-β1-α2-β2-β3-α3-β4-β5-α4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N:α1-β1-α2-β2-β3-α3-β4-β5-α4</td>
<td></td>
</tr>
<tr>
<td>FATCAT (109,110)</td>
<td>Apo acyl carrier protein from <em>E. coli</em> (dIT8k:A) [77 residues]</td>
<td>Score: 105.75 P-value: 4.61e-03 Opt: 3.13 Chain: 2.03</td>
<td></td>
</tr>
<tr>
<td>VAST (111)</td>
<td>Toluene-4-monoxygenase System Effector Protein from <em>P. mendocina</em> (3DHI:E) [103 residues]</td>
<td>61 Aligned Residues</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>3DHI Chain E M50 (1-168)</td>
<td>N:α1-β1-α2-β2-β3-α4-3-β5-β6-β7:C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N:---β5-α1-β1-α2-β3-α3-β4-β9-β8:C</td>
<td></td>
</tr>
<tr>
<td>PDBeFold (112)</td>
<td>No non-identical matches with 70% similarity or higher; Toluene-4-monoxygenase system protein D from <em>P. mendocina</em> (3Q14:E) [103 residues]</td>
<td>46% of M50 (1-168) secondary structure was identified in the target protein 55% of target chain secondary structure was identified in M50 (1-168).</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td>3Q14 Chain E M50 (1-168)</td>
<td>N:α1-β1-α2-β2-β3-β4-3-α4-3-β5-β6-β7:C</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N:---β5-α1-β1-α2-β3-α3-β4-β9-β8:C</td>
<td></td>
</tr>
<tr>
<td>MarkUs Protein Function</td>
<td>AlaX-M trans-editing enzyme, C-terminal domain from <em>P. horikoshii</em> (SCOP: d2e1ba2) [129 residues]</td>
<td>Protein Structural Distance (PSD) Score 0.34 Structural Alignment Score (SAS = RMSD x 100 / number aligned residues) 3.9</td>
<td>2.58</td>
</tr>
<tr>
<td>Annotation Server (113)</td>
<td>SCOP Domain:d2e1ba2 M50 (1-168)</td>
<td>N:α1-β1-β2-β3-β4-β4.5-α3-β5-β6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N:α1-β1-α2-β3-α3-β4-β5</td>
<td></td>
</tr>
</tbody>
</table>
two intervening strands (β6, β7) in the M50 (residues 1-168) fold that are part of opposing faces A and B that do not align with any elements in 3DHI:E.

Finally the Skan (113,114), run through the MarkUs Protein Function Annotation Server (113), returned the C-terminal domain of an AlaX-M trans-editing enzyme from *P. horikoshii* (SCOP: d2e1ba2). While superficially both have α-helical and β-stranded character, the aligned secondary structures show a number of differences. There are additional elements in the M50 (residues 1-168) fold (α2) and as well as in the d2e1ba2 SCOP domain (β3, β4, β4.5, α3) that are absent in the other.

Overall, none of the hits from the structural databases were similar to the overall M50 (residues 1-168) fold, suggesting that M50 shows novel structural character. Furthermore with the exception of VAST and PDBeFold, none of the databases gave the same or even proteins of similar function as a top hit. While from the standpoint of structural biology, the M50 (residues 1-168) fold expands the general knowledge of structural domains; from a biological standpoint, it does not allow us to draw any conclusions about the mechanism of nuclear egress.

### III.iii Dynamics of the M50 (residues 1-168) Structure

While some areas of the fold, such as the terminal helices showed fewer constraints to the body of the protein (Figure III-1A) and thus poorer convergence in the bundle of structures, this is not necessarily reflective of more motion or flexibility in this region. Standard longitudinal (T1) and transverse (T2) relaxation, and hetero-NOE experiments were therefore carried out in order to look at the internal dynamics of the M50 (residues 1-168) fold. Overall, the data (Figure III-4) and average T2 of approximately 50 ms are consistent with a relatively rigid 18 kDa protein.
The rigidity is implied by the rather flat profile of the data across the residues. As can be seen by alignment with the secondary structural elements found in the lowest energy CYANA-calculated structure (Figure III-4; top), the most mobile areas are the terminal helices and regions in or very near the unstructured loops. The data, however, does not give any strong indicators as to possible areas of conformational change that may occur upon binding of the heterodimeric protein partner. It is possible that conformational changes upon binding are subtle and therefore do not require a high degree of mobility.
Figure III-4: T1, T2, and Hetero-NOE values graphed by residue

T1, T2, and Hetero-NOE values are plotted versus residue number in the upper, middle, and lower panels respectively. Values were derived experimentally from the corresponding NMR experiments run at 291 K. Secondary structure designations derived from the lowest energy CYANA calculated structure are depicted along the top with regions corresponding to α-helices spanned by blue rectangles and to β-strands by red arrows.
III.iv  Threading of Other Sequences onto the M50 (residues 1-168) Structure

The sequential conservation of the nuclear egress complex across herpesviruses (Figure III-5) suggests that the structure of M50 (residues 1-168) may be more broadly conserved as well. In order to try and predict whether this might be true, with the help of Dave Filman (Harvard Medical School), we turned to a technique known as “threading.” Threading uses sequence homology to predict protein structure for sequences of proteins whose structure is still unknown.

