Title

Using Common Spatial Distributions of Atoms to Relate Functionally Divergent Influenza Virus N10 and N11 Protein Structures to Functionally Characterized Neuraminidase Structures, Toxin Cell Entry Domains, and Non-influenza Virus Cell Entry Domains

Authors

Arthur Weininger
Weininger Works Incorporated

Susan Weininger
Weininger Works Incorporated

Abstract

The ability to identify the functional correlates of structural and sequence variation in proteins is a critical capability. We related structures of influenza A N10 and N11 proteins that have no established function to structures of proteins with known function by identifying spatially conserved atoms. We identified atoms with common distributed spatial occupancy in PDB structures of N10 protein, N11 protein, an influenza A neuraminidase, an influenza B neuraminidase, and a bacterial neuraminidase. By superposing these spatially conserved atoms, we aligned the structures and associated molecules. We report spatially and sequence invariant residues in the aligned structures. Spatially invariant residues in the N6 and influenza B neuraminidase active sites were found in previously unidentified spatially equivalent sites in the N10 and N11 proteins. We found the corresponding secondary and tertiary structures of the aligned proteins to be largely identical despite significant sequence divergence. We found structural precedent in known non-neuraminidase structures for residues exhibiting structural and sequence divergence in the aligned structures. In N10 protein, we identified staphylococcal enterotoxin I-like domains. In N11 protein, we identified hepatitis E E2S-like domains, SARS spike protein-like domains, and toxin components shared by alpha-bungarotoxin, staphylococcal enterotoxin I, anthrax lethal factor, clostridium botulinum neurotoxin, and clostridium tetanus toxin. The presence of active site components common to the N6, influenza B, and S. *pneumoniae* neuraminidases in the N10 and N11 proteins, combined with the absence of apparent neuraminidase function, suggests that the role of neuraminidases in H17N10 and H18N11 emerging influenza A viruses may have changed. The presentation of E2S-like, SARS spike protein-like, or toxin-like domains by the N10 and N11 proteins in these emerging viruses may indicate that H17N10 and H18N11 sialidase-facilitated cell entry has been supplemented or replaced by sialidase-independent receptor binding to an expanded cell population that may include neurons and T-cells.

Introduction

The ability to identify the functional correlates of structural and sequence variation in proteins is especially critical in evaluating functional changes in emerging pathogens and interacting pathogen systems. Pathogenic influenza A viruses have emerged with expanded tissue preferences, reassortment opportunities with other viral species, and interactions with bacterial species. An avian-origin pathogenic H7N9 influenza A virus has emerged in China that causes severe pneumonia and has adapted to replicate in the human conducting and lower airways of humans [1]. A short period of viral shedding of H5N1 HPA1 influenza A virus indicates that emergent influenzas can reinfect a population of hosts can over several transmission cycles in naive hosts [2]. Bacterial neuraminidases have been found to rescue influenza virus replication from being inhibited by the neuraminidase inhibitor zanamivir [3]. Reassortment between avian and human influenza viruses has been found to be mainly between the matrix and neuraminidase gene segments [4].

South and Central American emergent influenza A viruses H17N10, isolated from bats in Guatemala, and H18N11, isolated from bats in Peru, have highly sequence divergent N10 and N11 proteins that do not process

the artificial substrate methylumbelliferyl-N-acetyl-α-D-neuraminic acid ("MUNANA") [5,6,7]. The N10 and N11 proteins were characterized as "neuraminidase-like" because the components of a functional active site were not identified in the structural reports and the proteins showed no activity by cleavage assays, e.g., MUNANA cleavage. No other N10 or N11 protein cell entry domains were identified in the reports of the x-ray crystal structures of these proteins [5,6,7] which were deposited to the Protein Data Bank. The lack of activity of the N10 and N11 proteins is problematic as the loss of sialidase activity, in the absence of some compensating change, would be expected to reduce the fitness of any influenza A virus that incorporates these proteins.

In this study, we used neuraminidase and non-neuraminidase structures deposited to the Protein Data Bank to interpret the N10 and N11 protein structures. We used the common relative spatial occupancy of atoms in N10 and N11 proteins and functionally validated influenza A, influenza B, and bacterial neuraminidases to superpose the structures. Using the superposed structures, we identified a previously unidentified site in the N10 and N11 proteins containing conserved neuraminidase active site residues. We identified variable loop regions in the N10 and N11 proteins that present residues forming domains associated with cell entry in non-neuraminidase proteins, such as toxins and hepatitis E and SARS viral coat proteins. The absence of demonstrated neuraminidase activity with the presence of new cell entry domain components in N10 and N11 proteins suggest that N10 and N11 protein-containing viruses may enter cells without a functioning sialidase, i.e., by binding to alternative receptors such as ACE2, acetylcholine, and MHC II receptors on an expanded receptive cell population, including cells such as neurons and T-cells.

Results

Spatial Alignment of Structures Using Distributed Common Spatial Occupancy of Atoms

Reported structures of N10 protein [5] ("N10P"), N11 protein [7] ("N11P"), N6 neuraminidase ("N6N") [8], influenza B neuraminidase ("IBN") [9] and a *S. pneumoniae* neuraminidase ("SPN)" [10] were spatially aligned by superposition of main chain oxygen atoms in each of the five structures that we found to have common distributed geometry. The atoms listed in Table 1 were found to have nearly identical spatial distribution in the N10P, N11P, N6N, IBN, and SPN structures. As the residue numbering varies among neuraminidase structures and sequences, a Consensus Numbering System for Atoms ("CNSA") and Residues ("CNSR") were created to relate corresponding atoms and residues in the structures. The N10P, N11P, IBN, and SPN structures, along with any associated non-protein atoms or molecules, were rotated and translated into a common reference orientation by superposition of the corresponding CNSA118, CNSA224, and CNSA276 atoms onto N6N atoms listed in Table 1. Superposition of the corresponding CNSA atoms listed in Table 1 allowed us to examine the spatial overlap of residues of N10P and N11P with residues in structures of neuraminidases whose function was established, i.e. N6N, IBN, and SPN.

Structurally Identical, Structurally Corresponding, and Structurally Divergent Residues

Each residue side chain in the superposed N10P and N11P structures was evaluated to determine if it had identity to, and spatial correspondence with, residues in the N6N or IBN structures. Structurally Identical Residues ("SIRs") are the identical residues at the same relative spatial position. Structurally Corresponding Residues ("SCRs") are different residues in structures with residue atom overlap and the same residue orientation relative to secondary structure. N6N, N10P, N11P, and IBN SIRs and corresponding SPN SIRs and SCRs are listed in Table 2 along with inter-residue offsets, i.e., the residue relative chain positions between successive entries in each column. Cysteines participating in disulfide bridges account for roughly one third (14 of the 44 (32%)) of the SIRs in N6N, N10P, N11P, and IBN shown in Figure 1 and listed in Table 2. SIRs represent the invariant core of the influenza virus neuraminidases in the study sample that is shared by N10P and N11P.

N10P and N11P Alignment With Influenza A, Influenza B, and Bacterial Neuraminidases

Figure 1 shows the annotated and structurally aligned sequences of N6N, IBN, N10P, N11P, and SPN. Sequences of Influenza A N1 [11-14], N2 [15,16], N3 [17], N6 [18], and N9 [19] neuraminidases were aligned by sequence identity to the structurally aligned N6N, N10P, N11P, and IBN residues. Table 3 lists the abbreviations

and descriptions of structures, sequences, and related references used in this study. Sequences used in this study are found in **S1 File**. The N11 [20] ("N11P+") sequence was used in Figure 1 instead of the sequence from N11P as the N11P+ sequence is identical to the N11P sequence with the exception that it contains an additional 10 residues missing from the N11P structure sequence. The use of structure to organize sequence, as illustrated in Figure 1, provides a compact summary of structural variation and its relationship to sequence variation within a class of proteins.

Structure Alignment Reveals Active Site Component Superposition

Despite a high degree of sequence divergence of N10P and N11P relative to other influenza A viruses, we determined that the N10P and N11P structures have shared secondary and tertiary structure with the functionally validated sialidases, N6N and IBN. The superposed structure ribbons of N6N, N10P, N11P, and IBN, with the overlapped CNSA118, CNSA224, and CNSA oxygen atoms, are shown in Figure 2A. Figure 2A shows the active sites of N6N and IBN superposed with N10P and N11P regions containing similar active site components. The CNSRs shown in Figure 2 are listed in Table 4. It can be seen by an inspection of Figure 2B-2E that, while the positions of key residues in the protein chains have not been conserved, many of the same residues are presented from alternative loops and contribute similar components to the N10P and N11P tertiary location superposed with the N6N and IBN active sites. Figure 2 panel E shows N10P with several side chains repositioned slightly to approximate the N6N contacts with the sialic acid. Figure 2C and Figure 2E show that N10P and N11P contain active site structural components common to active sialidases despite their highly divergent sequences. The superposition of N6N, with its sialic acid, and N10P provides a putative positioning of sialic acid relative to the N10P structure. This positioning of sialic acid relative to N10P is shown in Figure 2 panel F. Structural reports [5,6,7] without benefit of common spatial occupancy alignment were unable to identify a site in either N10P or N11P where key components of functional sialidases were present in a similar geometry and in a similar tertiary position of other sialidases.

Loop Switching of Residues in Sialidase Active Sites

Residues D151 and R152 in the D151 loop have historically been considered to be a distinctive feature of influenza A neuraminidases. As shown in Figure 3, and listed in Table 4, the positions of CNSR151, CNSR152, and CNSR178 residues in the protein chain can be different in functioning sialidases, such as N6N (Figure 3A) and SPN (Figure 3D). The SPN structure shows that there is structural precedent for the CNSR152 and CNSR178 having swapped positions in the chain of an active sialidase, i.e., a TRP is found at the chain position normally occupied by ARG, and an ARG is found at the chain position normally occupied by TRP. This switching of relative residue positions between proteins is signaled by negative offset values in Table 2. CNSR152 and CNSR178 residues in "swapped" positions in N10P (Figure 3B) can reach the substrate in a geometry similar to residues in N6N and IBN. The positions of CNSR151, CNSR152, and CNSR178 in N6N, IBN, and SPN illustrate how functional groups on a single loop in one functional neuraminidase are found distributed to different loops in other, functional, neuraminidases. The replacement of an ARG residue by a GLN residue at CSNR152 in N11P (Figure 3C) may be more significant than the swapping of CNSR152 and CNSR178 residues in the protein chain. Although N10P and N11P do not process MUNANA, their role as sialidases is uncertain as tests using exhaustive methods, such as the sialidase testing procedures suggested by Parker et al. [21], have not been reported.

Comparison of N6N, N10P, N11P, IBN, and SPN Structural Features

Figures 4 and 5 show the ribbon structures of N6N, N10P, N11P, IBN, and SPN monomers in a common reference orientation. The consensus active site relative positions, the regions of the proteins that contribute residues to the consensus active site area, and regions of the proteins that are not shared in N6N, IBN, SPN, N10P, and N11P can be seen by an examination of Figures 1, 4, and 5. As can be seen from Figures 4 and 5, the SPN structure has the same basic secondary structure and spatial position of the active site as N6N and IBN – but has several large structure extensions, colored deep teal and positioned in the sequence as shown in Figure 1. The N6N, N10P, N11P, and IBN SIRs and corresponding SPN SIRs and SCRs, listed in Table 2 and shown in Figures 4 and 5 as dark blue spheres, represent the invariant structural core of these related proteins. SPN has

non-cysteine residues in the same conserved positions as the N6N, N10P, N11P, and IBN SIRs forming disulfide bridges; the SPN residues structurally corresponding to these cysteines are also colored dark blue. The corresponding CNSR151, CNSR152, and CNSR178 consensus active site residue spheres are shown in medium blue. The consensus invariant structural cores identified in N6N, N10P, N11P, IBN, and SPN are flanked by variable loop regions ("VLRs") that are concentrated in two locations, which we refer to "Upside VLRs" and "Downside VLRs". Figure 4 shows the proteins with the "Upside" VLR presented. Figure 5 shows the proteins with the "Downside" VLR presented. Light brown spheres, shown in Figures 4A, 4B, 4D, 4F, 5A, 5B, 5D, and 5F, represent domains present on influenza A, influenza B, and bacterial neuraminidase structures; these domains are absent in N10P and N11P.

Variable Loop Regions Contain Domains Found in Non-neuraminidase Proteins

We found multiple, non-neuraminidase domains in the Upside VLRs: Staphylococcal Enterotoxin I ("SEI") [22] in the N10P Upside VLR; hepatitis E2S protein ("E2S") [23] and SARS spike protein [24] ("SARSSP") in the N11P Upside VLR; and substance P ("SUBP") [25] in the N6N Upside VLR. We also found toxin-like domains in N11P Downside VLRs; these toxin-like domains are present in alpha-bungarotoxin [26] ("ABT"), SEI [22], anthrax lethal factor [27] ("ALF"), clostridium botulinum neurotoxin [28] ("CBN"), and tetanus toxin [29] ("TTX"). Figure 4 panel A shows the superposed structures of N6N, N10P, N11P, INB and SPN with Upside VLR residue spheres. Figure 4 panel B shows the N6N Upside VLR residue spheres (colored red and green) with structural and sequence correspondences to SUBP. Figure 4 panel C shows the N11P Upside VLR residue spheres (colored green, orange and brown) with structural and sequence correspondences to E2S. Figure 4 panel C also shows the N11P Upside VLR residue spheres (colored yellow, red, and purple) with structural and sequence correspondences to SARSSP. Figure 4 panel E shows the N10P Upside VLR residue spheres (colored yellow and orange) with structural and sequence correspondences to SEI.

Figure 5 panel A shows the superposed structures of N6N, N10P, N11P, INB, and SPN with Downside VLR residue spheres from each structure. Figure 5 panel C shows the N11P Downside VLR residue spheres (colored magenta, hot, pink, cyan and salmon) with structural and sequence correspondences common to ABT, SEI, ALF, CBN, and TTX. As can be seen by Figure 5B, 5D, 5E, and 5F, this toxin-like domain is not found in N6N, N10P, IBN, or SPN.

SEI Domain in N10P

Figure 6 shows SEI, an N10 tetramer, and a N11 tetramer. In Figure 6, the blue structure ribbon is the N11P tetramer (from 4K3Y.pdb [7]), and the grey tetramer is comprised of N10P monomers translocated onto the N11P tetramer by superposing the CNSA118:O, CNSA224:O, and CNSA276:O atoms, given in Table 1 and located at the vertices of the blue dotted-line triangles. Figure 6 shows two sets of residues that comprise a domain that is common to N10P and SEI. In N10P, the SEI-like domain is highly structured and is distributed across two N10P monomers. For example, the N10P CB carbon atom of a A140 residue (spheres colored orange) fits into the six membered-ring of a W106 residue (spheres colored yellow) in an adjacent monomer. As shown in Figure 7, the SEI domain can be mapped onto a corresponding N10P domain by mapping SEI atoms (W51:NE1, E52:CA, and Q34:O) onto N10P atoms (W106:NE1; E109: CA and S139:O). Atom numbers for the superposed atoms are given in Table 5. This superposition serves as a detailed example for the use of similar tables to be discussed. Figure 7 shows the structural correspondence and orientation of the N10P and SEI proteins in a common reference orientation.

