Virology 537 (2019) 97–103

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Prediction of ligands to universally conserved binding sites of the influenza a virus nuclear export protein

Hershna Patel, Andreas Kukol*

School of Life and Medical Sciences, University of Hertfordshire, United Kingdom

ARTICLE INFO

Keywords:
Influenza a
Nuclear export protein
Sequence conservation
Binding site
Virtual screening

ABSTRACT

The nuclear export protein (NEP) of the influenza A virus exports viral ribonucleoproteins to the host cell cytoplasm following nuclear transcription. In this work conservation analysis of 3000 protein sequences and molecular modelling of full-length NEP identified ligand binding sites overlapping with high sequence conservation. Two binding hot spots were identified close to the first nuclear export signal and several hot spots overlapped with highly conserved amino acids such as Arg42, Asp43, Lys39, Ile80, Gln101 and Val109. Virtual screening with ~43,000 compounds against a binding site showed affinities of up to −8.95 kcal/mol, while ~1700 approved drugs showed affinities of up to −8.31 kcal/mol. A drug-like compounds predicted was ZINC01564229 that could be used as probe to investigate NEP function or as a new drug lead. The approved drugs Nandrolone phenylpropionate and Estropipate were predicted to bind with high affinity and may be investigated for repurposing as anti-influenza drugs.

Importance: The influenza A virus causes respiratory illness in humans and farm animals annually across the world. Antigenic shifts and drifts in the surface proteins lead to genome diversity and unpredictable pandemics and epidemics. The high evolution rate of the RNA genome can also limit the effectiveness of antivirals and is the cause of emerging resistance. From a human health perspective, it is important that compounds identified as potential influenza replication inhibitors remain effective long-term. This work presents results which are based on computational predictions that reveal interactions between available compounds and regions of the influenza A nuclear export protein which display high conservation. Due to a low probability of highly conserved regions undergoing genomic changes, these compounds may serve as ideal leads for new antivirals.

1. Introduction

Influenza A is known to be one of the most prevalent and significant human respiratory viral infections worldwide with previous pandemics resulting in remarkably high fatality rates (Kilbourne, 2006). This is largely due to the zoonotic nature of the virus which enables rapid transmission of new re-assortant strains between species (Reperant et al., 2012). The avian influenza caused by the same virus type A is a major problem for chicken- and geese-farming in various parts of the world (Canavan, 2019). The main form of prevention against influenza A is annual vaccination; however this does not always guarantee extensive protection or control of the virus (Chambers et al., 2015). Therefore treatment with antiviral drugs such as the neuraminidase and matrix protein 2 (M2) inhibitors is heavily relied upon. Since the approval of these drugs, particularly the M2 inhibitors, antiviral resistance due to amino acid mutations in the drug target site has been reported and remains a concern (Hayden and De Jong, 2011). The infectious life cycle requires several functional proteins encoded by eight RNA segments, which take over the host cell, allowing the virus to replicate its genome and suppress the immune response (Bouvier and Palese, 2008). The Influenza A nuclear export protein (NEP) is expressed from alternate splicing of the eighth RNA genome segment. NEP, also known as the Non-structural protein 2 (NS2) is found inside the virus particle and is 121 amino acid residues long. In the virion, it is present bound to the M1 matrix protein through interaction with the prominently surface exposed residue Trp78 (Akarsu et al., 2003). This interaction facilitates the main function of the NEP, which is to export newly synthesized viral ribonucleoproteins from the nucleus to the cytoplasm of infected cells to allow translation of the structural proteins, as well as packaging the genome into progeny virions (Akarsu et al., 2003; Li et al., 2015). This is an essential process during the replication cycle and is mediated by binding with the cellular export molecule chromosome region maintenance protein 1 (CRM1) and its co-factor ranGTP (Neumann et al., 2000). Several other key functions of the NEP have also been reported,

* Corresponding author.
E-mail address: a.kukol@herts.ac.uk (A. Kukol).

https://doi.org/10.1016/j.virol.2019.08.013
Received 10 July 2019; Received in revised form 9 August 2019; Accepted 13 August 2019
Available online 14 August 2019
0042-6822/

such as regulating accumulation of viral RNA during transcription and translation (Robb et al., 2009), interacting with cellular molecules such as ATPase to assist with viral budding (Paterson and Fodor, 2012) and several nucleoporins to enable nucleocytoplasmic transport (Chen et al., 2010; O'Neill et al., 1998).

A complete crystal structure of the NEP is currently unavailable, but experimental data shows that the NEP C-terminus has two helices and dimerises to form a four helix bundle (Akarsu et al., 2003). Another study reporting structural information from experimental data suggests that the protein adopts a flexible and highly mobile conformation, in particular the N-terminal fragment (Lommer and Luo, 2002). The N-terminus is recognised by the nuclear export machinery during replication and contains a leucine rich nuclear export signal (NES) consisting of residues 11–23 (O'Neill et al., 1998). Ser17 and Leu21 in this region were found to be highly conserved and have been proposed to be key residues of the NES (Darapaneni et al., 2009). A second NES signal in a leucine rich region between residues 22 and 45 has also been identified and is proposed to be involved in the export of vRNPs (Huang et al., 2013). In contrast, the C-terminal domain (residues 54–121) for which structural data is available (Akarsu et al., 2003), is highly structured and consists of several hydrophobic residues (Paterson and Fodor, 2012). Previous studies have also shown that the NEP displays high sequence conservation in known and unknown functional areas, indicating that the protein would be an ideal drug target due to less chance of resistance mutations developing (Darapaneni et al., 2009; Warren et al., 2013). Despite evidence showing that the number of influenza A replication inhibitors discovered targeting various viral proteins is rising (Shen et al., 2015), there have been no conclusive reports of inhibitors identified that target the NEP. The objective of this work has focused on investigating highly conserved regions of the nuclear export protein that overlap with potential ligand binding sites, and the identification of drug-like molecules that may bind to those sites.