In this case, we first used the Protein Homology/analogY Recognition Engine v2.0 (Phyre2) in BackPhyre mode to screen a host of bacterial and eukaryotic sequence databases (*Arabidopsis thaliana*, *Bdellovibrio bacteriovorus*, *Caenorhabditis elegans* [WS220.66], *Clostridium difficile*, *Drosophila melanogaster*, *Homo sapiens*, *Leishmania major strain Friedlin*, *Mus musculus*, *Mycobacterium tuberculosis* [CDC1551], *Plasmodium falciparum*, *Saccharomyces cerevisiae*, *Sulfolobus solfataricus* [P2], *Thermoplasma acidophilum*, *Agrobacterium tumefaciens* [C58 Cereon], *Bacillus subtilis*, *Bartonella henselae* [Houston-1], *Corynebacterium diphtheria*, *Desulfitobacterium hafniense* [Y51], *Escherichia coli* [K12], *Lactobacillus casei* [ATCC 334], *Neisseria meningitidis* [MC58], *Pseudomonas aeruginosa*, *Staphylococcus aureus* [COL], *Streptococcus pneumonia* [D39], *Streptomyces coelicolor*, *Synechococcus* [CC9311], *Yersinia pestis* [CO92]) against the M50 (residues 1-168) coordinates. None of the returned results had a confidence score higher than 42.1% (an *Arabidopsis* protein of unknown function), fitted more than 43 amino acids (out of a possible 244 total; *C. elegans* protein [F18E9.8]), or formed a compact domain with the aligned polypeptide segments of M50. Phyre2 was then used to predict structures using the sequences from known herpesviral homologues of M50, HCMV UL50 (100% confidence score) and HSV-1 UL34 (97.91% confidence score) (Figure III-6). Even the more distantly related HSV-1 sequence shows strong homology and a very similar
predicted structure to that of the M50 (residues 1-168) fold. The Phyre2 algorithm is based on the matching of hidden Markov models and therefore does not take energy optimization into account in the structure predictions. This means that the actual structural homologies will probably prove to be somewhat less striking than the confidence scores suggest. The high degree of similarity in the predicted structures, however, does argue for the general applicability of the overall protein fold presented here to the other herpesviral homologues, and the failure to find strong matches in the database searches lends support to the rare nature of the M50 (residues 1-168) structure.
Figure III-5: Sequence alignment of the human herpesvirus homologues with MCMV M50

The alignment of the human herpesvirus homologues with the MCMV M50 sequence (bold) was generated using Clustal Omega (72). * (asterisk) indicates positions that have an identical residue; : (colon) indicates conservation between groups of strongly similar properties – scoring >0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties – scoring ≤ 0.5 in the Gonnet PAM 250 matrix. Residues with non-zero normalized ratio of peak intensities lower than 0.4 in titration experiments with 100% UL53 peptide and isotopically enriched M50 (residues 1-168) are highlighted on the M50 sequence in red (Section IV.i). Secondary structural elements from the M50 (1-168) lowest energy fold are delineated by blue rectangles in areas of α-helical character and red arrows in areas of β-stranded character. The predicted transmembrane (TM) domain of M50 is highlighted by the grey box.
Figure III-5 (continued).
Figure III-6: Homologous residues from UL50 and UL34 threaded onto the M50 (residues 1-168) structure

Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/) (115) HCMV UL50 (red) and HSV-1 UL34 homologous residue threading results (blue) onto the M50 (1-168) lowest energy model (grey). were aligned and rendered in PyMol (116).
III.v Residual Dipolar Couplings, and M50 (residues 1-168) Structure Validation

Due to the lack of structural homologues, some effort was made to try and independently validate the structure. For this we initially turned to residual dipolar couplings (RDCs) to provide orientational constraints to fit to the calculated M50 (residues 1-168) fold with the help of Remy Sounier (formerly of the Chou Lab, Harvard Medical School). The RDCs calculated for the backbone N-H couplings were not inconsistent with the structure calculated from NOE-derived distance constraints, but the technique was insufficient to provide strong validation. Due to the assumptions of idealized amino acid geometry at the heart of the relatively fast torsion-angle-dynamics-based CYANA structure calculations, RDC values are often a poor fit, even for NMR structures verified to be correct by other methods such as x-ray crystallography, and therefore quality of fit is not necessarily an indicator of calculation accuracy.

Years of attempts to crystallize the NEC complex finally led to diffractable crystals of a chimeric MCMV/HCMV NEC provided by Ming Lye (Hogle Lab, Harvard Medical School). The M50 (residues 1-168) NMR structure was sufficient to be used in molecular replacement to build a partial map with density corresponding to the UL50 portion of the complex. This result is consistent with previous knowledge that the purified complex is more stable than the purified components alone, suggesting that some conformational changes likely occur upon binding to stabilize the proteins, and thus explaining why a full density could not be determined. Following the use of other techniques to obtain phases, Dr. Lye has been able to build an initial model of the truncated NEC (data unpublished), and the structure of the UL50 portion of the complex is consistent with the solution-state NMR structure of M50 (residues 1-168) presented here.
IV. HETERODIMERIZATION INTERFACE
[NMR titration and cross-saturation transfer experiments with the UL53 peptide were acquired and analysed by Kendra Leigh with the help of Haribabu Arthanari. The ITC data is the work of My Sam Mansueto. The viral growth curve data is the work of Mayuri Sharma. Purification of M53 (residues 103-333) was done by Kendra Leigh and the NMR titration experiments using the purified product were performed and analysed by Kendra Leigh with the help of Haribabu Arthanari.]
IV.i Initial Characterization of the Binding Interface

Initial characterization of the binding surface on M50 (residues 1-168) was done using NMR titration experiments (Figure IV-1A) with $^{15}$N-labelled M50 (residues 1-168) and a conserved minimal UL53 peptide (78.6% homology with the corresponding residues in M50). The sequence of this peptide encompasses the residues of the core UL53 binding domain and is predicted to have alpha-helical character by PHYRE (84), although it is unknown if the peptide has secondary structure without the larger context of the whole protein. The ratios of intensities were calculated for the resonances of each residue in the spectra acquired on $^{15}$N-M50 (residues 1-168) samples with and without the presence of equimolar UL53 peptide (Figure IV-1B). An arbitrary cut-off of 0.4 was used to identify residues “significantly affected” by peptide binding.

This method implicated a continuous patch of residues structurally proximal to the C-terminal helix (Figure IV-2) and coinciding with the largest cavity (out of sixteen) identified by SCREEN (113,117), a solvent-accessible cavity identification program, as being the core binding interface for formation of the NEC heterodimer.