E2S and SARSSP Domains in N11P

N11P Upside VLR residues correspond to residues in the reported structures of E2S [23] and SARS spike protein [24]. The E2S domain is formed from three sets of residues (A428-G433, Y138, and Y159). Movement of a N11P loop containing six E2S-like domain residues (A428-G433) exposes SARSSP-like domain residues (G105-G108, P166-P169, and N401-T403). Figure 8 shows the spatial relationship between the E2S-like and SARSSP-like domains. Figure 9 shows E2S and corresponding N11P Upside VLR residues presented in different

and common reference orientations. The common reference orientation of E2S and N11P residues is achieved by superposing the atoms with common distributed geometry listed in Table 6. Figure 10 shows SARSSP and corresponding N11P residues presented in different and common reference orientations. The common reference orientation of SARSSP and N11P residues is achieved by superposing the atoms with common distributed geometry that are listed in Table 7. The loops containing residues P105-P108 in N11P and residues P469-P472 in the SARSSP are mobile. The P469-P472 residues in SARSSP could easily reposition to bind within a monomer, instead of across monomers as shown in Figure 10.

Toxin Domains in N11P

N11P Downside VLR residues and residues in alpha-bungarotoxin dimers have common local spatial occupancy of residues as shown in Figure 11. Figure 11 shows three monomers of the N11P tetramer in positions A, C, and D. In place of the N11P monomer in the "B" position is a dimer of alpha-bungarotoxin superposed onto the N11P monomer, not displayed, in the "B" position. This alpha-bungarotoxin dimer was superposed onto the N11P monomer in the "B" position using the atoms listed in Table 8. The N11P monomer in the "C" position shows the N11P residues Y413A-S415 moved to a position relative to N11P residues D85-F87 that is the same as the relative position between ABT residues Y54-E56 and ABT residues D29-F31. As the residues in the N11P Downside VLR are flexible, the spatial relationship between the groups of residues is not fixed. As can be seen from Figure 11, there is a strong structural correspondence between the individual N11P domains mapped onto ABT, suggesting that movement of the mobile loops produces the same combined domain structure in N11P and ABT. This set of residues is present in other toxins suggesting its importance. Table 9 lists residue correspondences between N11P, SEI, ABT, ALF, CBN, and TTX. Figure 12 shows that these structurally characterized toxins present similar clusters of N11P Downside VLR residues on mobile loops.

Substance P Domain in N6N

Domains similar to those found in the substance P structure [25] were found in the Upside VLR of N6N. Figure 13 shows substance P and the presentation of a substance P-like domain in N6N. The small (11 amino acids) substance P is highly flexible and multiple N-terminal conformers can occupy the same volume. In order to map substance P domains to N6N domains, substance P residues R365, P366, and K367 (the first, second, and third residues of substance P) have been reoriented to place the substance P atoms in the same configuration as the corresponding residues (R438, P439, and K440) in N6N. All other residues and all main chain atoms in both structures are in the same relative positions as in the the crystal structures. The coordinates for this model-built and translocated structure were output as a pdb file, "WSUBP.pdb", available as supporting information **S2 File**. Figure 13 shows the overlap of the R365, P366, K367, Q369, and Q370 residue spheres of WSUBP.pdb with the corresponding residue spheres of N6N (R438, P439, and K440, Q407, and N408) after superposition using the atoms provided in Table 10.

Conclusions

Using common spatial occupancy of distributed and localized sets of atoms, divergent structures can be aligned and putative functional domains can be identified. Atoms with common distributed spatial occupancy can be used to superpose structures. Once structures are superposed in a common orientation, structural variation within the superposed structures can be identified and interpreted. Superposition of conserved atoms in related structures allows small molecules and atoms complexed with one structure to be mapped to another structure. Superposed bound molecules may suggest putative binding sites in functionally uncharacterized proteins. Once structures are superposed, structural invariants can be identified and used as markers for the alignment of non-homologous structures or sequences. In this way, structures or sequences that have little or no sequence similarity can be aligned. The identification of structural invariants allows the identification of any highly divergent areas of the protein. Proteins outside of a functionally related group may be used to interpret the function of divergent structures.

Proteins within a class, such as viral neuraminidases, can be interpreted using structures outside of the

class such as bacterial neuraminidases. This is a particularly important capability for the assessment of functional changes in emerging viruses. N10P and N11P structural reports could not identify an active site, functioning or disabled, without superposition of the structures. The bacterial neuraminidase structure, when aligned with influenza A and B neuraminidases, provides important information on the natural structural variation in active sites of functioning sialidases. Alternative configurations of residues performing the same function must be considered when evaluating proteins from emergent pathogens or divergent proteins within a class. Changes in geometry of an active site may impact substrate preference and must be considered when assessing divergent enzyme function by substrate processing, especially by artificial substrate processing.

Structural analysis using common spatial occupancy can identify structural features in proteins that may impact function. It is not clear whether N10P or N11P are functional or dysfunctional neuraminidases as this has not been adequately tested and reported. N10P may not be able to bind and process MUNANA as the large size and stereochemistry of MUNANA may prevent the closure of the loop containing the CNSR178 residue (N10P residue W154). Changes in presentation of the CNSR152 and CNSR178 residues would be expected to impact the ability of MUNANA to bind to a neuraminidase. In addition, N10P and N11P display structural features that suggest that, if they are even active, their substrate preference may be altered relative to other viral sialidases as key residue presentation to the active site has features found in bacterial sialidases (e.g. SPN). MUNANA is an artificial substrate consisting of a proto-fluorophore linked to Neu5Ac. S. *pneumoniae* produces three sialidases Nan A, Nan B, and Nan C with different substrate requirements. Nan A is promiscuous (accepting many sialosides), Nan B prefers alpha 2,3 sialosides, and NanC must process a substrate and hydrate it to form Neu5Ac before cleaving it. The Neu5Gc form of sialic acid is produced by many non-human mammals instead of the Neu5Ac (the most common sialic acid in humans) [21]. Alternatives to MUNANA as a substrate as an assay for N10P and N11P must be considered in light of the variation of presentation of active site components on alternate loops in their respective binding pockets.

Regardless of activity, or lack thereof, N10P and N11P have CNS R292T mutations that are expected to confer resistance to oseltamivir, zanamivir, and peramivir. If N10P-like and N11P-like proteins are present in infections with mixed virus populations, failure to identify functioning, drug-resistant viruses could facilitate spread of a resistant virus to health care workers that have prophylactically taken currently available antivirals with the mistaken belief that they are protected. In mixed virus populations, there is the possibility that H17N10 and H18N11 viruses can reassort and rescue drug sensitive viruses or be rescued by resistant bacterial neuraminidases. Resistant bacterial neuraminidases have been found to rescue sensitive influenza viruses from inhibition by the neuraminidase inhibitor, zanamivir [3]. H17N10 and H18N11 viruses should be carefully monitored for this reason. It remains to be determined whether N10P and N11P sites containing active site components are functionally active neuraminidase sites or represent vestigial active sites made obsolete by the incorporation of new cell entry domains.

Our identification of groups of localized residues in N10P and N11P having conserved spatial occupancy with non-influenza protein residues allowed the identification of putative cell entry domains in the N10P and N11P structures. These cell entry domains may be strategic in N10P and N11P in the absence of, change in, or reduction in sialidase activity. Loss of N10P and N11P sialidase activity coupled to the appearance of cell entry domains from bacterial toxins and other viruses is unprecedented.

The strong structural correspondence of SEI domains and N10P domains, even without altering the crystal structure positions of the residues in the corresponding domains to maximize overlap, suggest that H17N10 influenza virus may enter cells by binding to human MHC class II molecules in a manner similar to that of SEI. The SEI proteins bind to human MHC class II proteins and they were co-crystallized with the MHC class II proteins in the crystal structure [22].

The strong structural correspondence of E2S, SARSSP, and toxin-like domains and N11P domains, even without altering the crystal structure positions of the residues in the corresponding domains to maximize overlap, suggest that H18N11 influenza virus may enter cells by binding to an expanded set of human cellular receptors, including ACE2 and acetylcholine receptors. The identification of the similar residue domains in SEI, ABT, ALF,

CBN, and TTX suggests that these domains are important, conserved structures in these toxins. The fact that multiple toxins have similar domains to N11P domains suggests that the H18N11 influenza virus may, at the least, have the structural components necessary to enter cells via acetylcholine receptors. Whether these domains on multiple mobile loops enable viruses containing them to enter cells via the acetylcholine receptor should be investigated.

The strong structural correspondence of substance P residues and N6N residues, after altering the crystal structure positions of three of eleven of the highly flexible substance P residue side chains, suggest that N6N may have the ability to enter cells by binding to tachykinin receptors. The presentation of binding components that can bind simultaneously, as might occur when substance P-like domains are presented by an N6N tetramer, may cause a dramatic increase in binding affinity even if the number of residues in the individual binding domain is small. Multiple small binding domains, presented on an influenza virus, in a geometry where they can bind to more than one receptor simultaneously, would have an overall dramatically increased affinity. If n is the binding affinity of one domain, two domains binding simultaneously and cooperatively would be expected to produce approximately (n^2 - n) binding affinity. For this reason, clusters of atoms that can achieve a similar common spatial occupancy are significant even if the cluster is formed from atoms from small numbers of residues on different loops.

The non-influenza virus-like domains that we have identified in N10P and N11P are important to consider in developing diagnostic antibodies and therapeutic vaccines. The presence of these domains suggests that proteolytically released N11P may possibly be detected by anti-ABT and other toxin-related antibodies.

This method of relating distributed and local common spatial occupancy is general and can be applied to any set of structures, regardless of the size, complexity, particular orientation, distribution, or diversity of local structure of the components. The use of atom sets to superpose structures based on similar relative geometry of atoms is a powerful tool. The method of identifying the common and divergent spatial occupancy of atoms and proteins provides rapid, effective assessment of divergent emerging virus features and provides testable hypotheses about how viral sequence changes are related to viral trait changes such as cellular receptor preference. The method, demonstrated by the analysis of N10P and N11P, is a general method for evaluating proteins.

Materials and Methods

Overview of Method: Common Spatial Occupancy

Common spatial occupancy of atoms in structures consists of sets of atoms from structures having conserved distribution in space. The atoms with the same relative positions can be distributed or localized in one structure (e.g. in a monomer) or between multiple structures (e.g. across a dimer or other intermolecular association). We identified atoms with common spatial distribution and then used these atoms to orient structures (N6N, N10P, N11P, IBN, and SPN) relative to one another. These aligned structures were used to align sequences with no structures (e.g., other Influenza A sequences as shown in Figure 1) and to identify and characterize areas of structural similarity and deviation. Common spatial occupancy was further used to identify structural correlates of structural deviation.

Method for Determination of Common Distributed Relative Spatial Occupancy

The N6N, N11P, N11P, IBN, and SPN structures were reoriented relative to one another by superposition of atoms with common distributed relative spatial occupancy in these structures. The CNSR118:O, CNSR224:O, and CNSR276:O atoms, listed in Table 1, can be identified by calculating the distances between all atom pairs in each structure and then identifying sets of spatially distributed (not the same or contiguous) atoms with identical or near-identical spatial distribution. Figure 14 shows the standard deviation of the interatomic distances between the main chain oxygen atoms of specific residues in Table 2 that are found to be conserved among N6N, N10P, N11P, IBN, and the corresponding atoms in SPN; these residues were selected to be from rows in Table 2 that do not contain cysteines, prolines, or missing residues. S3 File lists, in the order of data calculation, these selected

main chain oxygen atoms, the interatomic distances between these oxygen atoms, and the standard deviation values of the set of these interatomic distances. Figure 14 highlights low standard deviation values (i.e. values under 0.300) in color: yellow by default and cyan and green for standard deviation values of distances between pairs of CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms. Figure 15 shows the spatial presentation of atom pairs that correspond to the cyan and green low standard deviation values in Figure 14. These atoms form a tetrahedron. Superposition of the corresponding CNSA118:O, CNSA224:O, and CNSA276:O atoms, listed in Table 1, is sufficient to place the N6N, N10P, N11P, IBN, and SPN structures into a common reference orientation; the use of the CNSA185:O atom does not improve the result. Figure 16 shows the N6N, N10P, N11P, IBN, and SPN structures oriented into a common spatial reference orientation using their corresponding CNSA118:O, CNSA224:O, and CNSA276:O atoms. Figure 16 also shows the CNSA185:O atom and the tetrahedron formed by the CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms for reference. Figures 14-17 illustrate a method used to select atoms with common spatial distribution that can be used to superpose sequence related and unrelated structures.

Method for Determination of Common Localized Relative Spatial Occupancy

N10P, N11P, and N6N structures were placed into a common reference orientation with non-influenza protein binding domains by superposition of atoms with common localized relative spatial occupancy. Local atomic correspondences can be identified independently from, and without use of, the chain position of corresponding residues. A detailed example of this method is given for tryptophan and glutamic acid residues having different relative chain positions in N10P and SEI. Figure 17 shows the standard deviation of N10P and SEI sets of the distances between TRP and GLU residues. The clustering of low standard deviation values indicates that these residues are in the same spatial position relative to one another in both structures –even though the residues are separated by two residues in N10P and one residue in SEI. **S4 File** contains excerpted atoms, interatomic distances, and standard deviation values presented in Figure 17. This method shows that it is the relative spatial positioning of atoms in residues rather than the chain positions of the residues that determines spatial structural correspondence. This use of common spatial occupancy between clustered atoms can be used to identify similar distributions of atoms, regardless of whether the atoms are contained in one monomer or distributed across any combination of molecules. Sets of atoms with localized common spatial occupancy are not restricted to the same molecule and can be distributed among associated molecules. Examples of localized intermolecular contacts identified are: N11P monomer and ABT dimer atoms presented in Figure 11 and Table 8; and N11P dimer and SARSSP monomer atoms presented in Figure 10 and Table 6. Substrate-protein atom contacts with sialic acid in the N6N-sialic acid complex in the superposed structures can also be evaluated using this method.

Molecular display programs can also be used to superpose molecules. Structural alignment and overlap can also be confirmed visually or by using standard deviation of atom pairs.

Supporting Information

S1 File. Figure Abbreviations, References, Sequence Identifiers, and Sequence Descriptions. List of sequences used, with sources and distribution of sequence groups in Figure 1.

S2 File. "WSUBP.pdb". Coordinates, in pdb format, of 11 residues of substance P with reoriented R1, P2, and K3 side chains.

S3 File. N6N, N10P, N11P, IBN, and SPN Example of Common Distributed Relative Spatial Occupancy. List of selected N6N, N10P, N11P, IBN, and SPN main chain oxygen atoms, interatomic distances, and interatomic distance population standard deviation values, as seen in Figures 14 - 16.

S4 File. N10P and SEI Example of Common Localized Relative Spatial Occupancy. List of atoms in specific TRP and GLU residues in N10P and SEI, interatomic distances, and population standard deviation values, as seen in Figure 17.