2. Results

2.1. NEP amino acid conservation

All available full-length NEP sequences (≈3000) of mainly human, swine and avian hosts from all geographic regions were initially dowloaded from the NCBI Influenza Virus Resource (Bao et al., 2008). 889 sequences remained after removing redundant sequences at 98% similarity. The alignment profile based on Valdar's scoring method showed a high level of conservation amongst sequences and the conservation scores for each amino acid are presented in Table 1. Scores ranged from 0.632 at position 89 (lowest) to 1 (highest). Gly30, Leu38 and Arg66 were 100% conserved, whilst other highly conserved residues (> 0.900) include Met16, Ser17, Tyr41, Leu69, Arg84 and Leu103. The most conserved region is from Thr90 to Glu110 which forms helix four at the C-terminus. The least conserved residues (< 0.750) were found to be: Leu14, Glu22, Gly26, Ser57, Asn60, Glu63 and Ile89; several of these residues form the interhelical loops and turns between the helices. Frequent amino acid substitutions observed for positions displaying low conservation include E, M, R, Q, K, R, V, T and A at position 14, and K, T, V, I, N, L, A, S and T at position 89. For display purposes the scores were re-scaled and mapped onto the NEP structure.

2.2. Protein modelling and molecular dynamics simulations

The I-TASSER server generated a full length NEP model based on the H1N1 input sequence. The C-terminal part of the experimental structure 1PD3 was used as the template and the remaining structure was modelled *ab-initio* as no suitable template was available. The RMSD between the C-terminal region of the model and 1PD3 was 1.98 Å, indicating a small difference between the two structures. The experimentally known coordinates of 1PD3 chain A (C-terminal residues 64–115) were combined with the N-terminal part of the I-TASSER model to generate a model consisting of four parallel helices linked by three turns. This model was used as the starting structure for molecular dynamics simulations. From three 100 ns simulation trajectories and cluster analysis, it was observed that the conformation of the N-terminal region (1–65) did not change significantly from the initial model; the experimentally known parts of the structure were kept restrained to their positions during the simulation. The backbone RMSD for the unrestrained part of the protein from each simulation is shown in Fig. 1. The trajectories were analysed to identify clusters of structures sampled starting from 10 ns with an RMSD cut-off of 0.2 nm. Simulation one produced eight clusters, simulation two produced nine clusters and simulation three produced ten clusters. The central structures from cluster one, which was the largest cluster from each simulation were within 3.0 Å of each other (Fig. 2). The RMSF of N-terminal residues was also analysed and the most flexible residues were found to be between positions 22–27 and 53–57.

2.3. Prediction of binding hot-spots

The representative structures from the largest cluster of each simulation (Fig. 2) were analysed with FTMap to identify potential binding hot spot locations. Common hot spots in the hydrophobic region between helix two and four were predicted for all three structures, although with different ranks assigned. The location of hot spot two from structure one was common amongst structures obtained from three simulations, and in a conserved region, therefore it was selected as the target site for virtual screening (Fig. 3).

2.4. Virtual screening

The predicted binding affinities for 42,348 molecules from the NCI library ranged from −8.95 kcal/mol to +20.63 kcal/mol using AutoDock 4. The majority of compounds were found to bind within the range of −4.0 kcal/mol and −5.0 kcal/mol; 46 compounds had a positive binding energy score. The predicted binding affinities ranged from −8.7 kcal/mol to +34.9 kcal/mol using AutoDock Vina. The majority of compounds were found to bind within the range of

Table 1
Conservation score for each amino acid from a multiple sequence alignment of NEP sequences. The scores were obtained from the Jalview AAcons web server using Valdar's scoring method.

Valdar conservation score	Residue position
0.90–1.00	Met1, Asp2, Pro3, Asn4, Thr5, Ser8, Phe9, Gln10, Asp11, Ile12, Leu13, Arg15, Met16, Ser17, Lys18, Met19, Gln20, Leu21, Ser23, Ser24, Ser25, Asp27, Leu28, Asn29, Gly30, Met31, Ile32, Thr33, Phe35, Glu36, Leu38, Lys39, Tyr41, Arg42, Asp43, Ser44, Leu45, Gly46, Glu47, Met50, Arg51, Gly53, Asp54, Leu55, His56, Leu58, Gln59, Arg61, Asn62, Trp65, Arg66, Glu67, Gln68, Leu69, Gln71, Lys72, Phe73, Glu74, Glu75, Ile76, Arg77, Trp78, Leu79, Ile80, Glu82, Arg84, Leu87, Thr90, Glu91, Asn92, Ser93, Phe94, Glu95, Gln96, Ile97, Thr98, Phe99, Met100, Gln101, Ala102, Leu103, His104, Leu105, Leu106, Leu107, Glu108, Val109, Glu110, Glu112, Ile113, Arg114, Phe116, Ser117, Phe118, Gln119, Leu120, Ile121
0.80–0.899	Val6, Ser7, Gln34, Ser37, Leu40, Val49, Met52, Lys64, Glu81, Val83, His85, Lys86, Lys88, Gln111, Thr115,
0.70–0.799	Glu22, Gly26, Ala48, Ser57, Gly70
0.60–0.699	Leu14, Asn60, Glu63, Ile89