Residues identified as potentially important in the NMR titration experiments were then targeted for site-directed mutagenesis to alanine. Purified protein was expressed containing the selected point mutations in M50 (residues 1-168) and in the homologous residues of UL50 (residues 1-169). These mutant proteins were then used for isothermal titration calorimetry (ITC) experiments by My Sam Mansueto (formerly of the Hogle Lab, Harvard Medical School) to measure the effects of each point mutation on binding affinity with the un-mutated heterodimeric partner. A subset of the corresponding UL50 mutations were then also introduced into the HCMV expression bacterial artificial chromosome (BAC) and used to determine the effect of the point mutation in the context of the HCMV infection of human foreskin fibroblasts (HFFs) by
Figure IV-1: HSQC spectra of M50 (residues 1-168) with and without UL53 peptide and bar graph of intensity ratios by residue

(A) The HSQC spectrum of $^{15}$N-M50 (residues 1-168) in red is overlayed with the HSQC spectrum of $^{15}$N-M50 (residues 1-168) with equimolar unlabelled UL53 peptide titrated into the sample in blue. The inset box shows the resonance for Y57 as an example of a peak that shows chemical shift perturbation upon addition of the peptide and S82 (asterisk) as an example of a peak that shows no chemical shift perturbation.

(B) Bar graph of the normalized ratio of peak intensities from the spectrum with equimolar unlabelled UL53 peptide present to peak intensities from the spectrum with no unlabeled UL53 peptide present for each residue. Areas highlighted by the grey boxes are residues that show marked reduction in peak intensity upon addition of peptide.
Figure IV-1 (Continued).
Figure IV-2: Residues affected by UL53 peptide binding mapped onto the M50 (residues 1-168) structure

Normalized ratios of peak intensities were calculated from the HSQC spectrum with $^{15}$N-M50 (residues 1-168) and equimolar unlabelled UL53 peptide present and the HSQC spectrum with no peptide present for each residue.

(A) Residues showing a non-zero normalized ratio of peak intensity less than 0.4 (red) were mapped onto a cartoon representation of the CYANA-calculated lowest energy structure of M50 (residues 1-168).

(B) Residues showing a non-zero normalized ratio of peak intensity less than 0.4 (red) were mapped onto a surface representation of the CYANA-calculated lowest energy structure of M50 (residues 1-168).
Table IV-1: Effect of alanine mutations on binding affinity of M50 (residues 1-168) and UL50 (residues 1-169) for their heterodimeric binding partners and on HCMV replication

Kₐ values ± standard deviations were calculated from isothermal calorimetry data collected using the method as described in Sam et al. (84). For wild type (WT) and single point mutant M50 (residues 1-168) proteins, M53 (residues 103-333) was used as a binding partner. For WT and single point mutant UL50 (residues 1-169) proteins, UL53 (residues 50-292) was used as a binding partner. NB stands for no detectable binding. Greyed-out cells are untested conditions. Data courtesy of My Sam Mansueto (unpublished; Harvard Medical School). HCMV replication phenotypes were characterized in viral growth kinetics experiments that were performed by Mayuri Sharma (unpublished; Harvard Medical School) using the method as described previously in Sharma et al. (19). Factor reductions in yield were calculated by dividing the supernatant viral titer of the mutant virus by the supernatant viral titer of wild type virus at day 7 post infection. No detectable virus in the supernatant at 7 days post infection resulted in characterization of the mutant as non-viable.

<table>
<thead>
<tr>
<th>M50 (1-168) Mutation</th>
<th>K₀ (µM)</th>
<th>UL50 (1-169) Mutation</th>
<th>K₀ (µM)</th>
<th>HCMV Replication Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild Type (WT)</td>
<td>0.97 ±0.12</td>
<td>WT</td>
<td>0.29</td>
<td>WT</td>
</tr>
<tr>
<td>V52A</td>
<td>NB</td>
<td>V52A</td>
<td>1.4 ±0.72</td>
<td>Reduced yield by a factor of 5 to 10</td>
</tr>
<tr>
<td>E56A</td>
<td>~30-90</td>
<td>Y57A</td>
<td>NB</td>
<td>Non-viable</td>
</tr>
<tr>
<td>Y57A</td>
<td>4.1 ±0.35</td>
<td>S125A</td>
<td>2.2</td>
<td>Reduced yield by a factor of 5 to 10</td>
</tr>
<tr>
<td>S60A</td>
<td>~17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I121A</td>
<td>~98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C124A</td>
<td>~76</td>
<td>L130A</td>
<td>7.9</td>
<td>Non-viable</td>
</tr>
<tr>
<td>V128A</td>
<td>0.89</td>
<td>V129A</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>L129A</td>
<td>~60-85</td>
<td>E154A</td>
<td>0.51 ±0.24</td>
<td></td>
</tr>
<tr>
<td>D153A</td>
<td>0.34 ±0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N154A</td>
<td>16 ±2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure IV-3: Heat map of ITC $K_D$ values mapped onto the M50 (residues 1-168) structure

ITC $K_D$ values from Table IV-1 mapped onto two orientations of the ribbon representation of the lowest energy structure from the M50 (residues 1-168) CYANA structure calculations. Higher $K_D$ values (weaker binding) are represented in stronger red colors whereas $K_D$ values closer to wild type are represented in greyer colors.
Mayuri Sharma (Harvard Medical School). These experiments found that single point mutations of residues identified in the NMR titration experiments were able to weaken the binding affinity, in some cases to the point of non-detection, and that the degree to which binding was weakened correlated to the strength of the mutation’s effect in the context of HCMV infection (Table IV-1). Mapping the $K_D$ of the alanine mutants as measured by ITC onto the surface of the structure (Figure IV-3) shows that residues which have greater effect (stronger red) coincide with the center of the region mapped in Figure IV-2A, running along the inside edge of the C-terminal helix and the bridging loop, suggesting the core of the interaction occurs in this region.

There were several mutations that had virtually no effect on $K_D$ as measured by ITC such as V128A and D153A, and others that had differential effects on M50 (residues 1-168) and UL50 (residues 1-169) binding such as M50 V52A, Y57A, C124A, or L129A. The residues with minimal effect on binding are found on the tip of loops at the edge of the binding surface. As peripheral regions of greater flexibility, it is expected that their influence on such a large binding interface would be less that residues found in the more rigid secondary structure at the core of the interaction site. It would be expected that were one of these minimal effect mutations be introduced into the context of the virus, that it would have little to no effect on nuclear egress and viral growth kinetics. These experiments, however, have not yet been performed.