References

1. Chan MCW, Chan RWY, Chan LLY, Mok CKP, Hui KPY, et al. (2013) Tropism and innate host responses of a novel influenza A H7N9 virus: an analysis of ex-vivo and in-vitro cultures of the human respiratory track. Lancet Respir. Med. 1(7):534-542. Available: http://dx.doi.org/10.1016/S2213-2600(13)70138-3
2. Wibawa H, Bingham J, Nuradji H, Lowther S, Payne J, et al. (2014) Experimentally Infected Domestic Ducks Show Efficient Transmission of Indonesian H5N1 Highly Pathogenic Avian Influenza Virus, but Lack Persistent Viral Shedding. PloS One 9(1):e83417. Available: http://dx.doi.org/10.1371/journal.pone.0083417
3. Nishikawa T, Shimizu K, Tanaka T, Kuroda K, Takayama T, et al. (2012) Bacterial Neuraminidase Rescues Influenza A Replication from Inhibition by a Neuraminidase Inhibitor. PloS One 7(9):e45371. Available: http://dx.doi.org/10.1371/journal.pone.0045371
4. Schrauwen EJA, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM, Herfst S. (2013) Reassortment between Avian H5N1 and Human Influenza Viruses Is Mainly Restricted to the Matrix and Neuraminidase Gene Segments. PloS One 8(3):e59889. Available: http://dx.doi.org/10.1371/journal.pone.0059889
5. Protein Data Bank ID: 4FVK. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=4FVK Li Q, Sun XM, Li ZX, Liu Y, Vavricka CJ, et al. (2012) Structural and functional characterization of neuraminidase-like molecule N10 derived from bat influenza A virus. Proc. Natl. Acad. Sci. USA. 109:18897-18902. Available: http://dx.doi.org/10.1073/pnas.1211037109
6. Protein Data Bank ID: 4GEZ Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=4GEZ Zhu X, Yang H, Guo Z, Yu W, Carney PJ, et al. (2012) Crystal structures of two subtype N10 neuraminidase-like proteins from bat influenza A viruses reveal a putative active site. Proc Nat Acad Sci USA Nov 13;109(46):18902-8. Available: http://dx.doi.org/10.1073/pnas.1212579109
7. Protein Data Bank ID: 4K3Y. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=4K3Y Tong S, Zhu X, Li Y, Shi M, Zhang J, et al. (2013) New World Bats Harbor Diverse Influenza A Viruses. PLoS Pathog. 9(10):e1003657. Available: http://dx.doi.org/10.1371/journal.ppat.1003657
8. Protein Data Bank ID: 1W1X. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=1W1X Rudino-Pinera E, Tunnah P, Crennell SJ, Webster RG, Laver WG, Garman EF. (2004)
9. Protein Data Bank ID: 1A4G. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=1A4G Taylor NR, Cleasby A, Singh O, Skarzynski T, Wonacott AJ, et al. (1998) Dihydropyrancarboxamides Related To Zanamivir: A New Series Of Inhibitors Of Influenza Virus Sialidases. 2. Crystallographic And Molecular Modeling Study Of Complexes 4-Amino-4H-pyran-6-carboxamides and Sialidase From Influenza Virus Types A and B. J. Med. Chem. 41(6):798-807. Available: http://dx.doi.org/10.1021/jm9703754
10. Protein Data Bank ID: 3H72. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=3H72 Hsiao YS, Parker D, Ratner AJ, Prince A, Tong L. (2009) Crystal structures of respiratory pathogen neuraminidases. Biochem. Biophys. Res. Commun. 380:467-471. Available: http://dx.doi.org/10.1016/j.bbrc.2009.01.108
11. GenBank: ADR32096.1. Available: http://www.ncbi.nlm.nih.gov/protein/ADR32096.1 neuraminidase, partial [Influenza A virus (A/Lyon/1364/2007(H1N1))]. Ottmann M, Duchamp MB, Casalegno JS, Frobert E, Moulés V, et al. (2010) Novel influenza A(H1N1) 2009 in vitro reassortant viruses with oseltamivir resistance. Antivir. Ther. (Lond.) (5):721-726. Available: http://dx.doi.org/10.3851/IMP1576
12. GenBank: ACJ47909.1. Available: http://www.ncbi.nlm.nih.gov/protein/ACJ47909.1 Influenza A virus (A/environment/Qinghai/1/2008(H5N1)). He H, Wu Y, Wang C. (2008) Direct Submission. Submitted (07-NOV-2008) National Research Center for Wildlife Born Diseases, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Datun Road, Chaoyang District, Beijing 100101, PR China.
13. GenBank: AF509109.2. Available: http://www.ncbi.nlm.nih.gov/nuccore/AF509109.2 Influenza A virus (A/Chicken/Hong Kong/873.3/01 (H5N1)). Guan Y, Peiris JSM, Lipatov AS, Ellis TM, Dyrting KC, et al. (2002) Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. Proc. Natl. Acad. Sci. U.S.A. 99(13),8950-8955. Available: http://dx.doi.org/10.1073/pnas.132268999
14. GenBank: AEV53435.1. Available: http://www.ncbi.nlm.nih.gov/protein/AEV53435.1 neuraminidase [Influenza A virus (A/Fukushima/09FY004/2009(H1N1))]. Dapat IC, Dapat C, Baranovich T, Suzuki Y, Kondo H, et al. (2012) Genetic Characterization of Human Influenza Viruses in the Pandemic (2009-2010) and Post-Pandemic (2010-2011) Periods in Japan. PLoS ONE. 7(6):E36455. Available: http://dx.doi.org/10.1371/journal.pone.0036455
15. GenBank: ADG59718.1. Available: http://www.ncbi.nlm.nih.gov/protein/ADG59718.1 neuraminidase [Influenza A virus (A/El Salvador/2-Q226L/1957(H2N2))]. Pappas C, Viswanathan K, Chandrasekaran A, Raman R, Katz JM, et al. (2010) Direct Submission. Submitted (27-APR-2010) WHO Collaborating Center for Surveillance, Epidemiology and Control of Influenza, Influenza Division, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333, USA
16. GenBank: AGW83423.1. Available: http://www.ncbi.nlm.nih.gov/protein/AGW83423.1 neuraminidase [Influenza A virus (A/Djibouti/N09200/2009(H3N2))]. Cosby MT, Pimentel G, Nevin RL, Ahmed SF, Klena JD, et al. (2013) Outbreak of H3N2 Influenza at a US Military Base in Djibouti during the H1N1 pandemic of 2009. PLoS ONE. (12):E 82089. Available: http://dx.doi.org/10.1371/journal.pone.0082089

17. GenBank: AAO62026.1. Available: http://www.ncbi.nlm.nih.gov/protein/AAO62026.1 neuraminidase [Influenza A virus (A/goose/HongKong/27404/78(H5N3))]. Webby RJ, Humberd JL, Krauss SL. (2002) Direct Submission. Submitted (24-DEC-2002) Division of Virology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA

18. GenBank: AAO62070.1. Available: http://www.ncbi.nlm.nih.gov/protein/AAO62070.1 neuraminidase [Influenza A virus (A/quail/Nanchang/4-034/2000(H4N6))]. Webby RJ, Humberd JL, Krauss SL. (2002) Direct Submission. Submitted (24-DEC-2002) Division of Virology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA

19. GenBank: AHA11501.1. Available: http://www.ncbi.nlm.nih.gov/protein/AHA11501.1 neuraminidase [Influenza A virus (A/Zhejiang/DTID-ZJU10/2013(H7N9))]. Chen Y, Cui DW, Zheng SF, Yu F, Yang SG, et al. (2013) Direct Submission. Submitted (21-OCT-2013) State Key Laboratory of Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, 79 Qingchun Road, Hangzhou, Zhejiang 310003, China

20. GenBank: CY125947.1. Available: http://www.ncbi.nlm.nih.gov/nuccore/CY125947.1 Influenza A virus (A/flat-faced bat/Peru/033/2010(H18N11)) neuraminidase-like protein (NA) gene. Tong S, Zhu X, Li Y, Shi M, Zhang J, et al. (2013) New World Bats Harbor Diverse Influenza A Viruses. PLoS Pathog. 9(10):E1003657. Available: http://dx.doi.org/10.1371/journal.ppat.1003657

21. Parker R, Mc Combs JE, Kohler JJ (2012) Sialidase Specificity Determined by Chemoselective Modification Of Complex Sialylated Glycans. ACS Chem. Biol. 2012 7(9):1509-1514. Available: http://dx.doi.org/10.1021/cb300241v

22. Protein Data Bank ID: 2G9H. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=2G9H Fernández MM, Guan R, Swaminathan CP, Malchiodi EL, Mariuzza, RA. (2006) Crystal Structure of Staphylococcal Enterotoxin I (SEI) in Complex with a Human Major Histocompatibility Complex Class II Molecule. J. Biol. Chem. 281:25356-25364. Available: http://dx.doi.org/10.1074/jbc.m603969200

23. Protein Data Bank ID: 3RKD. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=3RKD Tang XH, Yang CY, Gu Y, Song CL, Zhang X, et al. (2011) Structural basis for the neutralization and genotype specificity of hepatitis E virus. Proc. Natl. Acad. Sci. USA. 108(25):10266-10271. Available: http://dx.doi.org/10.1073/pnas.1101309108

24. Protein Data Bank ID: 3SCK. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=3SCK Wu K, Peng G, Wilken M, Geraghty R, Li F. (2007)

25. Protein Data Bank ID: 2KS9. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=2KS9 Gayen A, Goswami SK, Mukhopadhyay C. (2011) NMR evidence of GM1-induced conformational change of Substance P using isotropic bicelles. Biochim Biophys Acta. 1808(1):127-39 Available: http://dx.doi.org/10.1016/j.bbamem.2010.09.023

26. Protein Data Bank ID: 2ABX. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=2ABX Love RA, Stroud RM. (1986) The crystal structure of alpha-bungarotoxin at 2.5 A resolution: relation to solution structure and binding to acetylcholine receptor. Protein Eng. 1(1):37-46 Available: http://dx.doi.org/10.1093/protein/1.1.37

27. Protein Data Bank ID: 1YQY. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=1YQY Shoop WL, Xiong Y, Wiltsie J, Woods A, Guo J, et al. (2005) Anthrax lethal factor inhibition. Proc. Natl. Acad. Sci. 102(22):7958-7963 Available: http://dx.doi.org/10.1073/pnas.0502159102

28. Protein Data Bank ID: 3ZUQ. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=3ZUQ Masuyer G, Stancombe P, Chaddock JA, Acharya KR (2011) Structures of engineered *Clostridium botulinum* neurotoxin derivatives. Acta Cryst. F67:1466-1472 Available: http://dx.doi.org/10.1107/S1744309111034671

29. Protein Data Bank ID: 1DLL. Available: http://www.rcsb.org/pdb/explore/explore.do?structureId=1DLL Emsley P, Fontinou C, Black I, Fairweather NF, Charles IG, et al. (2000) The structures of the H(C) fragment of tetanus toxin with carbohydrate subunit complexes provide insight into ganglioside binding. J. Biol Chem. 275(12):8889-94 Available: http://dx.doi.org/10.1074/jbc.275.12.8889

461 **Table 1:** Conserved atom geometry for N6N, N10P, N11P, IBN, and SPN structures.

Atom Description	PDB	Protein	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 118: O	1W1X	N6	B	1124	ARG	3302	O
	4FVK	N10	B	118	ARG	3169	O
	4K3Y	N11	A	118	ARG	273	O
	1A4G	INB	A	115	ARG	312	O
	3H72	SPN	A	347	ARG	207	O
CNSA 224: O	1W1X	N6	B	1231	ARG	4122	O
	4FVK	N10	B	224	ARG	3974	O
	4K3Y	N11	A	224	ARG	972	O
	1A4G	INB	A	222	ARG	1160	O
	3H72	SPN	A	567	GLY	1957	O
CNSA 276: O	1W1X	N6	B	1283	GLU	4520	O
	4FVK	N10	B	276	GLU	4385	O
	4K3Y	N11	A	276	GLU	1364	O
	1A4G	INB	A	274	GLU	1559	O
	3H72	SPN	A	647	GLU	2569	O

462

463 **Table 2.** Consensus Structural Core Residues and Offsets.

N6N		N11P		N10P		IBN		SPN	
Residue	Offset	Residue	Offset	Residue	Offset	Residue	Offset	Residue	Offset
C 98	0	C 92	0	C 92	0	C 86	0	L 323	0
R 124	26	R 118	26	R 118	26	R 115	29	R 347	24
C 130	6	C 124	6	C 124	6	C 121	6	L 351	4
C 135	5	C 129	5	C 129	5	C 126	5	L 359	8
D 157	22	E 151	22	E 153	24	D 148	22	D 372	13
R 158	1	Q 152	1	R 178	25	R 149	1	R 400	28
W 185	27	W 178	26	W 154	-24	W 176	27	W 373	-27
S 186	1	S 179	1	S 179	25	S 177	1	D 417	44
C 190	4	C 183	4	C 183	4	C 181	4	V 421	4
D 192	2	D 185	2	D 185	2	D 183	2	D 423	2
G 193	1	G 186	1	G 186	1	G 184	1	P 424	1
G 203	10	G 196	10	G 196	10	G 194	10	G 441	17
Y 214	11	Y 207	11	Y 207	11	Y 205	11	Y 540	99
L 230	16	L 223	16	L 223	16	L 221	16	L 566	26
R 231	1	R 224	1	R 224	1	R 222	1	G 567	1
S 235	4	S 228	4	S 228	4	S 226	4	T 572	5
C 237	2	C 230	2	C 230	2	C 228	2	I 574	2
C 239	2	C 232	2	C 232	2	C 230	2	L 576	2
G 242	3	G 235	3	G 235	3	G 233	3	G 583	7
C 244	2	C 237	2	C 237	2	C 235	2	I 585	2
D 250	6	D 243	6	D 243	6	D 241	6	Y 590	5
G 251	1	G 244	1	G 244	1	G 242	1	T 591	1
G 267	17	G 260	16	G 260	16	G 258	16	G 613	22
E 283	16	E 276	16	E 276	16	E 274	16	E 647	34
C 287	4	C 280	4	C 280	4	C 278	4	V 650	3
C 296	9	C 289	9	C 289	9	C 288	10	L 660	10
R 299	3	T 292	3	T 292	3	R 291	3	R 663	3
R 307	8	R 300	8	R 300	8	R 299	8	-	-
P 308	1	P 301	1	P 301	1	P 300	1	G 664	1
C 325	17	C 318	17	C 318	17	C 317	17	V 679	15
D 331	6	D 324	6	D 324	6	D 323	6	D 684	5
R 334	3	R 327	3	R 327	3	R 326	3	R 687	3
C 343	9	C 338	11	C 338	11	C 336	10	V 696	9
G 355	12	G 348	10	G 348	10	G 346	10	-	-
G 358	3	G 351	3	G 351	3	G 349	3	-	-
W 368	10	W 361	10	W 361	10	W 363	14	Y 710	14
R 378	10	R 406	45	R 406	45	R 373	10	R 721	11
G 380	2	G 373	-33	G 373	-33	G 375	2	G 724	3
Y 412	32	F 277	-96	Y 277	-96	Y 408	33	Y 752	28
S 413	1	S 407	130	S 407	130	S 409	1	N 753	1
C 425	12	C 417	10	C 417	10	C 419	10	L 763	10
C 429	4	C 421	4	C 421	4	C 423	4	L 766	3
E 433	4	E 425	4	E 425	4	E 427	4	E 768	2
C 455	22	C 447	22	C 447	22	C 446	19	F 781	13
W 466	11	W 458	11	W 458	11	W 455	9	W786	5

464

Table 3. Abbreviations, Structure Sources, and Sequence Sources.