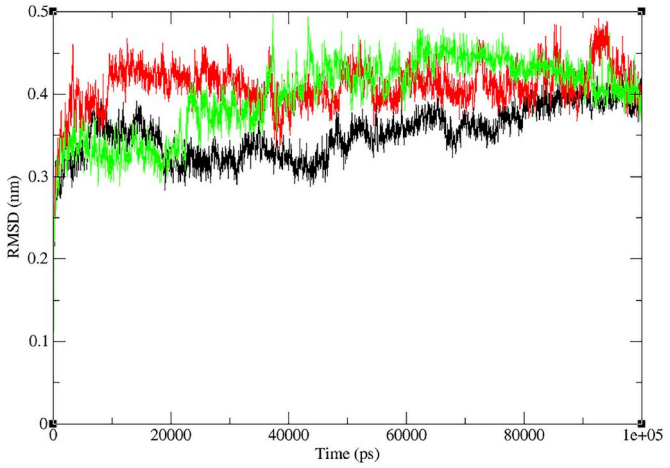

Fig. 1. Root mean square deviation of NEP backbone atoms of residues 1–65 for three 100 ns simulations, relative to the starting structure.

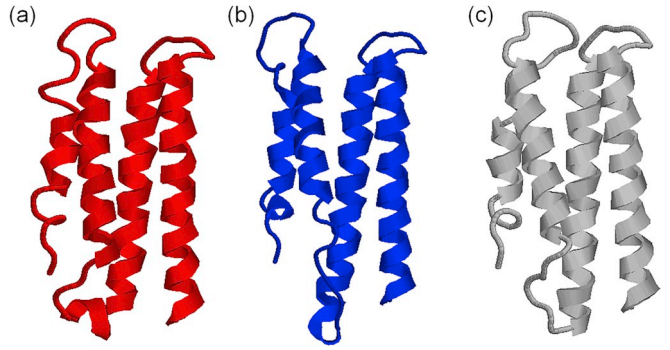

Fig. 2. Backbone representations of the NEP central structure in the largest cluster from (a) simulation one, (b) simulation two and (c) simulation three.

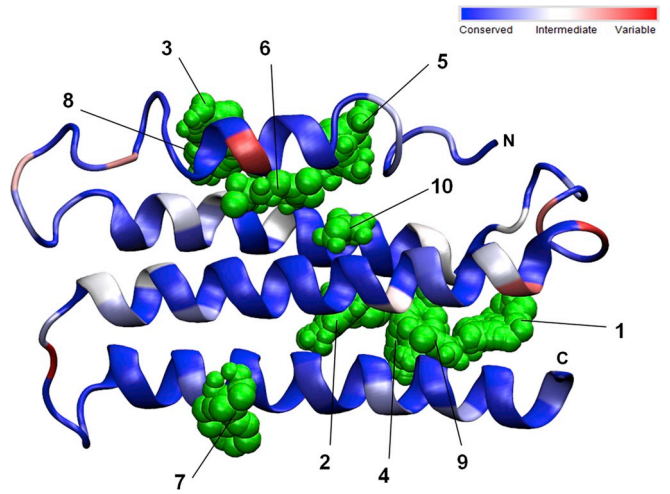

Fig. 3. Ligand binding hot spots (green) identified by FTMap shown together with the degree of NEP conservation for the structure from simulation one. The numbers indicate the site rank assigned by the algorithm.

−4.5 kcal/mol and −5.5 kcal/mol and 17 compounds had a positive binding energy score. Several of the same compounds were identified with positive binding affinities using both software. Common amino acid residues involved in binding some of these compounds through hydrogen bonding and hydrophobic interactions identified using LigPlot+ include Arg42, Asp43, Lys39, Ile80, Gln101, Leu105, Val109; all of which are highly conserved. The compound ZINC01717023 was identified as a top hit molecule with the most similar binding scores from both software. The docking model of compound ZINC01509994 in the NEP target site is shown in Fig. 4a. The binding affinities for 1738 compounds from the DrugBank-approved library ranged from −8.31 kcal/mol to +22.59 kcal/mol using AutoDock 4 and from −7.7 kcal/mol to +7.4 kcal/mol using AutoDock Vina. None of the approved drugs had a significantly stronger predicted binding affinity than the top ranked compound of the NCI library. The top ranked drug from AutoDock Vina was found to be the steroid Nandrolone phenylpropionate (ZINC03881613), while Estropipate was ranked highest with AutoDock 4. The binding affinities and properties of ten top compounds from the rank list are presented in Table 2.