The residues with differential effect are more difficult to explain without structural characterization of the HCMV NEC for comparison. In the MCMV M50 (residues 1-168) structure, V52 is positioned towards the center back of the region of interaction and mutation results in non-detectable binding. It is possible that in HCMV UL50, V52 is positioned slightly differently as a nearby proline (P54) found in M50 is replaced by a cysteine in the corresponding position in UL50. Y57A shows an effect in both M50 (residues 1-168) and UL50 (residues 1-
169), but a much stronger effect (non-detectable binding) in the human homologue. Y57 in M50 (residues 1-168) is positioned pointing outward on helix α3, at the core of the interaction. Conservation of this helix is very high between MCMV and HCMV; however, again the absence of P54 in HCMV may result in a slight shifting of the position of this helix, resulting in a corresponding re-positioning of Y57, giving it greater involvement in complex formation.

C124A weakens binding more effectively in M50 (residues 1-168) than the corresponding mutation S125A does in UL50 (residues 1-169). Cysteine is structurally similar to serine, with the replacement of the oxygen atom found in serine with sulfur leading to a slight decrease in polarity of the bond due to the weaker electronegative properties of sulfur. It is unlikely that this difference explains the difference in binding effects since cysteine is closer in hydrophobic character to alanine than serine is and therefore the mutation would be less of a disruption in terms of amino acid character, even if the residue is involved in some sort of modest charge-charge interaction. Since the corresponding residue in HCMV is shifted by one (residue 124 versus residue 125), this moves its position further into the loop and away from the interaction site, and this is a much more likely explanation for the weaker effect of S125A on binding.

L129A also has a greater effect in M50 (residues 1-168) than mutation of the corresponding residue L130 in UL50 (residues 1-169). Again, in M50 (residues 1-168), L129 is found at the core of the interaction site, at the edge of strand β8, near where the C-terminal helix comes to meet the body of the protein. Again, since the corresponding residue in HCMV is shifted by one (residue 130 versus residue 129), this positions the residue further into strand β8 and away from the interaction surface, possibly explaining the more modest effect. Once finished, the structural work on the HCMV NEC complex by Ming Lye will help to shed more light on the basis of these differences.
IV.ii Cross-Saturation Transfer Experiments with UL53 Peptide

Due to the possibility of allosteric effects secondary to a binding event also having an effect on the electronic environment of residues and thus being detectable in NMR titration experiments, we also ran a series of cross-saturation transfer (CST) experiments to try and eliminate residues that may not be involved in the direct binding event. The weakness of CST experiments is that resonances that are weak to begin with, that overlap with other resonances, or that broaden instead of shift can confound the results, since residues are identified based on a ratio of intensities. Also due to the assumptions upon which the experiment is designed, imperfect deuteration of the experimental protein can also result in some signal leakage as it means the on-resonance pulse will capture some protein resonances as well as those of the ligand.

In the case of experiments with M50 (residues 1-168) and equimolar amounts of UL53 peptide, very few residues were “positive hits.” The tendency of resonances to broaden rather than shift in the M50 (residues 1-168)-UL53 peptide interaction, indicative of intermediate exchange (µM affinity and msec timescale), means that in the context of a larger decrease of intensity due to the presence of peptide, the difference between the on-resonance and off-resonance intensities are less profound. Those that did show a change in intensity were clustered around significant residues and the C-terminal helix. This coincides with the larger interface identified in the NMR titration experiments (Figure IV-4) and thus confirms that area as the core of the binding surface with the nucleoplasmic heterodimer partner rather than being the result of indirect binding effects.
Figure IV-4: Residues affected by UL53 peptide binding in CST experiments mapped onto the M50 (residues 1-168) structure

Ratios of peak intensities were calculated from the on-resonance HSQC spectrum with $^{15}$N-M50 (residues 1-168) and equimolar unlabelled UL53 peptide present and the off-resonance HSQC spectrum with no peptide present for each residue. A mesh surface representation overlaying a cartoon representation of the lowest energy CYANA-calculated structure of M50 (residues 1-168) is depicted. Residues that showed chemical shift perturbation in the NMR titration experiments are colored in red. Residues that showed a ratio of peak intensity below 0.7 are colored in yellow. Residues that were affected both in the standard NMR titration and well as the cross-saturation transfer experiments are colored in orange.
IV.iii  Purification of M53 (residues 103-333)

M53 (residues 103-333) was purified using methods similar to those for M50 (residues 1-168). While the $^1$H-$^{15}$N TROSY of protonated M53 (residues 103-333) showed good dispersion (Figure IV-5), due to its slightly larger size (~27 kDa), it would be advantageous to use perdeuterated protein for the necessary NMR experiments. This is protein that is grown in D$_2$O, but purified in H$_2$O-based buffers similarly to the purification of standard protonated protein. Despite the entire purification process occurring at 4°C, we observed in the context of M50 (residues 1-168) purification that the resultant protein had the expected number of resonances in an $^{15}$N-HSQC suggesting that the backbone amide deuterons had sufficient time to exchange for protons over the course of the purification process.

However, when we acquired an initial test TROSY-selected $^{15}$N-HSQC spectrum of perdeuterated M53 (residues 103-333) we saw significantly fewer than expected resonances, suggesting an incomplete back-exchange of backbone protons during the purification process, unlike what is observed with expressed M50 (residues 1-168) protein. Furthermore, back-exchange could be facilitated and monitored by incubation of the sample at 298 K and acquisition of a series of TROSY-selected $^{15}$N-HSQC spectra, where we observed appearance of the additional expected signals over time. M50 (residues 1-168) was unstable in the spectrometer and precipitated immediately at temperatures higher than 291 K. The incomplete back-exchange combined with greater temperature stability suggests that M53 (residues 103-333) may form a less breathable, more compact and stable fold, unlike that of M50 (residues 1-168) since replacement of deuterons with protons in the core of the protein require molecular-motion based solvent access to the core of the protein. This increased stability will prove useful
Figure IV-5: TROSY-selected $^{15}$N-HSQC of M53 (residues 103-333)

The spectrum depicted is a TROSY-selected $^{15}$N-HSQC of $^{15}$N-labeled, protonated M53 (residues 103-333) acquired at 291 K on a 500 MHz Varian spectrometer.
for any NMR-based experiments that might be conducted in the future using M53 (103-333) alone.