Abbreviation	Ref.#	Sequence ID	Description
ABT	[26]	2ABX.pdb	alpha-bungarotoxin complexed to acetylcholine receptor
ALF	[27]	1YQY.pdb	anthrax lethal factor fragment
CBN	[28]	3ZUQ.pdb	clostridium botulinum neurotoxin type b
CNSR			consensus numbering system for residues (as found in 4FVK)
E2S	[23]	3RKD.pdb	hepatitis E virus E2S domain genotype I (complexed with a neutralizing Ab)
FIG2COL			coloring used in Figure 2 spheres
IBN	[9]	1A4G.pdb	neuraminidase [Influenza B virus B/Beijing/1/87] complexed with zanamivir
N10P	[5]	4FVK.pdb	N10 protein derived from bat influenza A virus fragment
N11P	[7]	4K3Y.pdb	N11 protein of A/flat-faced bat/Peru/033/2010(H18N11)
N11P+	[7,20]	4K3Y.pdb+	4K3Y.pdb [7] with 10 missing residues 139-148 from 1259-11 [19]
N6N	[8]	1W1X.pdb	neuraminidase duck subtype N6 complex with sialic acid (NANA, NEU5AC)
SARSSP	[24]	3SCK.pdb	SARS spike protein receptor-binding domain
SEI	[22]	2G9H.pdb	staphylococcal enterotoxin I (SEI) chain D and human MHC II molecule
SPATIAL			residue correspondence and coloring used in Figures 4-13
SPN	[10]	3H72.pdb	streptococcus pneumoniae D39 neuraminidase A precursor with NANA
SPN<>			residue loops not included in 'SPNSEQ' row
SPNSEQ			SPN sequence without loop regions
SPNRES#			SPN residue numbering
SUBP	[25]	2KS9.pdb	substance P with NK1R, substance P receptor tachykinin receptor 1
TTX	[29]	1DLL.pdb	receptor binding fragment H(C) of clostridium tetanus toxin
1150-09	[19]	AHA11501.1	neuraminidase [Influenza A virus (A/ZhejianG/DTID-ZJU10/2013(H7N9))]
1259-11	[20]	CY125947.1	N11 protein [Influenza A virus (A/bat/Peru/033/2010(H18N11))]
3209-01	[11]	ADR32096.1	neuraminidase [Influenza A virus (A/Lyon/1364/2007(H1N1))]
4790-01	[12]	ACJ47909.1	neuraminidase [Influenza A virus (A/environment/Qinghai/1/2008(H5N1))]
5091-01	[13]	AF509109.2	neuraminidase [Influenza A virus (A/Chicken/Hong Kong/873.3/01 (H5N1))]
5343-01	[14]	AEV53435.1	neuraminidase [Influenza A virus (A/Fukushima/09FY004/2009(H1N1))]
5971-02	[15]	ADG59718.1	neuraminidase [Influenza A virus (A/El Salvador/2-Q226L/1957(H2N2))]
6202-03	[16]	AAO62026.1	neuraminidase [Influenza A virus (A/Goose/HonGKonG/27404/78(H5N3))]
6207-06	[17]	AAO62070.1	neuraminidase [Influenza A virus (A/quail/NanchanG/4-034/2000(H4N6))]
8342-02	[18]	AGW83423.1	neuraminidase [Influenza A virus (A/Djibouti/N09200/2009(H3N2))]

467 **Table 4.** Consensus Numbering System for Key Residues (CNSR).

CNSR #	Protein	PDB	Residue #	Residue
118	N6N	1W1X	124	ARG
118	N10P	4FVK	118	ARG
118	N11P	4K3Y	118	ARG
118	IBN	1A4G	115	ARG
118	SPN	3H72	347	ARG
151	N6N	1W1X	157	ASP
151	N10P	4FVK	153	GLU
151	N11P	4K3Y	151	GLU
151	IBN	1A4G	148	ASP
151	SPN	3H72	372	ASP
152	N6N	1W1X	158	ARG
152	N10P	4FVK	178	ARG
152	N11P	4K3Y	152	GLN
152	IBN	1A4G	149	ARG
152	SPN	3H72	366	ARG
178	N6N	1W1X	185	TRP
178	N10P	4FVK	154	TRP
178	N11P	4K3Y	178	TRP
178	IBN	1A4G	176	TRP
178	SPN	3H72	373	TRP
276	N6N	1W1X	283	GLU
276	N10P	4FVK	276	GLU
276	N11P	4K3Y	276	GLU
276	IBN	1A4G	274	GLU
276	SPN	3H72	647	GLU
277	N6N	1W1X	412	TYR
277	N10P	4FVK	277	TYR
277	N11P	4K3Y	277	PHE
277	IBN	1A4G	408	TYR
277	SPN	3H72	752	TYR
292	N6N	1W1X	299	ARG
292	N10P	4FVK	292	THR
292	N11P	4K3Y	292	THR
292	IBN	1A4G	291	ARG
292	SPN	3H72	663	ARG
406	N6N	1W1X	378	ARG
406	N10P	4FVK	406	ARG
406	N11P	4K3Y	406	ARG
406	IBN	1A4G	373	ARG
406	SPN	3H72	721	ARG

468

469 **Table 5.** Conserved atom geometry for SEI and N10P structures.

Atom Description	PDB	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 106: NE1	4FVK	A	106	TRP	207	NE1
	2G9H	D	51	TRP	3509	NE1
CNSA 109: CA	4FVK	A	109	GLU	225	CA
	2G9H	D	53	GLU	3524	CA
CNSA 139: O	4FVK	A	139	SER	465	O
	2G9H	D	34	GLN	3373	O

470

471 **Table 6.** Conserved atom geometry for Hep E E2S and N11P structures.

Atom Description	PDB	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 138: CB	4K3Y	A	138	TYR	439	CB
	3RKD	A	557	TYR	754	CB
CNSA 431: CA	4K3Y	C	431	LYS	8001	CA
	3RKD	A	554	LYS	733	CA
CNSA 430: CB	4K3Y	C	430	THR	7997	CB
	3RKD	A	553	THR	729	CB

472

473 **Table 7.** Conserved atom geometry for SARSSP and N11P structures.

Atom Description	PDB	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 107: O	4K3Y	C	107	THR	5637	O
	3SCK	F	486	THR	12385	O
CNSA 403: O	4K3Y	C	403	THR	7764	O
	3SCK	F	425	THR	11880	O
CNSA 167: N	4K3Y	D	167	PRO	8738	N
	3SCK	F	470	PRO	12247	N

474

475 **Table 8.** Conserved atom geometry for Alpha-bungarotoxin and N11P structures.

Atom Description	PDB	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 30: O	4K3Y	D	185	ASP	8885	O
	2ABX	A	30	ASP	222	O
CNSA 55: OE(1,2)	4K3Y	D	414	GLU	10585	OE1
	2ABX	A	55	GLU	407	OE2
CNSA 20: CA	4K3Y	D	89	GLU	8224	CA
	2ABX	B	20	GLU	681	CA

476

477 **Table 9.** Toxin and Lethal Factor Domains Found in N11P.

Domain	Fig.12 Panel	PDB	Chain	Residue	Residue	Residue	Residue	Residue
Domain I (cyan)	A	4K3Y	A	(G)	E 89	N 90	L 91	
	B	2G9H	D	(D)	(K)	N 26	L 27	
	C	2ABX	A	(G)	E 20	N 21	L 22	
	D	1YQY	A	(G)	S 722	N 723	L 724	
	E	3ZUQ	A	(I)	(K)	N 385	L 386	
	F	1DLL	A	(F)	N 1219	N 1200	L 1221	
Domain II (salmon)	A	4K3Y	A	Y 414	E 415	S 416		
	B	2G9H	D	Y 195	E 196	D 197		
	C	2ABX	B	Y 54	E 196	E 197		
	D	1YQY	A	Y 650	E 651	Q 652		
	E	3ZUQ	A	Y 421	E 422	E 423		
	F	1DLL	A	Y 1258	D 1259	D 1260		
Domain III (deep pink)	A	4K3Y	A	(C)	S 125	D 126	K 127	E 128
	B	2G9H	D	S 109	T 110	D 111	K 112	(I)
	C	2ABX	A	S 61	T 62	D 63	K 64	(C)
	D	1YQY	A	(L)	(L)	D 701	K 702	N 703
	E	3ZUQ	A	(I)	S 401	D 402	K 403	D 404
	F	1DLL	A	(L)	K 1295	D 1296	K 1297	(I)
Domain IV (magenta)	A	4K3Y	A	D 185	G 186	F 187		
	B	2G9H	D	D 63	I 64	F 65		
	C	2ABX	B	D 30	A 31	F 32		
	D	1YQY	A	D 716	I 718	F 719		
	E	3ZUQ	A	D 488	I 489	F 490		
	F	1DLL	A	D 1187	S 1188	F 1189		

478

479 **Table 10.** Conserved atom geometry for Substance P and N6N structures.

Atom Description	PDB	Chain	Residue #	Res.Type	Atom #	Atom Type
CNSA 2438: CA	1W1X	C	2438	ARG	8715	CA
	WSUBP	B	365	ARG	5869	CA
CNSA 2439: CA	1W1X	C	2439	PRO	8726	CA
	WSUBP	B	366	PRO	5895	CA
CNSA 2440: CA	1W1X	C	2440	LYS	8733	CA
	WSUBP	B	367	LYS	5909	CA

480

481 **Figure 1.** Structural alignment of N10P and N11P with influenza and bacterial neuraminidases.

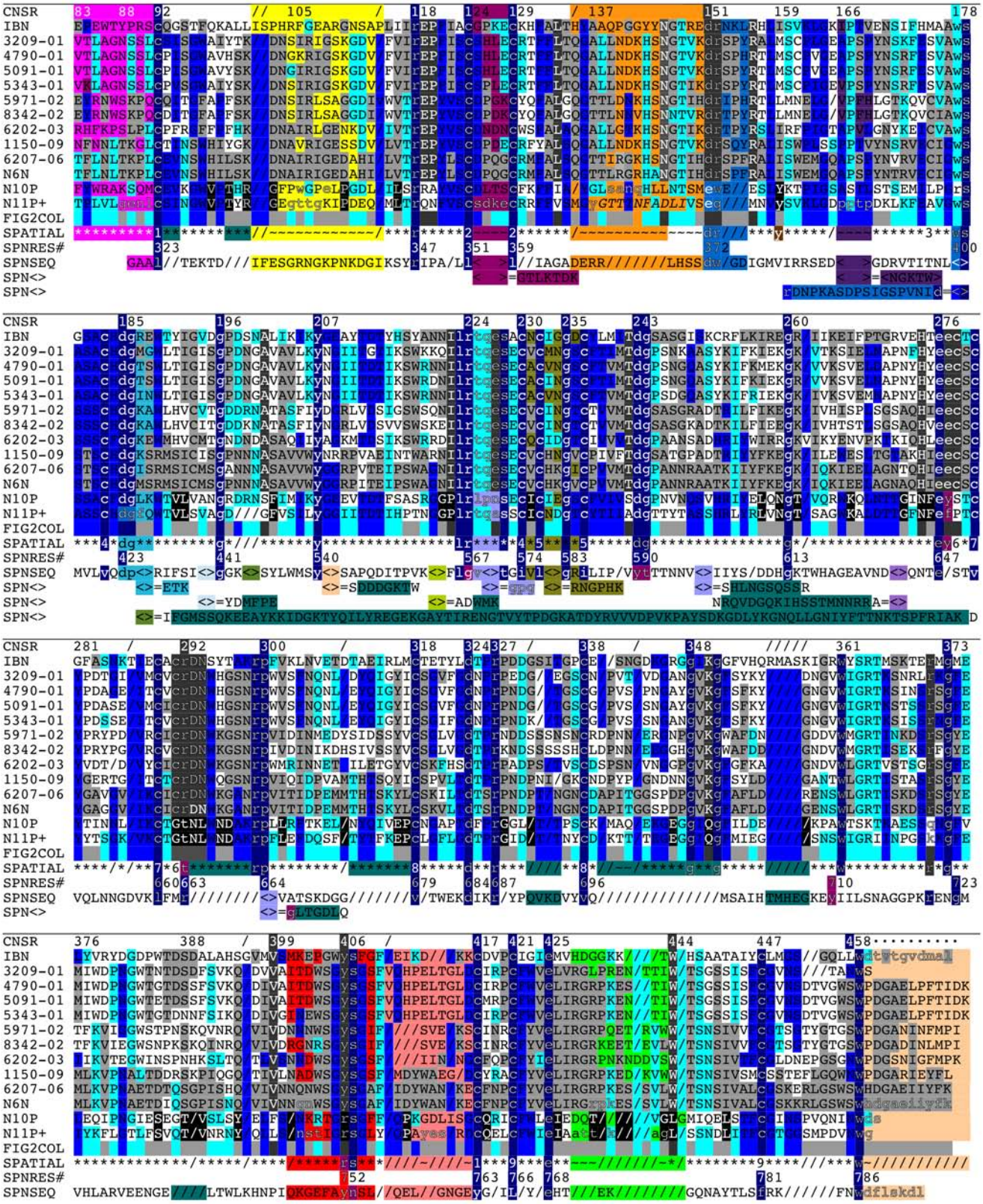


Figure 1 (previous page). Structural alignment of N10P and N11P with influenza and bacterial neuraminidases.

Figure 1. Sequences of N6N, N10P, N11P, IBN, and SPN are shown aligned by common spatial occupancy of residues in their superposed structures. Other influenza A sequences are shown aligned by sequence identity to the structurally aligned N6N, N10P, N11P, and INB residues. Structurally invariant residues above the 'SPATIAL' row and spatially corresponding SPN residues below the 'SPATIAL' row are shaded dark blue. If both the N10P and N11P residues are present in the rest of the column above the 'SPATIAL' row, then the corresponding residues are shaded medium blue. If either, but not both, of the N10P or N11P residue are present in the rest of the column above the 'SPATIAL' row, then the corresponding residues are shaded light blue. If the non-N10P, non-N11P residues are identical to each other but do not match either N10P or N11P, then the non-N10P, non-N11P residues are shaded dark grey. If either of the non-N10P/N11P residues match either IBN or N6P but do not match either N10 or N11, then the residue is shaded light grey. If the N10P and N11P residues are identical to each other but do not match any other residue in that column, then the N10P and N11P residues are shaded black. Upside VLR residues, not shaded as above, are shaded yellow, orange, brown, purple, red, green, and light brown, according to position in the protein and this color is also shown in that column in the 'SPATIAL' row. Residues shaded deep teal are residues in protein loops that deviate spatially from structure common to N6N, N10P, N11P, IBN, and SPN. In the 'SPNSEQ' row, "<>" means that there is an insertion of additional residues that are listed after the "<>=" in the 'SPN<>' row(s) directly below the 'SPNSEQ' row. In the 'SPATIAL' row, the symbol: "*" means spatially conserved, "/" means missing, "~" means not spatially conserved; and a number indicates corresponding cysteines in disulfide bridges. Lowercase residues represent residues shown as spheres in Figs. 4-13.

503 **Figure 2.** Consensus active site components of N6N, N10P, N11P, and IBN.