3. Discussion

3.1. NEP conservation

Conservation results show that the influenza A nuclear export protein (NEP) is highly conserved as 112 out of 121 amino acid residues scored above 0.800 based on the Valdar conservation scores from the sequence alignment. The most conserved region with several residues scoring over 0.950 was helix four of the C-terminal, which has been reported to stabilise the structure of helix three by interacting with the M1 protein through intermolecular bonds between side chains (Akarsu et al., 2003). Gly30, Leu38 and Arg66 were found to be 100% conserved and could therefore be most resistant to change due to potential evolutionary adaption of the virus. With regards to known NEP functions, the N-terminal region is reported to contain two leucine rich nuclear export signal (NES) motifs at position 11–23 (O'Neill et al., 1998) and 22–45 (Huang et al., 2013). In this work, the second NES region was found to consist of four highly conserved leucine residues at position 21, 28, 38 and 45. This feature assists recognition and binding of cellular CRM1-RanGTP to the NES (Akarsu et al., 2003) and is a critical stage in transport of vRNPs out of the nucleus. Furthermore, a stretch of conserved hydrophobic residues between position 31–40 (Met31, Phe35, Leu38 and Leu40) of helix two have an effect on the nuclear and cytoplasmic distribution of NEP, as mutation of these residues to alanine resulted in NEP nuclear retention and also reduced virus growth (Huang et al., 2013). The host protein AIMP2 which is a tumour suppressor was found to interact with NEP in the cytoplasm. AIMP2 binds at the NEP N-terminal to positively regulate virus replication by preventing degradation of the M1 protein required for vRNP export (Gao et al., 2015). Additionally, the cellular protein human nucleoporin 98 (hNup98) has been found to interact with NEP of H5N1 subtypes at the 22–53 region which causes NEP to co-localise in the nucleoli, but the significance of this interaction requires further study (Chen et al., 2010). It is suggested that the high conservation of the N-terminal region could be attributable to these functions.

The viral M1 protein binding site on helix three of the C-terminal region involves Trp78 and surrounding glutamate residues which are also highly conserved. This region is recognized by a nuclear localization signal (NLS) on the M1 protein (Akarsu et al., 2003; Shimizu et al., 2011) which in turn is bound to vRNP's to form the export complex. However, mutation of Trp78 was also reported to have no effect on vRNP export and NEP function (Robb et al., 2009). Other highly conserved residues may have a role in maintaining the structure of the N-terminus, and forming recognition sites for other biomolecules.

Regions of low conservation may indicate recognition sites by the host immune system. Sequence variations may also be a result of subtype or species specific codon changes which have circulated and consequently lowered the alignment conservation score. Alanine at position 48 which displays intermediate conservation with a score of 0.791 enables efficient interaction with CRM1 and has been shown to reduce nuclear aggregation of NEP in a H1N1 strain (Gao et al., 2014). Significant differences between sequences could also lead to changes in biological functions and enable the protein with additional features that

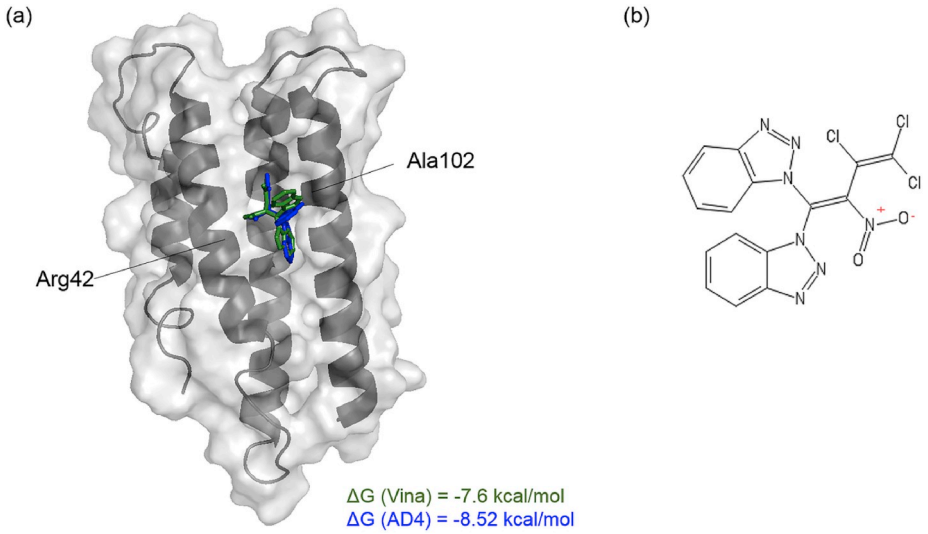

Fig. 4. (a) Docking model of compound ZINC01509994 in the NEP target site predicted with Autodock 4 in blue and Vina in green and (b) the chemical structure of ZINC01509994.

Table 2
Properties of top compounds from the NCI and DrugBank libraries identified by virtual screening including binding affinity (ΔG), number of Hydrogen bond donors (H Don) and acceptors (H Acc), molecular weight and partition coefficient (xLogP) at pH 7 from the ZINC database.

Compound (ZINC ID)	ΔG (kcal/mol) AD4	Vina	xLogP	H Don	H Acc	Mol Wt (g/mol)
NCI library						
ZINC01564229	−8.95	−7.4	−3.26	0	2	360.460
ZINC01717023	−8.09	−8.7	5.29	0	5	417.420
ZINC01509994	−8.52	−7.6	4.31	0	9	436.646
ZINC02035101	−6.9	−8.3	3.48	0	5	357.365
ZINC01592475	−8.10	−6.5	1.63	0	8	345.283
ZINC01646246	−7.34	−8.2	5.62	0	2	332.358
ZINC01717021	−7.07	−8.2	5.29	0	5	417.420
ZINC01561925	−8.02	−7.5	5.53	0	6	406.444
ZINC08617587	−7.93	−6.2	4.21	0	1	270.416
ZINC01645196	−7.83	−6.6	3.57	2	7	405.892
DrugBank library						
Nandrolone -phenylpropionate	−6.83	−7.7	5.51	0	3	406.566
Estropipate	−8.31	−7.1	0.84	0	5	349.428
Pimozide	−4.56	−7.6	5.62	2	4	462.564
Tretinoin	−7.8	−6.0	5.80	0	2	299.434
Ergotamine tartrate	−6.28	−7.6	2.08	3	10	581.673
Adapalene	−7.5	−6.6	7.69	0	3	411.521
Azelastine Hydrochloride	−6.88	−7.5	4.82	1	4	382.915
Tasosartan	−7.50	−7.1	2.48	0	8	410.461
Azelastine hydrochloride	−6.84	−7.5	4.82	1	4	382.915
Zuclopenthixol	−7.33	−6.4	4.69	2	3	401.983