IV.iv  Titration Experiments with M53 (residues 103-333)

In order to compare the effects on the M50 (residues 1-168) structure of binding UL53 peptide versus a much larger portion of the binding partner, NMR titration experiments were conducted using M53 (residues 103-333). Unlike with UL53 peptide, which showed a compact core region proximal to the C-terminus, residues affected by binding of M53 (residues 103-333) were located across the body of M50 (residues 1-168) (Figure IV-6) even with addition of only 10% of the equimolar amount of M53 (residues 103-333). The C-terminal helix showed perturbations that were close to, but not below, the arbitrary ratio cut-off of 0.6, and there were virtually no residues unaffected by addition of M53 (residues 103-333). The use of the UL53 peptide allowed recognition of the core binding site, since other residues were not present to draw focus; however, with the use of M53 (residues 103-333), we see that upon binding, the greatest perturbations in structure are seen elsewhere in the protein fold. Since NMR titration experiments are very sensitive to any changes in the electronic environment of the residues, including as a result of allosteric effects, these global perturbations suggest that at the very least formation of the heterodimer results in subtle conformational change across the entirety of the protein fold. This conclusions is consistent with the knowledge that the purified NEC components are much more stable in the context of the complex than alone as well as the observation that the M50 (residues 1-168) structure could not be used to completely phase x-ray diffraction data collected from crystals of the protein heterodimer as touched on in section III.v.
Figure IV-6: Residues affected by titration of M53 (residues 103-333) in comparison to residues affected by titration of UL53 peptide

Depicted in grey is a cartoon representation of the lowest energy CYANA-calculated structure of M50 (residues 1-168). Ratios of peak intensities were calculated from the reference HSQC spectrum of $^{15}$N-M50 (residues 1-168) and the HSQC spectrum of $^{15}$N-M50 (residues 1-168) with M53 (residues 103-333) at a concentration 10% that of M50 (residues 1-168). Residues affected in the NMR titration experiments with UL53 peptide are shown in red. Residues affected in the NMR titration experiments with M53 (residues 103-333) using a peak intensity ratio cut-off of 0.6 are shown in yellow. Residues that are affected in both the UL53 peptide and M53 (residues 103-333) experiments are shown in blue.
A key set of residues around V52 and E56 are affected in both the UL53 peptide and M53 (residues 103-333) NMR titration experiments. This is also the area, where mutation shows some of the greatest effects on binding affinity and virus viability (Table IV-1). Further structural and biochemical data will be needed to fully reconcile this broader effect with that of the validated core binding site identified using the minimal UL53 peptide.
V. SMALL MOLECULE INHIBITION OF HETERODIMERIZATION
[In silico screening of the ZINC database was done by B. Jayaram. NMR titration experiments with the small molecules were done and analysed by Kendra Leigh with the help of Haribabu Arthanari. Cell culture based experiments with the small molecules were performed by Mayuri Sharma.]
Once a core binding site had been identified, this gave a specific interface that could be targeted for disruption by small molecule inhibitors. The existence of lethal single amino acid mutations that could also be shown to abrogate binding between the two proteins composing the NEC suggested that the binding interface is targetable. In collaboration with B. Jayaram from the Indian Insitute of Technology Delhi, the coordinates of 119,380 bioactive molecules and a further 1,043,549 molecules from the ZINC database (118) were screened in silico against the M50 (1-168) fold using the RASPD algorithm (119) and knowledge of important residues on the M50 binding surface. The RASPD protocol uses calculated physicochemical properties of the targeted residues and the ligand to calculate a protein-ligand binding energy. Docking was accomplished using ParDOCK software (120,121) and the binding free energy of the resulting structure was scored using the BAPPL scoring function (122), which takes into account the energy of electrostatics, van der Waals forces, hydrophobicity, and loss of conformational entropy of protein side chains upon binding.

Ten screened molecules (with a formal charge between -1 and +1 inclusive and fewer rotatable bonds) from the bioactive molecule grouping and from the remaining small molecule grouping were subjected to further atomic docking and scoring. Top scoring hits were subjected to further molecular dynamics (MD) simulations using the AMBER 12 software package (123-126). Simulations began with neutralization of the simulated docked complex with a commensurate number of counter ions and solvation with a TIP3P intermolecular potential function (127) in an octahedral box (leaving a distance of 8 Å between the complex surface and the box boundary). The system was heated from 100 K to 300 K (using a Berendsen thermostat), with a solute restrain force of 25 kcal/mol using a constant volume condition for 100 ps. This was followed by a gradual relaxation of restrain force (5.0, 4.0, 3.0, 2.0, 1.0, and 0.5 kcal/mol) on the solute for
a total period of 300 ps at 300 K, while the periodic boundary condition in the NPT (isothermal-isobaric) ensemble was maintained using the Berendsen thermostat for temperature control, a constant pressure of 1 atm, and isotropic position scaling for pressure control. These conditions were maintained for a 5 ns equilibration followed by a 100 ns unrestrained production simulation, using the SHAKE algorithm (128) to fix all hydrogen-containing covalent bonds and the particle mesh Ewald (PME) method (129) to address long-range electrostatic interactions. The time step was set to 2 fs using a cut-off radius of 8 Å for non-bond interactions, and trajectory coordinates were saved at 5 ps intervals. The generated coordinate files were processed using the PTRAJ AMBER module (130). The PTRAJ module generates 10,000 snapshots over the course of the 100 ns production run, from which the final 1000 systems were taken for calculation of an average BAPPL score. An average BAPPL score below -10 kcal/mol (∼ 50 nm K_D) was taken to be a positive hit.

Of the initial top ten positive hits, only two were still available for purchase, so an additional two purchasable hits were selected. These four available hits were ordered for further testing (Figure V-1; Table V-1). Of these four, so far two have been tested for binding to M50 (residues 1-168) in NMR titration experiments (Figure V-2, Figure V-3, Figure V-4) and three have been tested for their effect on HCMV titer in cell culture-based experiments performed by Mayuri Sharma (Figure V-5; Harvard Medical School). NMR titration experiments showed that at the tested compound concentration of 150 µM, ZINC38147030 and ZINC01192203 show chemical shift perturbations in the same region identified as important for binding of the UL53 peptide (Figure V-4A, C and Figure V-4B, respectively), indicating that not only binding of the small molecule occurs, but that it also occurs at the desired binding interface. In the case of ZINC38147030, residues closest to the MD-modelled binding site of the small molecule consist of hydrophobic
Figure V-1: Chemical structures of the purchasable top-hits from in silico screen of the ZINC database against the M50 (residues 1-168) fold

Table V-1: Binding free energy values of the purchasable top-hits from *in silico* screen of the ZINC database against the M50 (residues 1-168) fold

For each of the small molecules listed (as identified by their ZINC database identification number), the value of the calculated binding free energy after the ParDOCK ligand binding simulation, the value of the calculated binding free energy after the 100 ns production step of the molecular dynamics (MD) simulation, and the final average BAPPL-calculated score are given. All values are in kcal/mol.