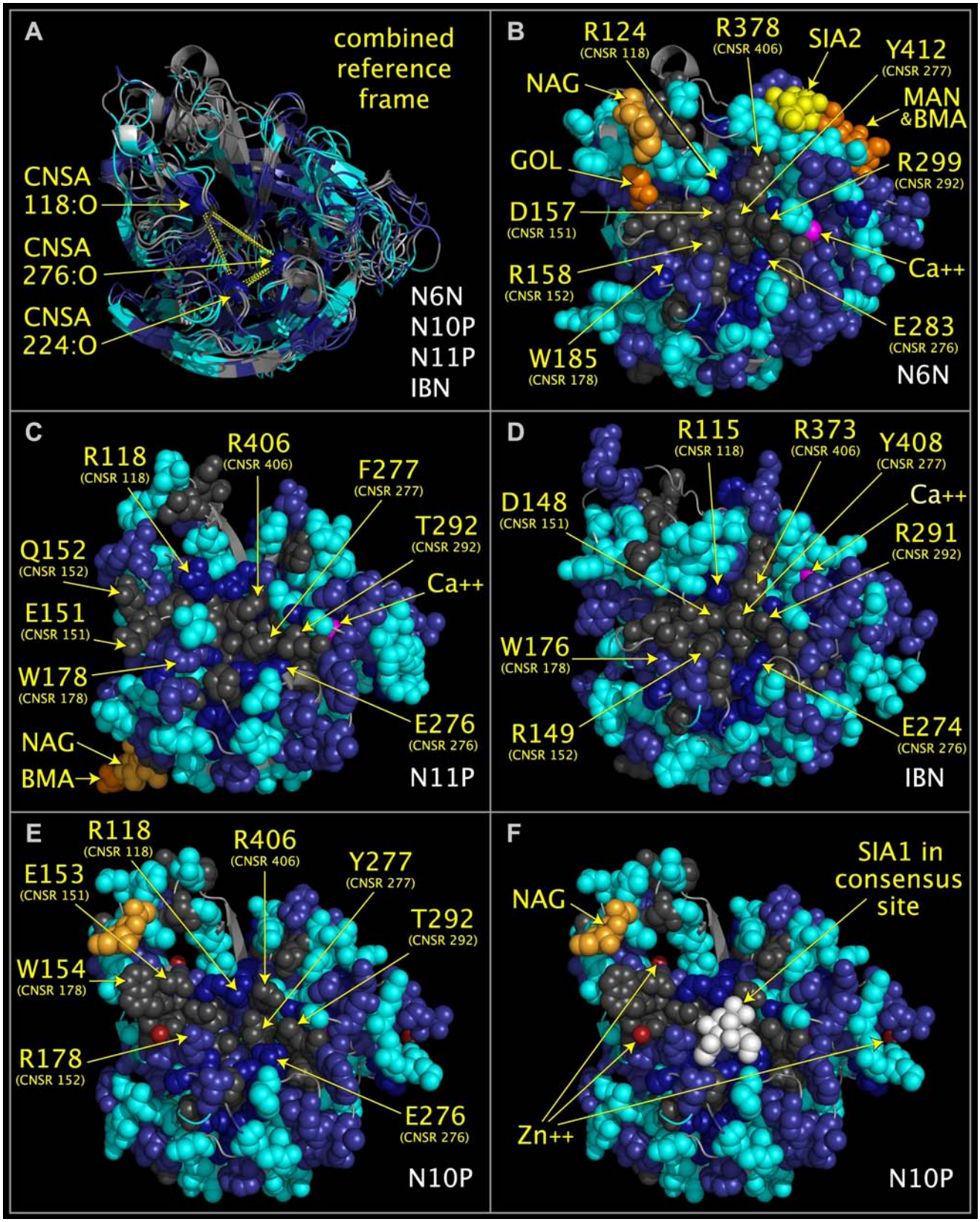
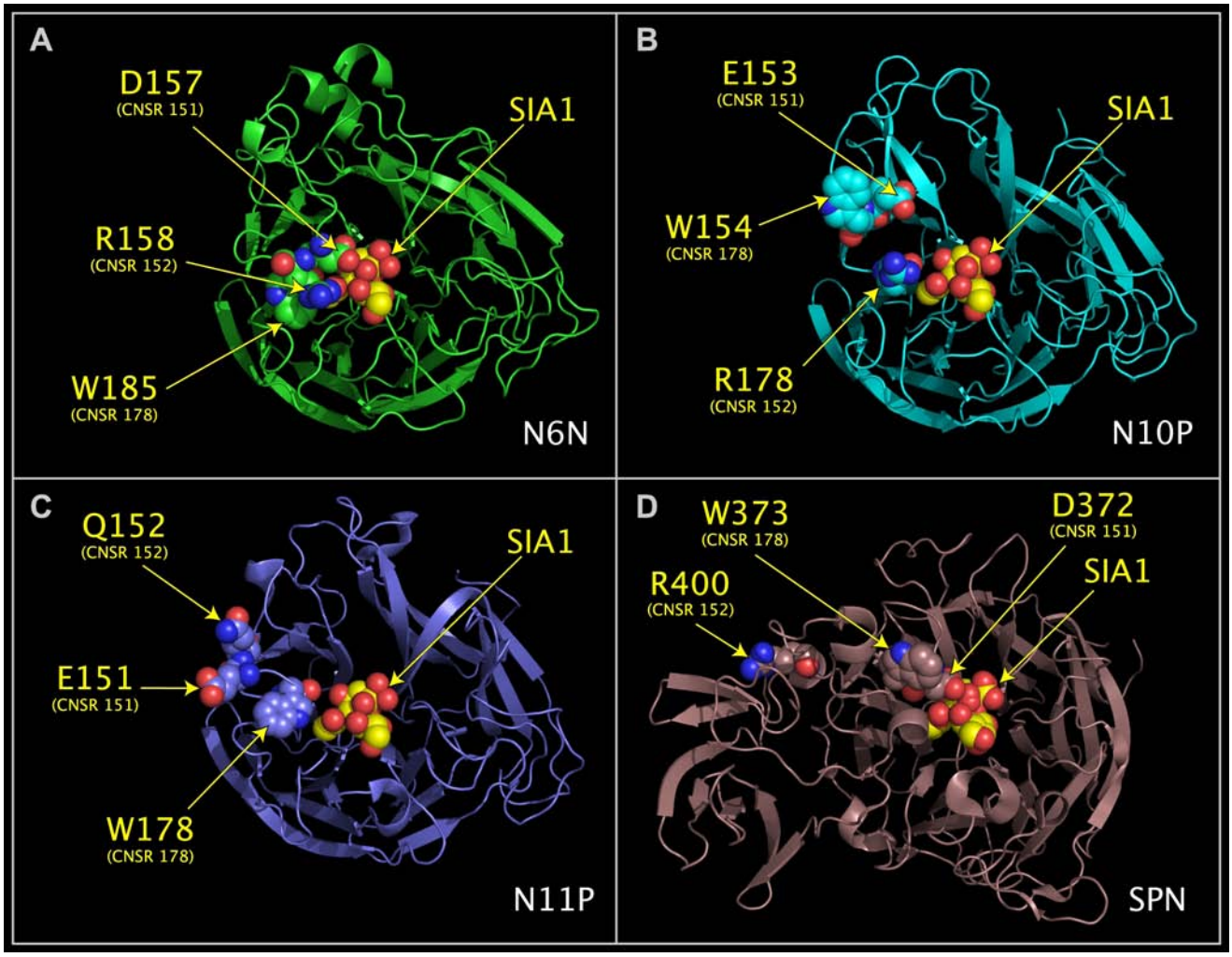


Figure 2 (previous page). Consensus active site components of N6N, N10P, N11P, and IBN.

Figure 2. Structure ribbon and residue spheres are color-coded at each structurally aligned position as in the 'FIG2COL' row in Fig. 1. Fig. 2A shows structure ribbons representing influenza A N10P [1], N6N [2], N11P [3], and IBN [4]) structures superposed using CNSA118:O, CNSA224:O, and CNSA276:O atoms. Figs. 2B-2F show structure ribbons and residues spheres of N6N (Fig. 2B), N11P (Fig. 2C), IBN (Fig. 2D), and N10P (Figs. 2E-2F). Fig. 2F also shows sialic acid (white spheres) positioned relative to N10P by its superposition onto N6N, which was crystalized with sialic acid in its binding pocket. Fig. 2E and 2F residues side chains in the area of the sialic acid have been repositioned slightly from the crystal structure positions to approximate the positions of the corresponding superposed N6N residues but no other side chain or main chain atoms have been moved.

514 **Figure 3.** Loop swapping of CNSR151, CNSR152, and CNSR178.



516 Figure 3. The structure ribbons of N6N (Fig. 3A), N10P (Fig. 3B), N11P (Fig. 4C), and SPN (Fig. 4D) are shown
517 individually in their superposed positions. N6N residues D151, R152, and W178, shown in Fig. 4A, correspond to:
518 N10P residues E153, R178, and W154, shown in Fig. 4B; N11P residues E151, Q152, and W178, shown in Fig.
519 4C; and SPN residues D372, R400 and W373, shown in Fig. 4D.

520 **Figure 4.** Consensus active site components and Upside VLR domains of N6N, N10P, N11P, IBN, and SPN.

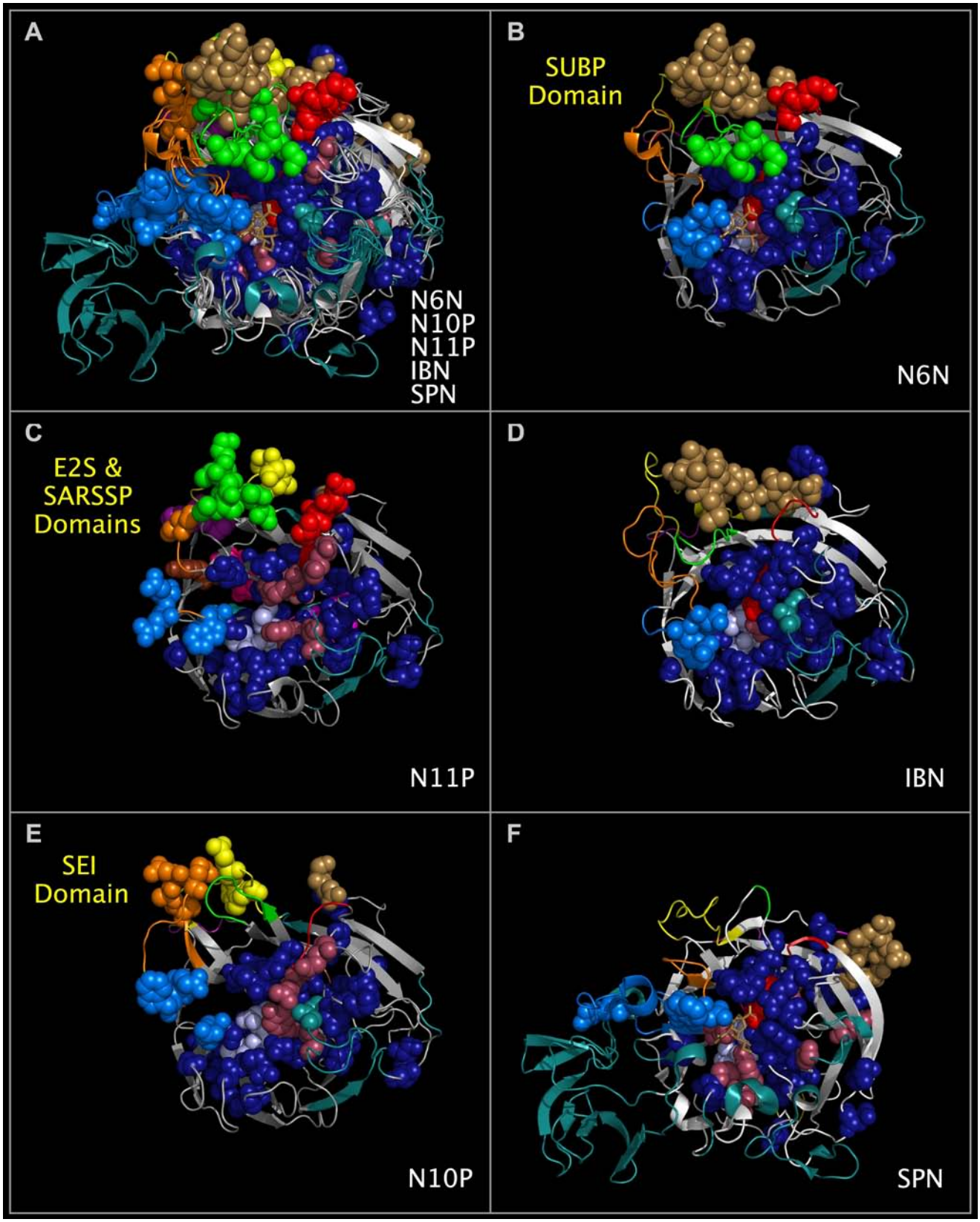


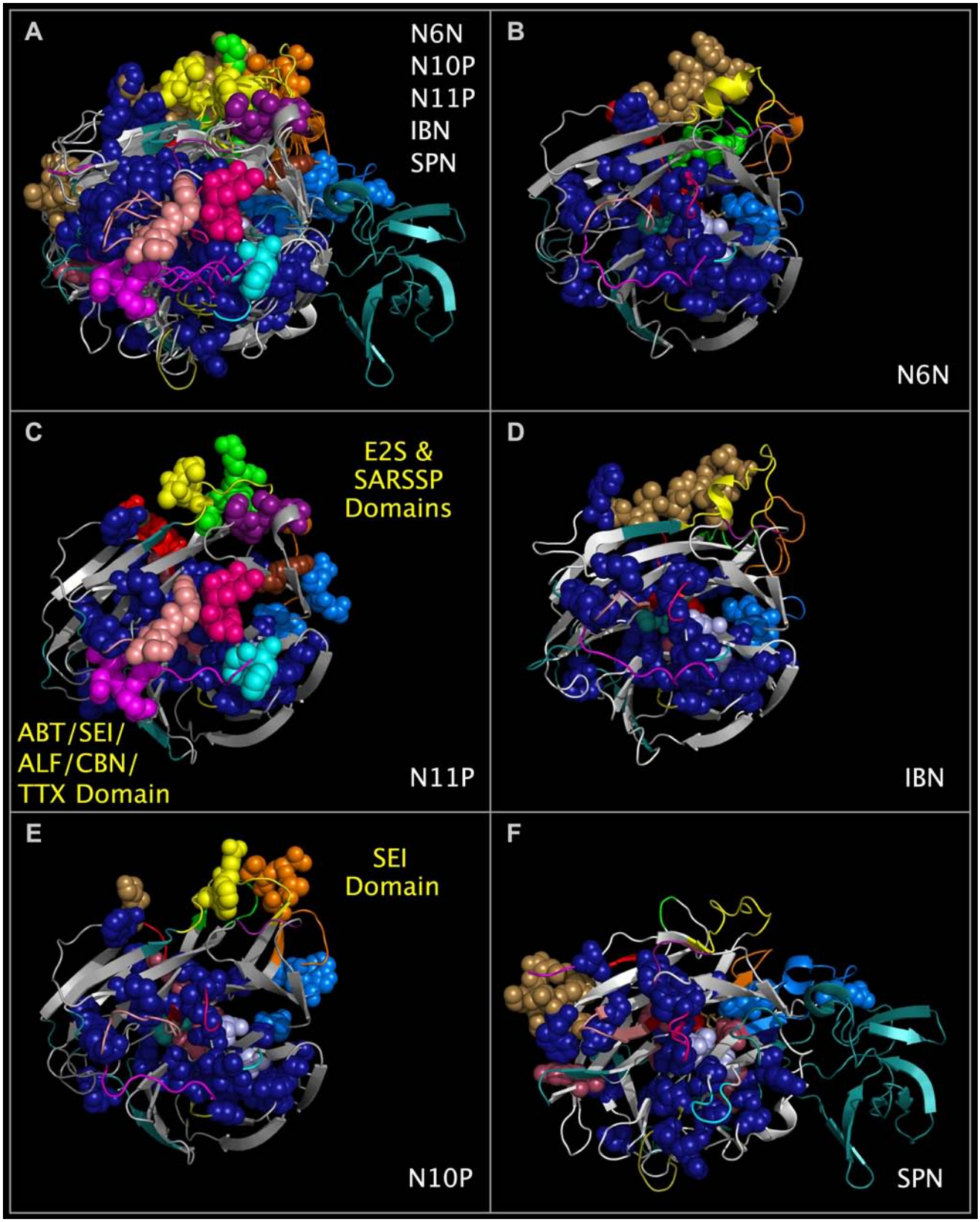
Figure 4 (previous page). Consensus active site components and Upside VLR domains of N6N, N10P, N11P, IBN, and SPN.