could result in host adaptations. This is exemplified by a compensatory mutation found in the NEP sequence of a human H5N1 isolate at position 16 (M16I) which originated in birds and was found to enhance the activity of avian derived polymerases in human cultured cells (Mänz et al., 2012).

3.2. Structure and predicted binding site locations of the NEP

To investigate binding hot spots as drug target sites in conserved regions, a full length NEP structure is required, which was not available from the PDB, hence protein structure prediction and molecular dynamics simulations were performed. Contrary to an experimental study focusing on NEP structure, the simulation trajectories and RMSD based cluster analyses reveal that the NEP is a stable structure and that the N-terminal does not display significant conformational flexibility (Figs. 1 and 2); a feature which is proposed to assist in formation of the nuclear export complex (Lommer and Luo, 2002).

Compared to the initial starting structure, helix one becomes slightly distorted throughout the 100 ns simulations with the end of the helix unravelling, whereas the structure of helix two mostly remains intact. The distances between the apex of the N-terminal and C-terminal helices also varies a little between simulations, presumably due to higher flexibility of the residues that form the loop connecting helix one and two, opening up a pocket at one end of the protein. Otherwise, no major structural changes occur that alter the predicted four helix bundle structure. Conventional MD simulations of nanosecond time scale provide an insight into the local structural dynamics of proteins, although, they do not provide complete conformational sampling. In the future, enhanced sampling methods, such as accelerated MD (Hamelberg et al., 2004) or metadynamics (Laio and Parrinello, 2002) might be required to achieve a complete picture. Different binding hot spots were predicted for the three largest cluster structures by the FTMap server, despite similar cluster structures being identified from each simulation based on the protein backbone conformation. This was most likely caused by different orientations of amino acid side chains such as those at positions 21–30, which permit different interactions with molecular probes. However, between the three structures, many binding hot spots were predicted in similar locations and close to the same residues. Most hot spots were located in conserved areas of the protein and several sites were predicted close together. Two hot spots were identified near the first nuclear export signal region (residues 11–23), either side of helix one on all three structures. The specific NEP-CRM1 binding location which has also been shown to be independent of the nuclear export signals (Huang et al., 2013; Neumann et al., 2000), may involve N-terminal residues on the outer face of the protein which form these hot spots. One hot spot on all structures was identified in close spatial proximity to residues 74–82 of helix three, which forms the viral M1 protein binding site.

Several other NEP binding proteins have been discovered, such as ATP5E, HINT2, SMC3 and F1F0-ATPase; however, the NEP interaction sites with those proteins are unknown (Gao et al., 2015; Gorai et al., 2012). As interaction with these proteins may have indirect effects on influenza replication, it is possible that their binding locations may coincide with predicted hot spots identified in this work. Those hot spots which are the same across the three structures are presumably important interaction sites with other biomolecules, as they are unaffected by different conformational states. As common hot spots in the

hydrophobic region between helix two and four were predicted for all three structures, and consisted of several highly conserved residues, this part of the protein was selected as the target site for screening. This region was also identified to interact with several different FTMap probes.

3.3. Virtual screening

From the NCI library, compounds such as ZINC01717023 and ZINC01561925, which have comparable high binding affinities and similar docking poses using both AutoDock Vina and AutoDock 4 are predicted to be the best candidate inhibitors. All of the top compounds from the combined screening rank list contain aromatic rings but their overall structures are diverse. The predicted binding modes show that many of the top compounds bend around a cleft on the surface of the protein. Specific molecular interactions were analysed using the software LigPlot+ (Laskowski and Swindells, 2011). The compound ZINC0150994 (Fig. 4) specifically interacts with Arg42 on helix two and Ala102 on helix four via hydrogen bonding, as well as Leu106, Leu105, Val109, Gln101 and Lys39 through hydrophobic interactions. These residues are part of several known interaction motifs, such as the nuclear export signal (residues 22–45) and a postulated region of interactions with the PB2 subunit of the viral polymerase complex (C-terminal helix 2, residues 94–118) (Brunotte et al., 2014).

The AutoDock Vina top compound from the DrugBank library (Nandrolone phenylpropionate) is held in place between helices two and four through hydrophobic interactions only, with several surrounding residues of the target site. The DrugCard entry (DB00984) in the DrugBank database (Law et al., 2014; Wishart et al., 2006) states that this drug targets androgen receptors and can be used to treat haematological disorders, growth failure and Turners syndrome. Estropipate, a form of the hormone estrogen was ranked highest with AutoDock 4. The drug binds to the NEP between the middle part of helices two and four in a vertical orientation and also interacts through several hydrophobic interactions, as well as hydrogen bonding with Arg42. Binding of larger compounds such as ZINC01561925 could partly block recognition of the second nuclear export signal on helix two, or accessibility of other binding proteins to this region.