<table>
<thead>
<tr>
<th>ZINC Database ID (Source)</th>
<th>Binding free energy after docking</th>
<th>Binding free energy after 100 ns MD simulation</th>
<th>Simulation score after 100 ns MD simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC38147030 (ChemDiv)</td>
<td>-11.12</td>
<td>-10.91</td>
<td>-43.7</td>
</tr>
<tr>
<td>ZINC38144557 (Life Chemicals)</td>
<td>-14.40</td>
<td>-10.26</td>
<td>-43.4</td>
</tr>
<tr>
<td>ZINC01192203 (ChemBridge)</td>
<td>-13.99</td>
<td>-11.33</td>
<td>-40.47</td>
</tr>
<tr>
<td>ZINC08592215 (ChemDiv)</td>
<td>-11.93</td>
<td>-9.64</td>
<td>-27.52</td>
</tr>
</tbody>
</table>
Figure V-2: HSQC spectra of M50 (residues 1-168) with and without addition of ZINC38147030

The HSQC spectrum of $^{15}$N-M50 (residues 1-168) in red is overlaid with the HSQC spectrum of $^{15}$N-M50 (residues 1-168) with 3x the amount of ZINC38147030 titrated into the sample in blue. The inset boxes show a larger view of some of the affected resonances and their residue assignment.
Figure V-3: HSQC spectra of M50 (residues 1-168) with and without addition of ZINC01192203

The HSQC spectrum of $^{15}$N-M50 (residues 1-168) in red is overlaid with the HSQC spectrum of $^{15}$N-M50 (residues 1-168) with 3x the amount of ZINC01192203 titrated into the sample in blue. The inset boxes show a larger view of some of the affected resonances and their residue assignment.
Figure V-4: M50 (residues 1-168) residues affected by addition of small molecules identified by in silico screening of the ZINC database

(A) Residues affected by addition of three-times molar excess of ZINC38147030 as measured by chemical shift perturbation in NMR-based titration experiments mapped (red) onto the lowest energy structure from the M50 (residues 1-168) CYANA structure calculation.

(B) Residues affected by addition of three-times molar excess of ZINC01192203 as measured by chemical shift perturbation in NMR-based titration experiments mapped (red) onto the lowest energy structure from the M50 (residues 1-168) CYANA structure calculation.

(C) Residues affected by addition of three-times molar excess of ZINC38147030 as measured by chemical shift perturbation in NMR-based titration experiments mapped (red) onto the M50 (residues 1-168) structure resulting from the molecular dynamics (MD) simulation of protein-ligand (ZINC38147030; green) binding.

(D) A close-up of the representation depicted in panel C rotated 30° around the PyMol y-axis.
Figure V-5: Effect of *in silico* small molecule inhibitors on HCMV viral titer

\[
\log_{10} \text{Viral Titer (PFU/ml)}
\]

Log$_{10}$ of the measured viral titer of supernatant harvested six days post infection from human foreskin fibroblasts (HFFs) infected at an MOI of 0.1 graphed versus the concentration of small molecule inhibitor used to treat the infected cells two hours post virus adsorption. Data courtesy of Mayuri Sharma (unpublished; Harvard Medical School). DMSO: dimethyl sulfoxide; GCV: Ganciclovir; MCV: Maribavir; 487: ZINC38147030; 511: ZINC08592215; 549: ZINC01192203
and non-charged polar amino acids (Figure V-4D; V128, T130, N131, A149, F150, L161), which is consistent with binding a ligand with aromatic ring character as well as carbonyl, amide, and ester moieties. Initial experiments in human foreskin fibroblasts infected with HCMV show that while not as effective as known anti-herpetic drugs such as ganciclovir and maribavir, the in silico hits do have moderate effects on viral titer in a way that correlates to their average score after MD simulation. These molecules have undergone no experimental optimization such as structure-activity relationship (SAR) analysis or medicinal chemistry-based exploration of similar compound structures. In light of this fact, the experimental success of even low levels of binding and inhibition are encouraging not only for the potential identification of inhibitors for this key heterodimeric interaction, but also more generally for the technique of in silico screening as a method for identifying initial promising inhibitor scaffolds.
VI. DISCUSSION
While a great deal of headway has been made in understanding herpesviral nuclear egress through a combination of techniques, including mutagenesis, biochemistry, cell biology, mass spectrometry, immunocytochemistry, and transmission electron microscopy, the lack of structural information has meant that many of those results could not be put in molecular context. By solving the solution-state NMR structure of M50 (residues 1-168), we begin to address this stumbling block. Our discovery of the novel structural character of the resulting protein fold not only hints at the rather unique nature of this process but also begins to explain previous failures to search for sequence-structure-function relationships (73). Unfortunately, the lack of existing similar structures means that no inferences can be drawn about the mechanistic function of the nuclear egress complex based on structural homology.