Figure 4. Panel 4A shows structure ribbons representing influenza A N10P [5], N11P [7], N6N [8], IBN [9]), and SPN [20] structures superposed using CNSA118:O, CNSA224:O, and CNSA276:O atoms. Atom spheres shown for each structure are identified by lowercase letters in Fig. 1. Structure ribbons and residue spheres of N10P, N11P, N6N, and IBN are color-coded as in the 'SPATIAL' row in Fig. 1. SPN residues are color coded as in the 'SPNSEQ' and 'SPN<>' rows in Fig. 1. Panel 4B shows substance P-like domains (green and red spheres) in the Upside VLR of N6N. Panel 4C shows E2S-like (green, brown, and orange spheres) and SARSSP-like (purple, red, and yellow spheres) in the Upside VLR of N11P. Panel 4D shows a C-terminal domain common to influenza B viruses (light brown colored spheres) in the Upside VLR of IBN. Panel 4E shows SEI-like domains (orange and yellow spheres) in the Upside VLR of N10P. Panel 4F shows a C-terminal domain common to bacterial viruses (light brown colored spheres) in the Upside VLR of SPN. Medium blue spheres adjacent to the consensus active site region in Panels 4A-4F are CNSR151, CNSR152, and CNSR 178 residues. Sialic acid sticks (colored medium brown) are shown in the consensus active site region of N6N, N10P, N11P, IBN, and SPN (Panel 4A), N6N (Panel 4B), and SPN (Panel 4F).

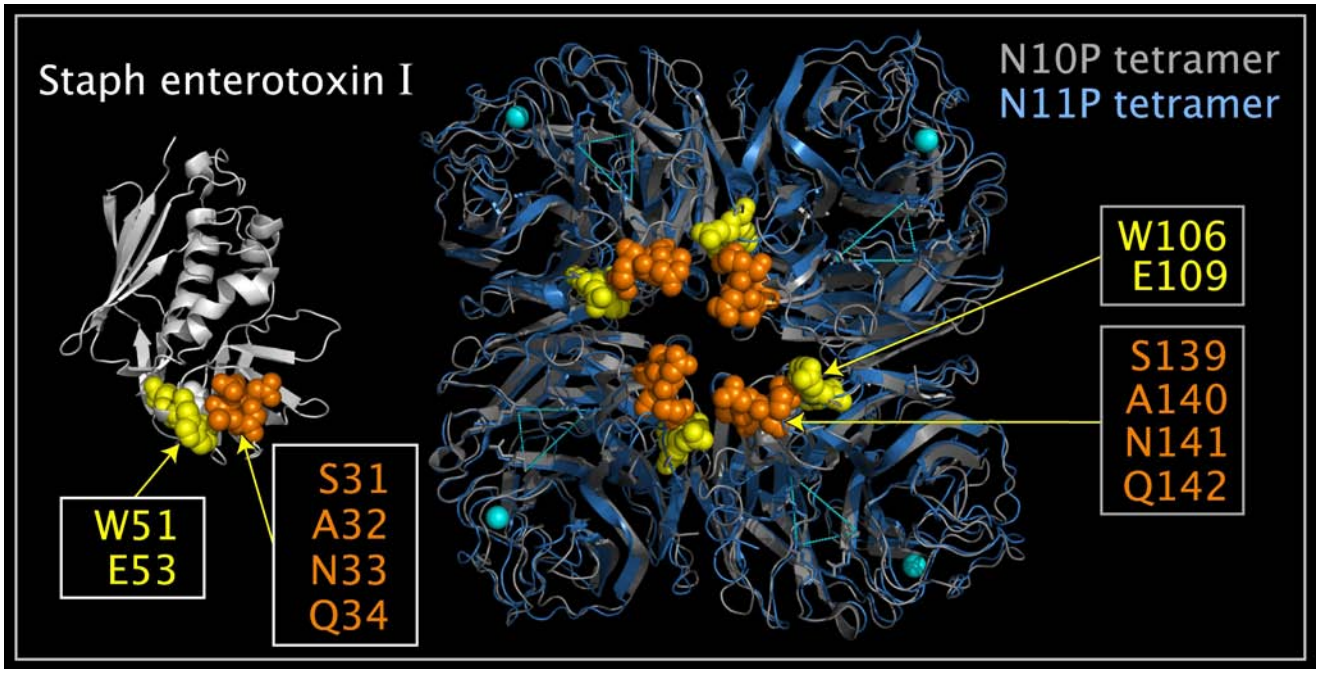
Figure 5 (next page). Downside VLR domains of N6N, N10P, N11P, IBN, and SPN.

Figure 5. Panel 5A shows structure ribbons representing influenza A N10P [5], N11P [7], N6N [8], IBN [9]), and SPN [20] structures superposed using CNSA118:O, CNSA224:O, and CNSA276:O atoms. Atom spheres shown for each structure are identified by lowercase letters in Fig. 1. Structure ribbons and residue spheres of N10P, N11P, N6N, and IBN are color-coded as in the 'SPATIAL' row in Fig. 1. SPN residues are color coded as in the 'SPNSEQ' and 'SPN<>' rows in Fig. 1. Panel 5B shows the Downside VLR loops of N6N (colored magenta, hot pink, cyan, and salmon) in the foreground and the Upside VLR C-terminal light brown spheres also shown in Panel 4B in the background. Panel 5C shows residues found in ABT and SEI (colored magenta, hot pink, cyan, and salmon spheres) on the Downside VLR of N11P in the foreground and the Upside VLR E2S-like and SARSSP-like residues residues also shown in Panel 4C in the background. Panel 5D shows the Downside VLR loops of IBN (colored magenta, hot pink, cyan, and salmon) in the foreground and the Upside VLR C-terminal light brown spheres also shown in Panel 4D in the background. Panel 5E shows the Downside VLR loops of N10P (colored magenta, hot pink, cyan, and salmon) in the foreground and the Upside VLR SEI-like domain orange and yellow spheres also shown in Panel 4E at the top. Panel 5F shows the Downside VLR loops of SPN (colored magenta, hot pink, cyan, and salmon) in the foreground, an extra domain (colored teal) on the right hand side, and the Upside VLR C-terminal light brown spheres also shown in Panel 4F in the background.

555 **Figure 5.** Downside VLR domains of N6N, N10P, N11P, IBN, and SPN.

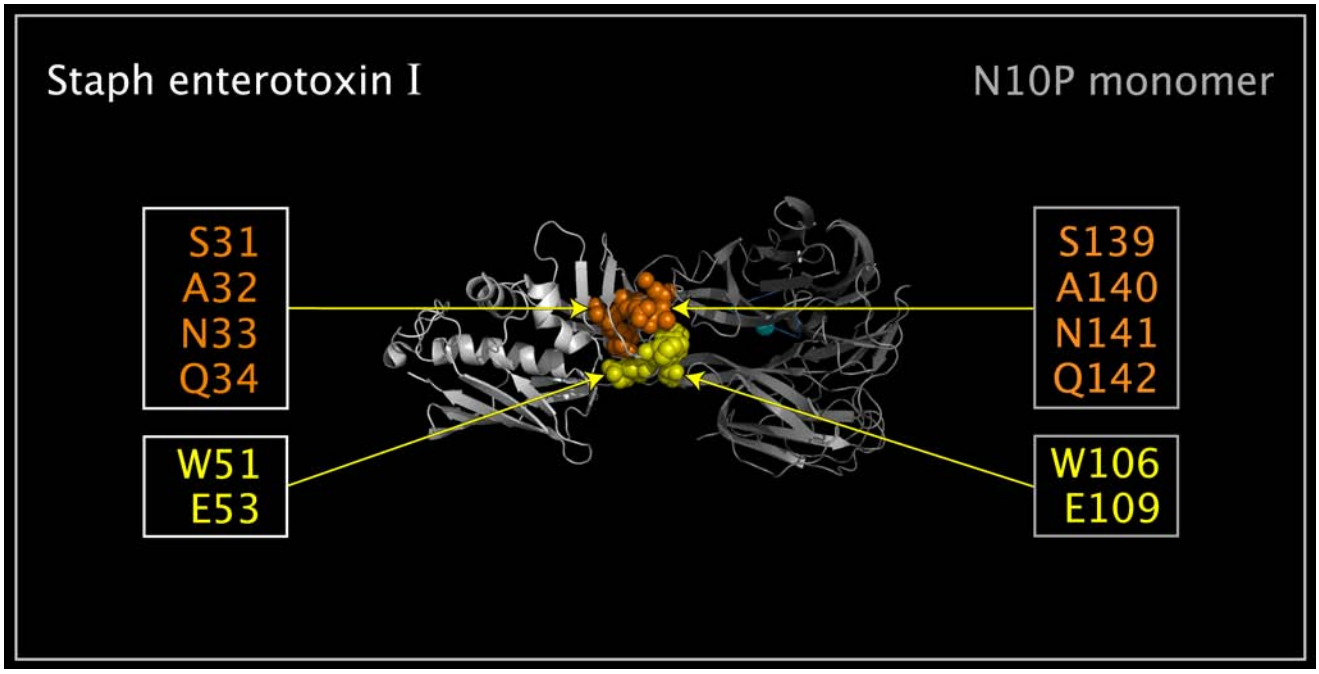


557 **Figure 6.** SEI domain and corresponding N10P tetramer Upside VLR residues with N11P tetramer reference.



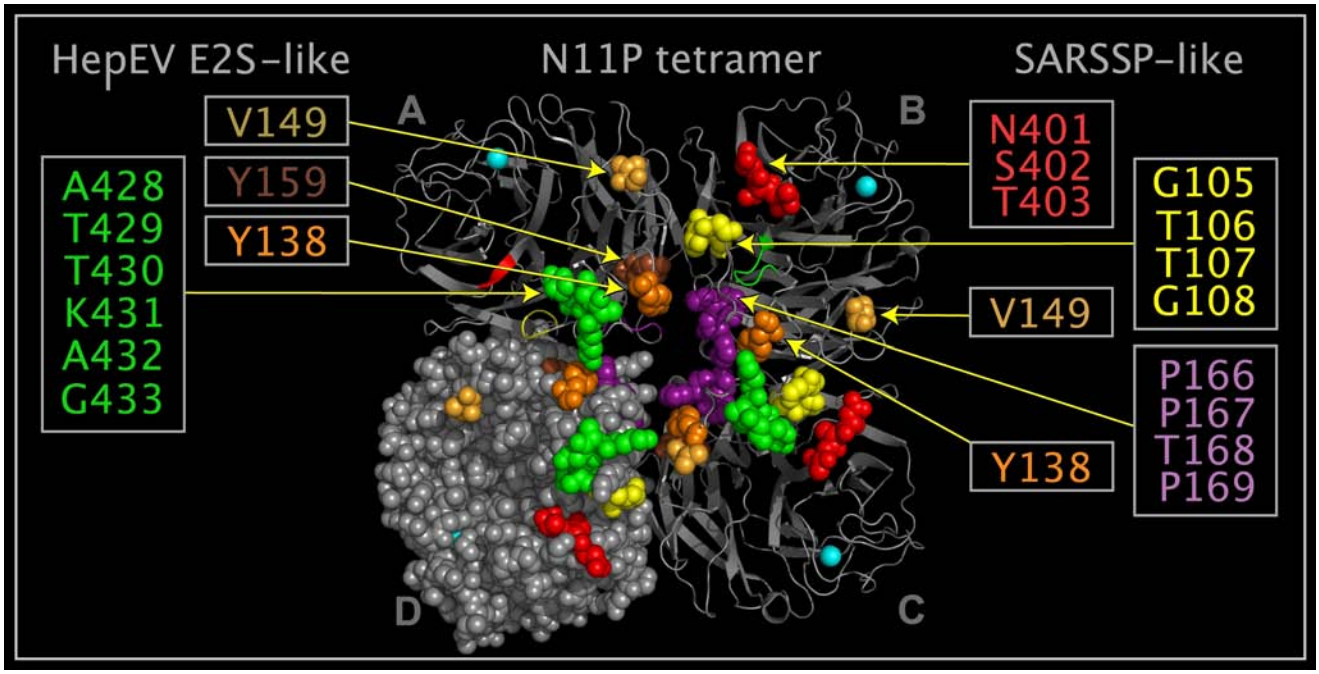
559 Figure 6. Shown are structure ribbons for SEI monomer (colored white), N10P tetramer (colored grey), and N11P
560 tetramer (colored blue). The SEI monomer and N11P tetramer are unaltered crystal structures. The N10P
561 tetramer was formed by translocating N10P monomers onto N11P monomers in the N11P tetramer. The yellow
562 triangles on the N10P monomers are lines between CNSA118:O, CNSA224:O, and CNSA276:O whose
563 superposition was used to orient N10P monomers into N11P tetramer positions. Residue spheres colored orange
564 represent: SEI residues S31, A32, N33, and Q34; and corresponding N10P Upside VLR residues S139, A140,
565 N141, and Q142. Residue spheres colored yellow represent: SEI residues W51and E53; and corresponding
566 N10P Upside VLR residues W106 and E109.
567

568 **Figure 7.** SEI and N10P Upside VLR residues in common spatial reference orientation.