The overall binding affinity distribution for molecules from the NCI library and DrugBank library are similar between AutoDock 4 and AutoDock Vina. The majority of binding affinities are around −5.0 kcal/mol to the target site. However, the distribution amongst outliers is less comparable between the software for the NCI library, as 716 molecules were identified with a binding affinity between −7.0 kcal/mol and −9.0 kcal/mol with AutoDock Vina, compared to 176 with AutoDock 4. These molecules may have structural features more specific to the target site, although many of these could be false positive predictions. Also, for both libraries, a higher number of molecules with scores between −3.0 kcal/mol and 0 kcal/mol screened with AutoDock 4 are outliers from the bulk of binding affinities compared with AutoDock Vina.

In conclusion, this research has identified specific regions of the nuclear export protein which display the highest levels of sequence conservation and how these regions are implicated in protein function. Also drug-like compounds predicted to bind with strong affinity to a region of the NEP protein consisting of highly conserved amino acid residues have been identified. Due to the low probability of the targeted regions undergoing genetic changes amongst different virus subtypes and hosts, such compounds may remain viable long-term as universal influenza A inhibitors. Future work should be directed towards *in vitro* investigations to verify inhibition of virus replication and whether the predicted compounds could be developed into successful antiviral drugs.

4. Materials and methods

4.1. NEP sequence analysis

Full length NEP sequences from all hosts, regions and subtypes until September 2015 were downloaded from the National Centre for Biotechnology Information (NCBI) Influenza Virus Resource database (Bao et al., 2008). Identical sequences and sequences containing non-standard residues were removed. The CD-HIT web server (Huang et al., 2010) was used to remove redundancy of the sequences at a similarity threshold of 98% as this gave an sizable number of sequences for further analysis. Multiple sequence alignments were performed with the Clustal Omega web server version 1.2.1 (Sievers et al., 2011). The default settings for all parameters remained unchanged whereby the mBed algorithm was used to generate the guide tree and the number of combined guide tree and Hidden Markov Model iterations remained at 0. The level of amino acid conservation was calculated using the Valdar scoring method (Valdar, 2002) and the scores were re-scaled and mapped to the protein structure as previously described (Patel and Kukol, 2017). The minimum score used for re-scaling was 0.632.

4.2. Protein modelling

The crystal structure of the NEP M1 binding domain (PDB ID 1PD3) from a H1N1 sequence (A/Puerto Rico/8/1934) contained missing regions 1–62 at the N-terminus and 117–121 at the C-terminus. Therefore the I-TASSER server (Zhang, 2008) was used to generate a full model of this sequence. The I-TASSER model and chain A of the experimental structure 1PD3 were aligned by structural superposition in the protein structure viewer PyMol using the 'align' command and the coordinates of the alignment were saved as a PDB file. This PDB file was then modified to replace the structural information of residues 64–115 of the model with the experimentally known coordinates of 1PD3 using a text editor. Atom coordinates for residues 86 and 104 from the model were not replaced.

4.3. Molecular dynamics simulations

Coordinates of the full length NEP model were used for MD simulations using Gromacs version 2016 (Van Der Spoel et al., 2005). The AMBER99SB – ILDN force field was selected as it demonstrated high accuracy in a systematic evaluation of eleven MD force fields (Beauchamp et al., 2012) along with the TIP3P water model. The protein was solvated in explicit water in a cubic box with periodic boundary conditions, and the charge of the chemical system was neutralised with five sodium ions and additional NaCl at 100 mM concentration. Energy minimisation with 500 steps of the steepest descent algorithm followed by 100 ps MD simulation with position restraints on the protein non-hydrogen atoms were performed to equilibrate the system. The temperature of the system was set to remain constant at 300 K using the velocity rescaling thermostat with a coupling constant of 0.5 ps and pressure coupling was set at 1.0 atm with the Berendsen barostat. The Particle Mesh Ewald (PME) algorithm for calculating electrostatic interactions was used with the real space cut-off set at 1.0 nm with the Verlet cut-off scheme. For the production run, three replicate partially position restrained simulations over 100 ns (50 million steps) were performed. The same settings for equilibration were used, except the Parinello-Rahman barostat was used for pressure coupling and position restraints were released on N-terminus residues 1–65 by modifying the forces in the posre.itp file as shown in Table 3. The simulations were performed at a time step of 2 fs.

Simulation trajectories were analysed by creating an index file specifying backbone atoms for residues 1–65 to calculate the root mean square deviation for the unrestrained part of the protein. The implementation of position restraints was verified by analysing the root mean square fluctuation for each residue. For cluster analysis the group

Table 3
Implementation of position restraint forces in the x, y, z axis on the non-hydrogen atoms of the NEP protein. A force of 1000 indicates full restraint on an atom and 0 indicates full motion.

Atoms	Residue	fx	fy	fz
1–971	1–62	0	0	0
972–1032	63–65	500	500	500
1033 onwards	66–121	1000	1000	1000

selected for least squares fitting and RMSD calculation was residues 1–65, and the group for output was selected as protein.