While it is highly probable that both components of the nuclear egress complex have multiple functions, some which may even occur simultaneously, as supported by the existence of mutations that render the protein non-functional while maintaining heterodimerization capability (69,131), and even in other areas of the viral lifecycle (44,64,65), one of the most predominantly studied functions of M50 and its homologues is its ability to bind its soluble partner. This interaction has been demonstrated to be necessary for proper localization of the nuclear egress complex and facilitation of nucleocapsid primary envelopment (34,37,68,73,132). In this study, using NMR spectroscopy in combination with biophysical and biological validation, a core binding interface has been mapped onto an area structurally proximal to the C-terminus of M50 (residues 1-168). M50 residues identified as important in NMR titration experiments were shown to have an effect on binding affinity and virus viability when single alanine mutations were introduced into purified M50 (residues 1-168), purified UL50 (residues 1-169), and UL50 in the context of HCMV infection.
A lot of emphasis is given in the literature to mutations in and around E56/Y57 in HCMV and their homologous residues in other herpesviruses. E56 in particular is conserved across the human herpesviruses (Figure III-5). Bubeck et al. showed that when co-transfected into NIH3T3 cells and detected by immunofluorescence, M50 and M53 will localize to the nuclear rim; however, M53 shows a diffuse nuclear distribution when expressed in the absence of its membrane-bound partner. The introduction of mutations E56A or Y57A into M50, result in an M53 phenotype similar to that seen in the absence of M50 (34). An insertion between E56 and Y57 results in a dominant negative phenotype in M50 (37), and Milbradt et al. showed this residue pair to be crucial in UL50 for binding to UL53 (68). Mutations in the homologous residues in HSV-1 (E67 and Y68) have also shown their importance for viral viability (64,73). The studies presented here similarly identified those residues as being a part of the core binding surface as well as correlating their abrogation of heterodimerization with their importance on viral replication.

Beyond these two highly conserved residues, previous studies have also highlighted the importance of other regions in the N-terminal domain as being important for function. Rupp et al. described an insertion between D125 and K126 of M50 that results in a dominant negative phenotype (37), Liang and Baines describe a minimal region in HSV-1 UL34 (residues 137-181) that is sufficient for binding to the M53/UL53 homologue, UL31 (75), and Paßvogel et al. highlight the importance of a dileucine pair L166/L167 (homologous to L161/L162 in M50) through pseudorabies virus UL34-complementation assays (131). A large number of the mutants described in Bubeck et al. that fall between residue 1 and 168 and that are unable to rescue an M50 null phenotype coincide with those highlighted in our NMR titration experiments with
UL53 peptide (Figure IV-2). While all of these mutants are sequentially separate, they all map to a spatially proximal region that is encompassed by our NMR-described interface.

There are, however, some mutations that have been identified in the literature as having an effect on NEC function that fall within the region covered by our construct, but do not fall within the minimal binding surface described by the peptide, such as a dominant negative insertion between Y40 and S41 (37) or non-complementing mutational insertions between K36 and N37 and C43 and D44 (34). These residues fall in a loop between helix $\alpha_2$ and strand $\beta_2$ that is distal to the C-terminus and the UL53 peptide interaction interface. Mutations in comparable regions of other M50 homologues with detrimental effects on virion egress from the nucleus also appear in the literature (69,73), although given the differences that exist even in this more highly conserved portion of the protein, it is impossible to say in anything but the broadest terms (proximity to the N-terminal helix) how those mutations fall structurally without more structural studies on the M50 homologues. It is possible that the phenotypes associated with these residues is indicative of a secondary interaction with another binding partner, such as one of the other viral or host cell proteins hypothesized to play a role in nuclear egress (13,14,19,28,44,68), particularly as the latter two mutations still permit binding to the soluble partner, M53 (34).

Importance has also been attributed to the N-terminal helix in the previous literature, based on the non-functionality of N-terminal truncation mutants lacking 9 or more residues in pseudorabies virus UL34 (76), the failure of a HSV-1 UL34 charged cluster CL02 mutant to complement a UL34 null phenotype (73), and the effect of single alanine mutations in the helix of UL50 on binding to and recruitment of UL53 (68). Milbradt et al. speculate that these effects are due to protein folding and stability rather than direct binding interactions. It is interesting that in our NMR titration studies with M53 (residues 103-333) we see effects primarily on the N-
terminal proximal and central portion of the protein as well as in the region immediately around the conserved E56/Y57 pair discussed above. The mapping of residue perturbations shows effects across the body of the protein. When this result is combined with evidence that the NEC components are more stable as a complex (27), this suggests that an overall shifting of conformation likely occurs upon binding and that mutations, even in the N-terminally proximal region of the structure, can have an effect on overall heterodimer formation and this global conformational adjustment.

Some of the important regions in M50 or its homologues as characterized in the literature are in the latter half of the protein sequence, which is beyond the scope of the solved structure. Primarily the literature focuses on two areas: the area immediately after the end of the M50 (residues 1-168) construct, which in M50 contains the proline rich region (approximately from residue 178 to 207), and the region around the transmembrane domain (TM). Studies have already shown that deletion or mutation in the TM region are detrimental to NEC function (29,34), but that the domain can be substituted with a TM from an alternative inner nuclear membrane protein and NEC function is preserved (70,71) suggesting that membrane targeting ability is more important than actual sequence preservation. In M50, the proline rich region has been implicated in binding the SH3 domain of endophilin-A2 (31) and a deletion of this region results in a failure to reconstitute viral progeny (34). The alpha- and gammaherpesviruses do not have equivalent proline enrichment in this region as conservation across the herpesviruses drops off rapidly in the latter half of the protein (Figure III-5); however, a region from residue 186-220 in HSV-1 UL34 has been implicated in facilitating binding of the NEC to membranes (27).

Yeast two-hybrid assays show binding between UL50, PKC, and p32 (15), both of which have been implicated in lamina disruption (16) and particle release (133). Co-immunoprecipitation
assays with UL50 truncation mutants narrowed the region of interaction with PKCα down to residues 100-280 and the region of interaction with p32 to residues 100-358 (32). Both of these regions extend beyond the domain covered by our structure into sequence that is predicted to be semi- or unstructured. It is possible that binding to these proteins may induce secondary structure in this region or that binding is dependent upon specific linear motifs such as the two proline stretches or the cysteine rich region found in the latter half of the protein sequence. While conservation of sequence is minimal in the latter half of the M50 homologues, the persistence of this additional sequence argues for its function. Given some of the already documented interaction sites, it may be that some of the differences in viral and cellular factor requirements for nuclear egress amongst the different herpesviruses may be attributable to the variability in this latter half of the protein.