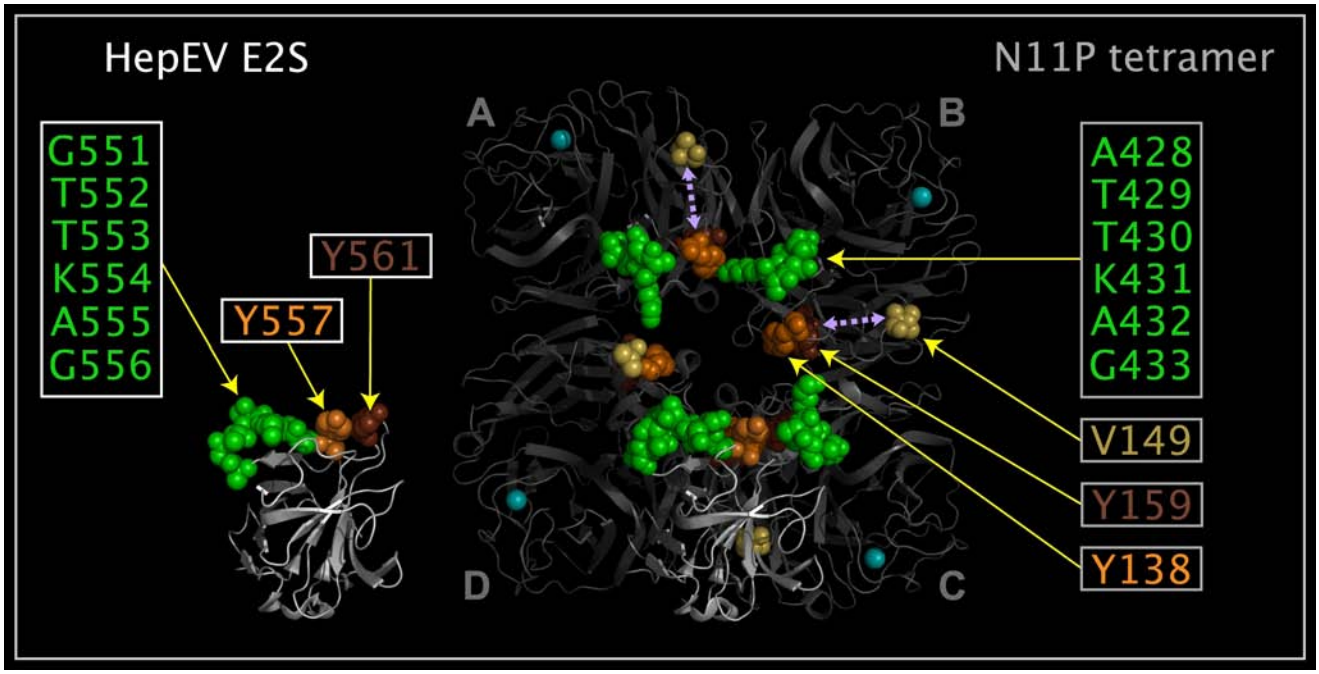
570 Figure 7. Shown are structure ribbons for superposed crystal structure SEI monomer (colored white) and N10P
571 monomer (colored grey). Residue spheres colored orange represent: SEI residues S31, A32, N33, and Q34; and
572 corresponding N10P Upside VLR residues S139, A140, N141, and Q142. Residue spheres colored yellow
573 represent: SEI residues W51and E53; and corresponding N10P Upside VLR residues W106 and E109. The SEI
574 monomer shown was superposed onto N11P tetramer Upside VLR residues using the atoms listed in Table 5.

575 **Figure 8.** E2S domains, SARSSP domains, and corresponding N11P Upside VLR residues.



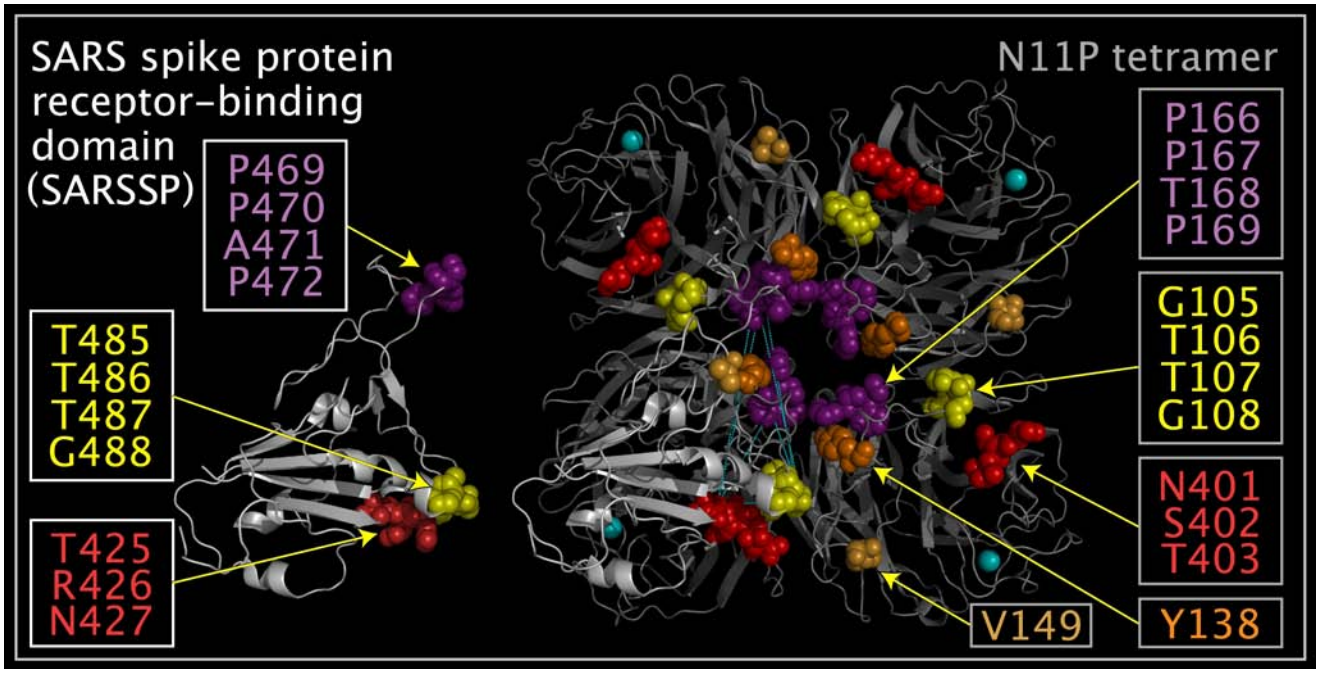
577 Figure 8. Shown are N11P tetramer crystal structure ribbons colored grey, associated calcium atoms are colored
578 cyan, and spheres depicting selected N11P Upside VLR residues colored grey with the following exceptions: in
579 monomers in positions “A”, ”B”, ”C”, and ”D”, Y138 is colored orange, V149 is colored light tangerine and calcium
580 atoms are colored cyan; in monomers in positions “A”, ”C”, and ”D”, ALA428-G433 spheres are colored green,
581 and Y159 is colored brown; and in monomers in positions ”B”, ”C”, and ”D”, G105-G108 are colored yellow, P166-
582 P169 are colored purple, and N401-T403 are colored red. Residues between Y138 and V149 are missing in the
583 crystal structure in monomers in positions “A”, ”B”, and ”D”, and the structure is disjoint. In monomer in position
584 “C”, residue G139 is between residues Y138 and V149 and the crystal structure of this monomer chain is
585 presented as contiguous. Green, brown, and orange spheres correspond to an E2S-like domain. Yellow, purple,
586 and red spheres correspond to a SARS spike protein-like domain.

587 **Figure 9.** E2S and N11P Upside VLR residues in common spatial reference orientation.



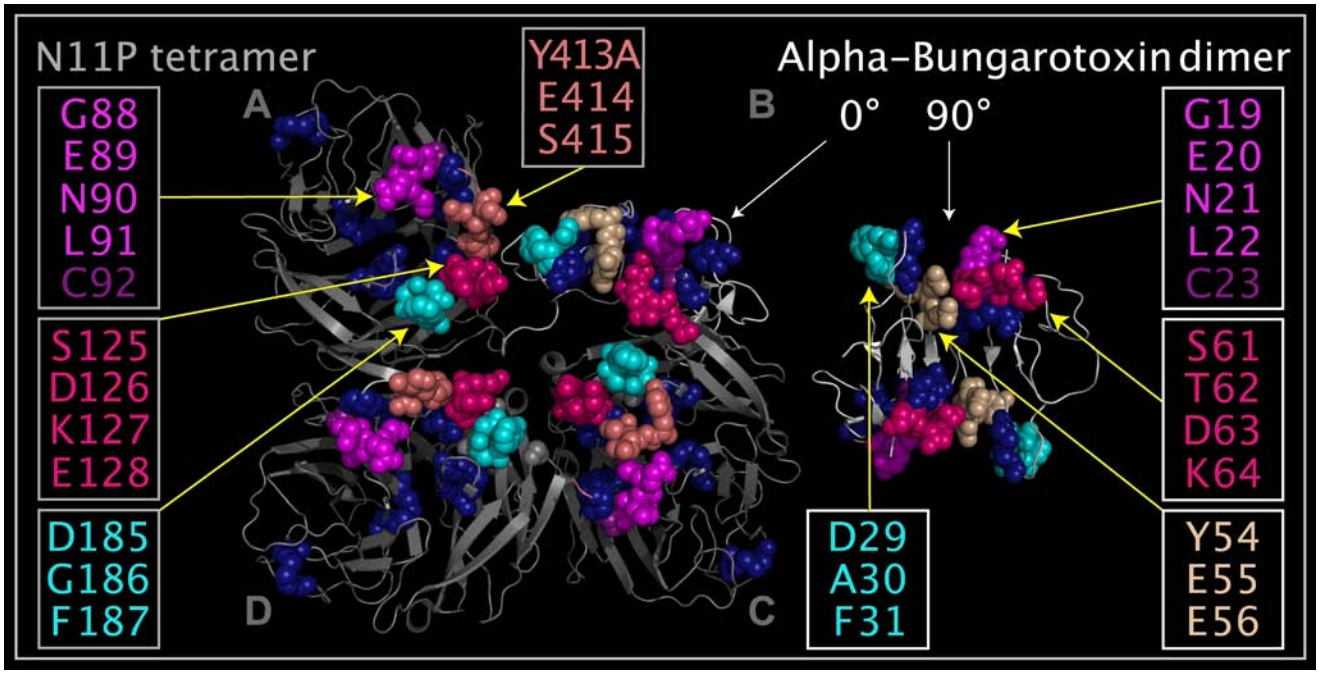
589 Figure 9. Shown are two E2S monomers (structure ribbons colored white) and a N11P tetramer (structure ribbons
590 colored grey and associated calcium atoms colored cyan) from crystal structures. One E2S monomer is shown
591 apart from the N11P tetramer and the other E2S monomer is shown with E2S residues superposed onto N11P
592 tetramer Upside VLR residues using the atoms listed in Table 6. N11P Upside VLR residue spheres depict: Y138
593 colored orange, V149 colored light tangerine, Y158 colored brown, and ALA428-G433 spheres colored green.
594 E2S residue spheres in the stand-alone and superposed monomers depict: Y557 colored orange, Y561 colored
595 brown, and G551-G556 colored green. V149 light tangerine residue spheres in N11P monomers in positions “A”,
596 ”B”, ”C”, and ”D” and arrows in monomers in positions “A” and ”B” are shown as a reference to residues missing
597 between Y138 and V149 in the N11P monomers and have no structural correspondence in E2S. The E2S
598 monomer residues spheres shown are superposed onto N11P residue spheres from two N11P monomers.

599 **Figure 10.** SARSSP and N11P Upside VLR residues in common spatial reference orientation.



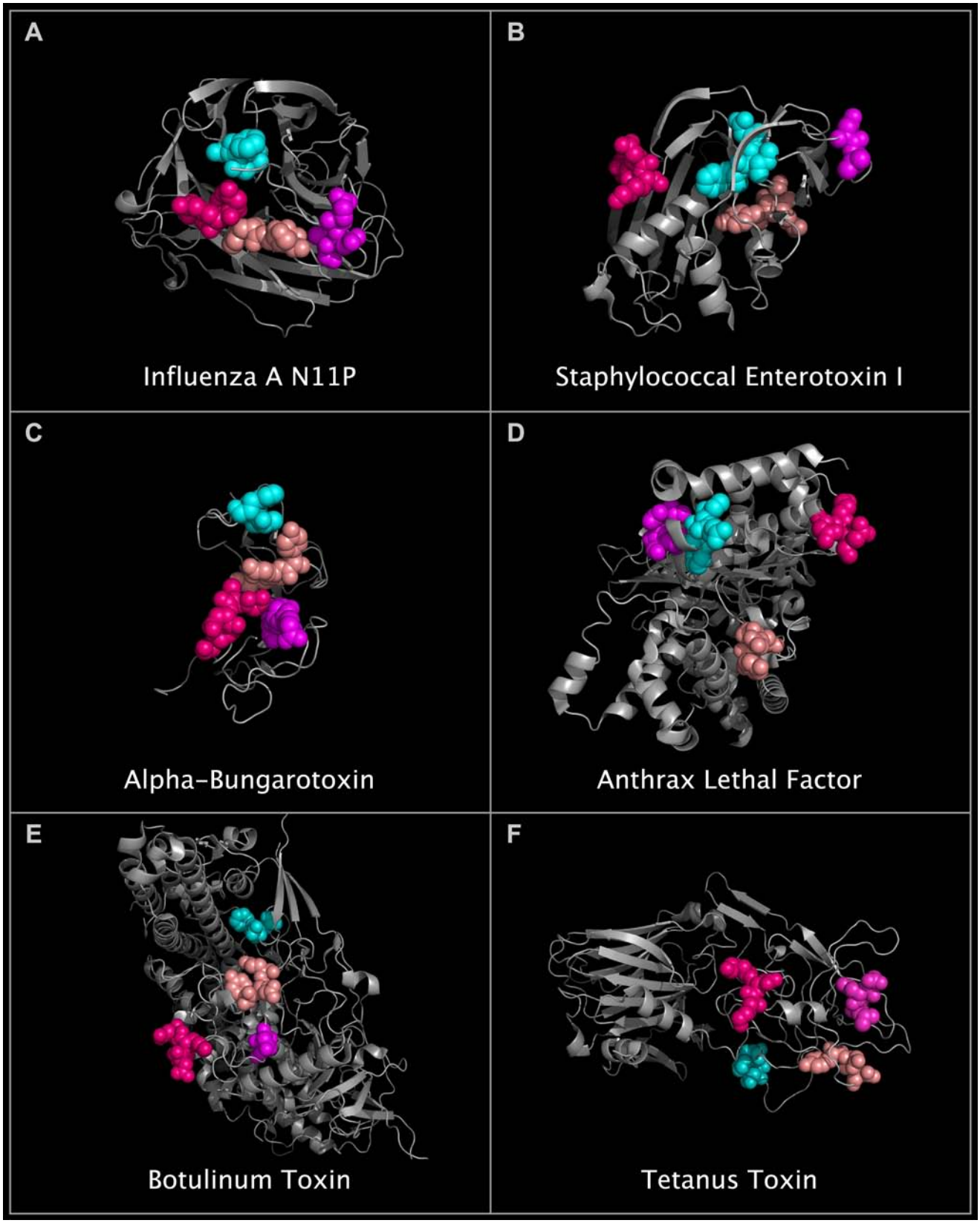
601 Figure 10. Shown are two SARSSP monomers (structure ribbons colored white) and a N11P tetramer (structure
602 ribbons colored grey and associated calcium atoms colored cyan) from crystal structures. One SARSSP
603 monomer is shown apart from the N11P tetramer and the other SARSSP monomer is shown with SARSSP
604 residues superposed onto N11P tetramer Upside VLR residues using the atoms listed in Table 7. N11P Upside
605 VLR residue spheres depict: Y138 colored orange, V149 colored light tangerine, G105-G108 colored yellow,
606 P166-P169 colored purple, and N401-T403 colored red. SARSSP residue spheres in the stand-alone and
607 superposed monomers depict: T485-G488 colored yellow, P469-P472 colored purple, and T425-N427 colored
608 red. Y138 orange and V149 light tangerine residue spheres in N11P are shown as a reference to residues
609 missing between Y138 and V149 in the N11P monomers and have no correspondence in the SARSSP monomer.