4.4. Prediction of binding hot spots

Following cluster analysis of the three MD trajectories, the central structures from the largest simulation clusters were selected. Binding hot spots were identified with the FTMap web server (Brenke et al., 2009)

4.5. Virtual screening

A binding hot spot predicted in the same location between helix two and four on each NEP structure following cluster analysis was selected as the target site for virtual screening. The grid box parameters for docking using AutoDock Vina (Trott and Olson, 2010) and AutoDock 4 (Morris and Huey, 2009) were set using AutoDock Tools as follows, with a grid spacing of 0.375 Å:

center_x = 11.756

center_y = 5.141

center_z = −0.693

size_x = 15.75

size_y = 16.5

size_z = 14.25

exhaustiveness = 12

The compound library used for virtual screening against the target site has been previously described (Patel and Kukol, 2017). Briefly, the chemical structures of all molecules at pH 6–8 were downloaded from the NCI compound library of the ZINC database (http://zinc.docking.org/). A chemically diverse subset of molecules based on structural similarity within a Tanimoto cut-off at 80% was extracted from this library using a script. The compounds were then filtered using the software Openbabel 2.3.1 (O'Boyle et al., 2011) to eliminate compounds with molecular weight over 500 g/mol and partition coefficient over five. Compounds considered as 'frequent hitters' were removed by passing the library through the Pan Assay Interference Compounds filter with the online FAF-Drugs3 (Free ADME-Tox Filtering Tool) program (Baell and Holloway, 2010; Lagorce et al., 2015). A total of 42,348 compounds remained. The DrugBank-approved compound library (version 4.0) containing 1738 drugs between pH 6 and 8 (Law et al., 2014) was downloaded from the ZINC database and also screened against the NEP target site in order to find any FDA approved drugs which may also bind to the NEP. Molecular interactions were analysed with Ligplot + version 1.4.5 (Laskowski and Swindells, 2011).

Data availability

The protein structure, list with predicted binding affinities and scripts used in this work are available from Mendely Data (doi: 10.17632/7bdp3ck9v6.2).

Declarations of interest

None.

Acknowledgements

This work was supported by a PhD studentship of the University of Hertfordshire, United Kingdom to HP and has made use of the University of Hertfordshire's high-performance computing facility.

References

Akarsu, H., Burmeister, W.P., Petosa, C., Petit, I., Müller, C.W., Ruigrok, R.W.H., Baudin, F., 2003. Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). EMBO J. 22, 4646–4655. https://doi.org/10.1093/emboj/cdg449.

Baell, J.B., Holloway, G.A., 2010. New substructure filters for removal of Pan Assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J. Med. Chem. 53, 2719–2740. https://doi.org/10.1021/jm901137j.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., Lipman, D., 2008. The influenza virus resource at the national center for Biotechnology information. J. Virol. 82, 596–601. https://doi.org/10.1128/JVI.02005-07.

Beauchamp, K.A., Lin, Y.S., Das, R., Pande, V.S., 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. J. Chem. Theory Comput. 8, 1409–1414.

Bouvier, N.M., Palese, P., 2008. The Biology of Influenza Viruses. Vaccine 26, D49–D53.

Brenke, R., Kozakov, D., Chuang, G.Y., Beglov, D., Hall, D., Landon, M.R., Mattos, C., Vajda, S., 2009. Fragment-based identification of druggable "hot spots" of proteins using Fourier domain correlation techniques. Bioinformatics 25, 621–627. https://doi.org/10.1093/bioinformatics/btp036.

Brunotte, L., Flies, J., Bolte, H., Reuther, P., Vreede, F., Schwemmle, M., 2014. The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. J. Biol. Chem. 289 20067–20077. https://doi.org/10.1074/jbc.M114.569178.

Canavan, B.C., 2019. Opening Pandora's Box at the roof of the world: landscape, climate and avian influenza (H5N1. Acta Trop. 196, 93–101.

Chambers, B.S., Parkhouse, K., Ross, T.M., Alby, K., Hensley, S.E., 2015. Identification of hemagglutinin residues responsible for H3N2 antigenic drift during the 2014-2015 influenza season. Cell Rep. 12, 1–6. https://doi.org/10.1016/j.celrep.2015.06.005.

Chen, J., Huang, S., Chen, Z., 2010. Human cellular protein nucleoporin hNup98 interacts with influenza A virus NS2/nuclear export protein and overexpression of its GLFG repeat domain can inhibit virus propagation. J. Gen. Virol. 91, 2474–2484. https://doi.org/10.1099/vir.0.022681-0.

Darapaneni, V., Prabhaker, V.K., Kukol, A., 2009. Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. J. Gen. Virol. 90, 2124–2133. https://doi.org/10.1099/vir.0.011270-0.

Gao, S., Wang, S., Cao, S., Sun, L., Li, J., Bi, Y., Gao, G.F., Liu, W., 2014. Characteristics of nucleocytoplasmic transport of H1N1 influenza A virus nuclear export protein. J. Virol. 88, 7455–7463. https://doi.org/10.1128/JVI.00257-14.

Gao, S., Wu, J., Liu, R.-Y., Li, J., Song, L., Teng, Y., Sheng, C., Liu, D., Yao, C., Chen, H., Jiang, W., Chen, S., Huang, W., 2015. Interaction of NS2 with AIMP2 facilitates the switch from ubiquitination to SUMOylation of M1 in influenza A virus-infected cells. J. Virol. 89, 300–311. https://doi.org/10.1128/JVI.02170-14.