Targeting the NEC to the nucleus post-translation requires accommodation by the complex of interaction or signalling events beyond occupation of the heterodimerization interface. Kinetics suggest the components of the NEC traffic to the nucleus separately (77,79), but Schmeiser et al. has suggested that once in the nucleus the canonical nuclear localization signal on the nucleoplasmic partner helps retain the membrane-bound partner, and thus the whole NEC, at the nuclear rim (77). This would suggest that the NLS remains exposed, concurrently with the heterodimeric interaction. The NLS in HCMV UL53 (residues 18-27) (77) and the NLS in MCMV M53 (residues 16 to 106) (36), do fall before the minimal binding domain (UL53 residues 58-85) characterized in Sam et al. (84) and the homologous residues in M53. More structural data about the complex using constructs that include these motifs would help confirm “visibility” of these targeting sequences, and lend further support to the Schmeiser model.
There are a number of studies that identify possible protein interactions with the NEC through immunoprecipitation or similar techniques, but fail to elucidate any potential NEC binding site. Studies by Ye et al. using GST-tagged UL34 show a possible interaction with major capsid protein (a.k.a. ICP5 or VP5) in HSV-1 infection, but the design of the experiments do not allow conclusions to be drawn as to whether the interaction is a direct one or via the other NEC component, UL31 (44). The fact that UL31 is pulled down in addition to major capsid protein does suggest that wherever ICP5 is binding, it is at an interface other than that needed for NEC heterodimer formation. Similarly, other capsid-component interaction studies in HSV-1 by Yang et al. implicate an interaction between capsid components, UL17 and UL25, and UL31 that persists in the presence or absence of UL34, but do not delve into regions or residues of UL31 that may be important for this interaction (26,45). Immunofluorescence co-localization studies also suggest that there may be an interaction between the NEC and UL47 in HSV-1 (47), and again the presence of both components of the NEC suggests that an alternative binding site is needed for interaction with UL47 or that this interaction is indirect. The interaction between MCMV M50 and IRE1 requires the N-terminal domain of the protein (65), which is encompassed by our expression construct, but the interaction site has not been narrowed down further. There have been a number of studies that implicate other viral and cellular proteins in facilitating or regulating nuclear egress, but interactions with the NEC have not yet been demonstrated (24,51,53,62,133-141). Phosphorylation of the NEC by kinases such as US3 (53) and UL97 (62) would necessitate direct contact between these proteins and components of the NEC, presumably in heterodimeric form rather than in its component parts. Not only would the kinases need to be able to bind the NEC, but they would also need access to phosphorylation sites that permit regulation of function, either on an exposed site that allows binding of a
secondary protein, or on a site that has structural effects on the complex as a whole. The HCMV UL50 UL97 phosphorylation site (S216) (62) supports the idea of access over conformational change as it falls outside of the globular N-terminal domain where more structural flexibility may be available. As a facilitator, the NEC needs to help coordinate a number of factors to smoothly transition filled capsids from the nucleus, across the nuclear envelope, into the cytoplasm. Simply given surface area limitations, it is likely that direct interactions with the NEC are part of a larger temporal orchestration of binding events and that the majority of interactions are indirect (i.e. via other interactions with other proteins), which means that many of the factors implicated (often broadly) in the early stages of virion progression from the nucleus to the exterior of the cell may never come in direct contact with the NEC at all.

While nuclear egress is now known to no longer be unique to herpesviruses (80), it so far still seems to be incredibly rare as a cellular mechanism. Our work in combination with previous literature has demonstrated that disruption of binding affinity can be correlated to viral viability. The success of point mutations such as UL50 E56A in rendering HCMV non-viable, suggests that this is an interaction that can be potentially targeted with a small molecule inhibitor. Our initial work to use in silico screening to identify candidate scaffolds that can later be optimized has yielded initial hits that show modest ability to bind the target protein in the region of the defined interface as well as some effect on HCMV titer in cell culture experiments. To improve on these results, chemistry-based optimization of the small molecule scaffolds needs to be done followed by assessment for binding using NMR. This assessment can be done in a higher-throughput fashion using saturation transfer difference spectroscopy, which is also used in NMR fragment-based drug discovery, followed by NMR-based HSQC titration experiments to check for binding at the heterodimeric interface. A biophysical assay such as a fluorescence-
polarization assay or ITC can be used to determine if the small molecule is capable of disrupting the NEC and cell-culture-based assays such as those already performed by Mayuri Sharma can be used to evaluate the effectiveness in the context of viral growth. One would expect that mutagenesis of residues in direct contact with the small molecule either to a much bulkier residue or a residue with significantly different charge character (hydrophobic to charged or vice-versa) would also have an effect on inhibitor binding and activity. Mutations at the inhibitor binding site can be assayed for their effect on $K_D$ via NMR since the experiments using unmutated protein can be used as reference and the use of single components rather than the complex means that mutants that affect NEC formation itself do not become a confounding factor. Ideally residues mutated in the context of the virus and tested in cell culture would be ones showing little to no effect on formation of the heterodimer itself again since otherwise a growth or egress defect could be due to failure of the mutation to affect inhibitor binding or simply the result of disrupting the NEC itself.

The lack of treatment options available for cytomegalovirus and other herpesviruses mean that there is a definite need for targeted therapeutics and the importance of the egress process across the herpesvirus family makes the NEC, which lives at the core of this process, a very attractive therapeutic target. Future work should take advantage of the new structural information presented here in the form of the solution-state NMR structure of M50 (residues 1-168) to not only explore interactions of the NEC with other potential binding partners, but to also pursue further structure-guided development of small molecule inhibitors.
References


102. Zerbe, O. and Jurt, S. *Applied NMR spectroscopy for chemists and life scientists*.


cartesian equations of motion of a system with constraints: molecular dynamics of n-

A smooth particle mesh Ewald method. *The Journal of chemical physics*, 103, 8577-
8593.

and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and
Computation*, 9, 3084-3095.

Identification of conserved amino acids in pUL34 which are critical for function of the

UL31 and UL34 Proteins of Herpes Simplex Virus Type 1 Form a Complex That
Accumulates at the Nuclear Rim and Is Required for Envelopment of Nucleocapsids.
*Journal of Virology*, 75, 8803-8817.

133. Marschall, M., Marzi, A., aus dem Siepen, P., Jochmann, R., Kalmer, M., Auerochs, S.,

simplex virus UL33 gene product is required for the assembly of full capsids. *Virology*,
180, 380-388.