Gorai, T., Goto, H., Noda, T., Watanabe, T., Kozuka-Hata, H., Oyama, M., Takano, R., Neumann, G., Watanabe, S., Kawaoka, Y., 2012. F1Fo-ATPase, F-type proton-translocating ATPase, at the plasma membrane is critical for efficient influenza virus budding. Proc. Natl. Acad. Sci. U.S.A. 109, 4615–4620. https://doi.org/10.1073/pnas.1114728109.

Hamelberg, D., Mongan, J., Mccammon, J.A., 2004. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J. Chem. Phys. 120, 11919–11929. https://doi.org/10.1063/1.1755656.

Hayden, F.G., De Jong, M.D., 2011. Emerging influenza antiviral resistance threats. J. Infect. Dis. 203, 6–10. https://doi.org/10.1093/infdis/jiq012.

Huang, S., Chen, Jingjing, Chen, Q., Wang, H., Yao, Y., Chen, Jianjun, Chen, Z., 2013. A second CRM1-dependent nuclear export signal in the influenza A virus NS2 protein contributes to the nuclear export of viral ribonucleoproteins. J. Virol. 87, 767–778. https://doi.org/10.1128/JVI.06519-11.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Kilbourne, E.D., 2006. Influenza pandemics of the 20th century. Emerg. Infect. Dis. 12, 9–14. https://doi.org/10.3201/eid1201.051254.

Lagorce, D., Sperandio, O., Baell, J.B., Miteva, M.A., Villoutreix, B.O., 2015. FAF-Drugs3: a web server for compound property calculation and chemical library design. Nucleic Acids Res. 43. https://doi.org/10.1093/nar/gkv353.

Laio, A., Parrinello, M., 2002. Escaping free-energy minima. Proc. Natl. Acad. Sci. U.S.A. 99, 12562–12566. https://doi.org/10.1073/pnas.202427399.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786. https://doi.org/10.1021/ci200227u.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt,

D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z.T., Han, B., Zhou, Y., Wishart, D.S., 2014. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 42, D1091–D1097. https://doi.org/10.1093/nar/gkt1068.

Li, J., Yu, M., Zheng, W., Liu, W., 2015. Nucleocytoplasmic shuttling of influenza A virus proteins. Viruses 7, 2668–2682. https://doi.org/10.3390/v7052668.

Lommer, B.S., Luo, M., 2002. Structural plasticity in influenza virus protein NS2 (NEP). J. Biol. Chem. 277, 7108–7117. https://doi.org/10.1074/jbc.M109045200.

Mänz, B., Brunotte, L., Reuther, P., Schwemmle, M., 2012. Adaptive mutations in NEP compensate for defective H5N1 RNA replication in cultured human cells. Nat. Commun. 3, 802. https://doi.org/10.1038/ncomms1804.

Morris, G., Huey, R., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791. https://doi.org/10.1002/jcc.21256.AutoDock4.

Neumann, G., Hughes, M.T., Kawaoka, Y., 2000. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. EMBO J. 19, 6751–6758. https://doi.org/10.1093/emboj/19.24.6751.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: an open chemical toolbox. J. Cheminf. 3, 33. https://doi.org/10.1186/1758-2946-3-33.

O'Neill, R.E., Talon, J., Palese, P., 1998. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J. 17, 288–296. https://doi.org/10.1093/emboj/17.1.288.

Patel, H., Kukol, A., 2017. Evolutionary conservation of influenza A PB2 sequences reveals potential target sites for small molecule inhibitors. Virology 509, 112–120. https://doi.org/10.1016/j.virol.2017.06.009.

Paterson, D., Fodor, E., 2012. Emerging roles for the influenza A virus nuclear export protein (NEP). PLoS Pathog. 8, e1003019. https://doi.org/10.1371/journal.ppat.1003019.

Reperant, L.A., Kuiken, T., Osterhaus, A.D.M.E., 2012. Adaptive pathways of zoonotic influenza viruses: from exposure to establishment in humans. Vaccine 30, 4419–4434. https://doi.org/10.1016/j.vaccine.2012.04.049.

Robb, N.C., Smith, M., Vreede, F.T., Fodor, E., 2009. NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. J. Gen. Virol. 90, 1398–1407. https://doi.org/10.1099/vir.0.009639-0.

Shen, Z., Lou, K., Wang, W., 2015. New small-molecule drug design strategies for fighting resistant influenza A. Acta Pharm. Sin. B 5, 419–430. https://doi.org/10.1016/j.apsb.2015.07.006.

Shimizu, T., Takizawa, N., Watanabe, K., Nagata, K., Kobayashi, N., 2011. Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. FEBS Lett. 585, 41–46. https://doi.org/10.1016/j.febslet.2010.11.017.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., Sö Ding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7. https://doi.org/10.1038/msb.2011.75.

Trott, O., Olson, A., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.AutoDock.

Valdar, W.S.J., 2002. Scoring residue conservation. Proteins Struct. Funct. Genet. 48, 227–241. https://doi.org/10.1002/prot.10146.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718. https://doi.org/10.1002/jcc.20291.

Warren, S., Wan, X.F., Conant, G., Korkin, D., 2013. Extreme evolutionary conservation of functionally important regions in H1N1 influenza proteome. PLoS One 8, 1–14. https://doi.org/10.1371/journal.pone.0081027.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J., 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D668–D672. https://doi.org/10.1093/nar/gkj067.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinf. 9, 40. https://doi.org/10.1186/1471-2105-9-40.