610 **Figure 11.** Corresponding residues in ABT dimer and N11P Downside VLR.



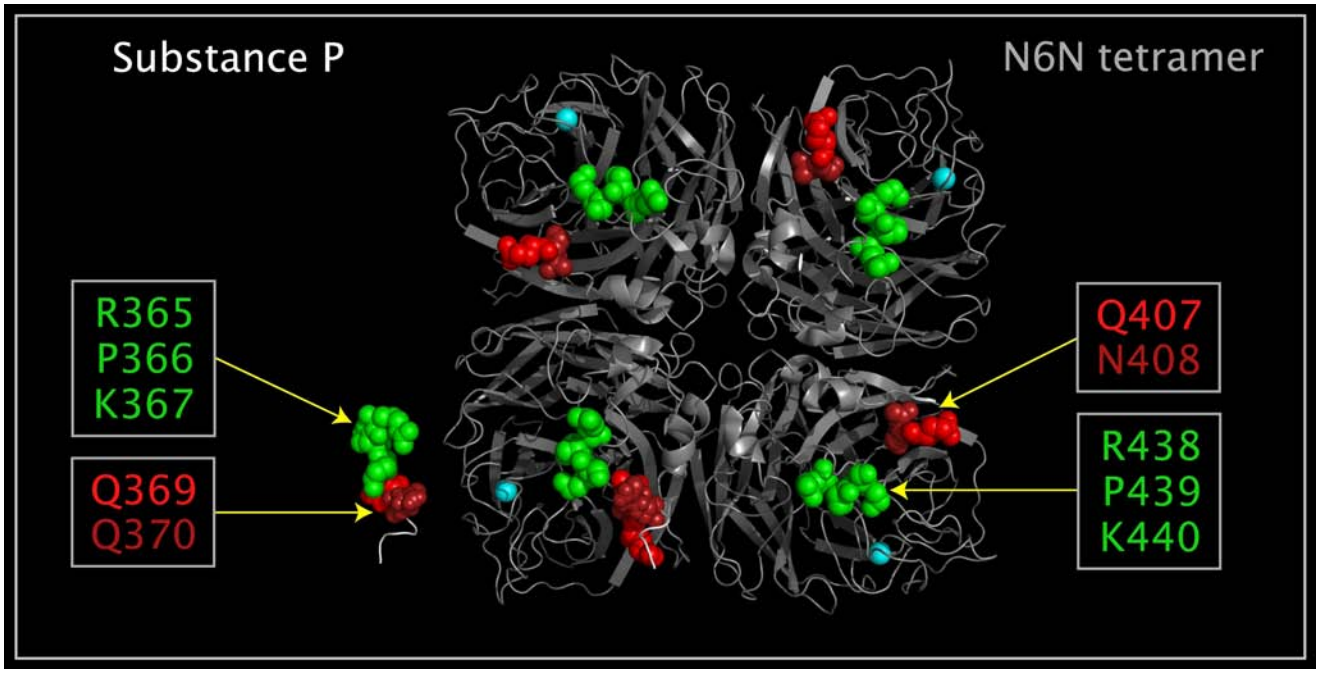
612 Figure 11. Shown are two ABT dimers (structure ribbons colored white with dark blue spheres representing
613 disulfide bridges) and a N11P tetramer (structure ribbons colored grey and with dark blue spheres representing
614 disulfide bridges) from the crystal structures. One ABT dimer is shown apart from the N11P tetramer and the
615 other ABT dimer is shown at an angle that is 90 degrees rotated from the first dimer in position “B”, i.e. positioned
616 with the ABT dimer superposed onto and substituting for the N11P monomer in position “B” in the tetramer.
617 Atoms used to superpose the ABT dimer into the position that would be occupied by the N11P “B” monomer are
618 listed in Table 8. N11P Downside VLR residue spheres in monomer positions “A”, “C”, and “D” depicting: G88-
619 L91 colored magenta, C92 colored plum, S125-E128 colored hot pink, D185-F187 colored cyan, and Y413A-S415
620 colored salmon. Corresponding ABT residue spheres in the two dimers depicting: G19-L22 colored magenta, C23
621 colored plum, S61-K64 colored hot pink, D29-F31 colored cyan, and Y54-E56 colored tan. N11P monomers in the
622 “A” and “D” positions are shown in the crystal structure positions. In the N11P monomer in the “C” position, the
623 Y413A-S415 residues (colored salmon) have been rotated into the same position relative to G19-L22 (colored
624 magenta) and D29-F31 (colored cyan) as in corresponding residues of ABT crystal structure dimer (shown in the
625 “B” position) in order to illustrate that small movements of the mobile N11P Downside VLR residues can produce
626 nearly identical relative residue presentation to ABT.

627 **Figure 12.** N11P Downside VLR residues and corresponding SEI, ABT, ALF, CBN, and TTX residues.



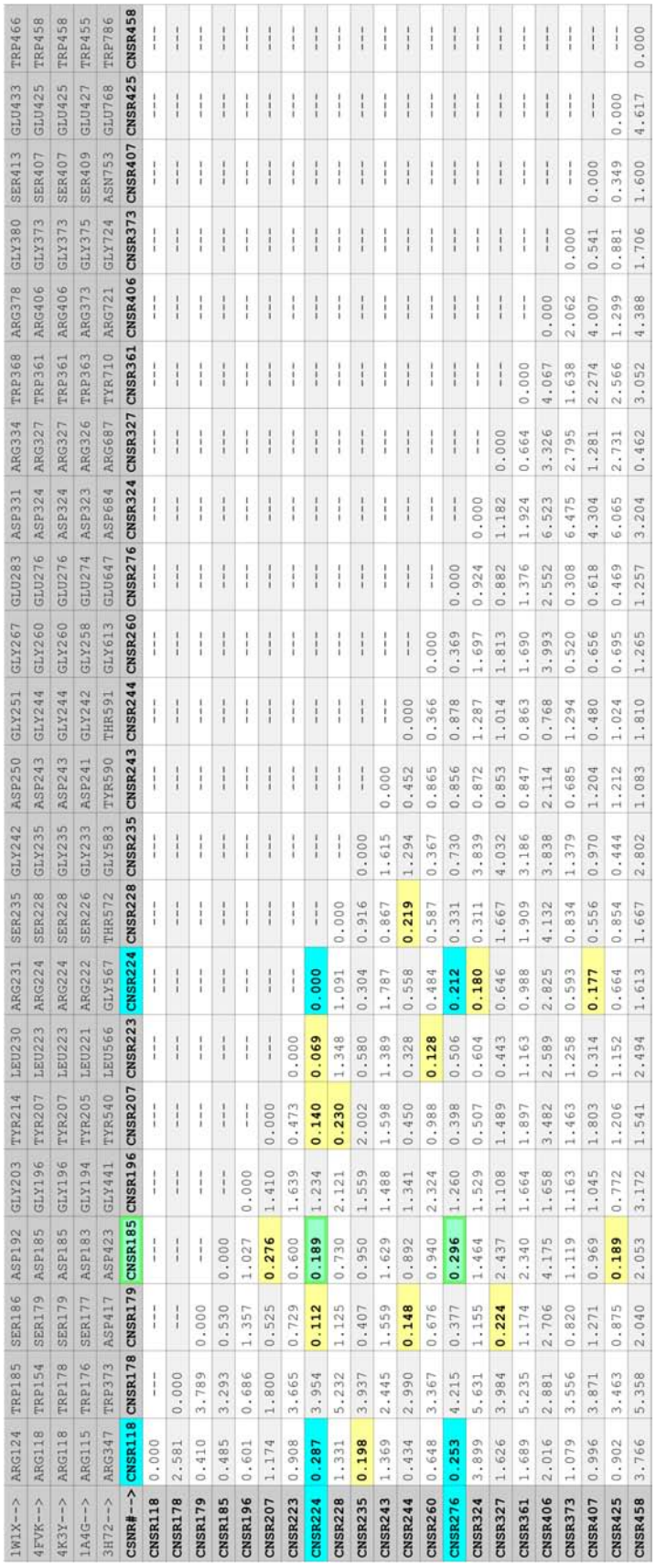
629 **Figure 12 (previous page).** N11P Downside VLR residues and corresponding SEI, ABT, ALF, CBN, and TTX residues.
630 Shown are grey structure ribbons depicting: N11P monomer (panel A), SEI monomer (panel B), ABT dimer (panel
631 C), ALF monomer (panel D), CBN monomer (panel E), and TTX monomer (panel F). Corresponding residue
632 spheres in each panel are identified and colored according to Table 9.
633

634 **Figure 13.** Reoriented Substance P residues and corresponding N6N Upside VLR residues.



636 Figure 13. Shown are a crystal structure N6N tetramer (structure ribbons colored grey), a crystal structure
637 substance P apart from the N6N tetramer, and a model-built substance P monomer superposed on a monomer in
638 the N6N tetramer. The model built structure coordinates are given as a PDB file, "WSUBP.pdb", in supporting
639 information (**File S2**). The atoms used to map WSUBP.pdb onto the N6N monomer are given in Table 10. The
640 model building consisted of reorienting the first 3 residues of substance P, R365-K367 (colored green), to exactly
641 match the corresponding N6N residues, R438-K440 (colored green), and leaving the rest of the structure as in the
642 crystal structure. N11P residues, Q 407 and N408 (colored red), correspond to substance P residues, Q369 and
643 Q370 (colored red).

644 **Figure 14.** Interatomic distance population standard deviations for selected N6N, N10P, N11P, IBN, and SPN
645 atoms.
646

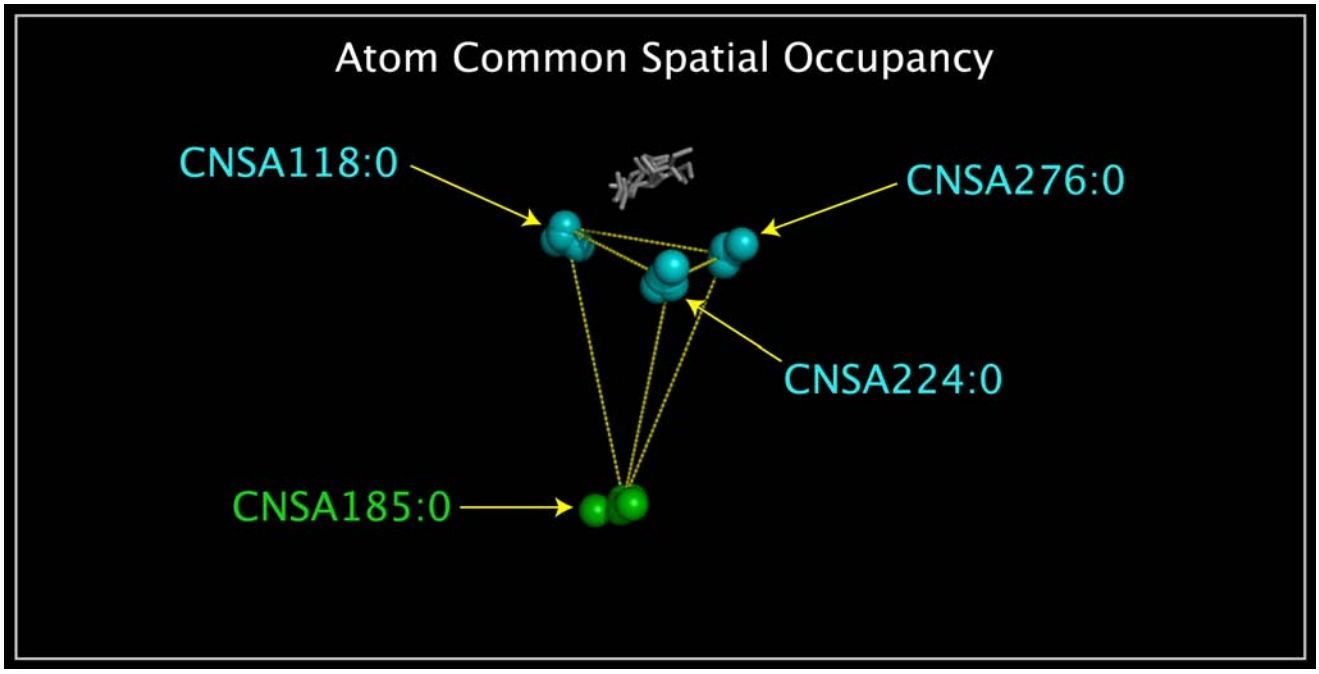


1W1X-->	ARG124	TRP185	SER186	ASP192	GLY203	TYR214	LEU230	ARG231	SER235	GLY242	ASP250	GLY251	GLY267	GLU283	ASP331	ARG334	TRP368	ARG378	GLY380	SER413	GLU433	TRP466
4FVK-->	ARG118	TRP154	SER179	ASP185	GLY196	TYR207	LEU223	ARG224	SER228	GLY235	ASP243	GLY244	GLY260	GLU276	ASP324	ARG327	TRP361	ARG406	GLY373	SER407	GLU425	TRP458
4R3Y-->	ARG118	TRP178	SER179	ASP185	GLY196	TYR207	LEU223	ARG224	SER228	GLY235	ASP243	GLY244	GLY260	GLU276	ASP324	ARG327	TRP361	ARG406	GLY373	SER407	GLU425	TRP458
1A4G-->	ARG115	TRP176	SER177	ASP183	GLY194	TYR205	LEU221	ARG222	SER226	GLY233	ASP241	GLY242	GLY258	GLU274	ASP323	ARG326	TRP363	ARG373	GLY375	SER409	GLU427	TRP455
3H72-->	ARG347	TRP373	ASP417	ASP423	GLY441	TYR540	LEU566	GLY567	THR572	GLY583	TYR590	THR591	GLY613	GLU647	ASP684	ARG687	TYR710	ARG721	GLY724	ASN753	GLU768	TRP786
CSNR#-->	**CNSR118**	**CNSR178**	**CNSR179**	**CNSR185**	**CNSR196**	**CNSR207**	**CNSR223**	**CNSR224**	**CNSR228**	**CNSR235**	**CNSR243**	**CNSR244**	**CNSR260**	**CNSR276**	**CNSR324**	**CNSR327**	**CNSR361**	**CNSR406**	**CNSR373**	**CNSR407**	**CNSR425**	**CNSR458**
CNSR118	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR178	2.581	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR179	0.410	3.789	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR185	0.485	3.293	0.530	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR196	0.601	0.686	1.357	1.027	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR207	1.174	1.800	0.525	**0.276**	1.410	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR223	0.908	3.665	0.729	0.600	1.639	0.473	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR224	**0.287**	3.954	**0.112**	**0.189**	1.234	**0.140**	**0.069**	**0.000**	---	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR228	1.331	5.232	1.125	0.730	2.121	**0.230**	1.348	1.091	0.000	---	---	---	---	---	---	---	---	---	---	---	---	---
CNSR235	**0.198**	3.937	0.407	0.950	1.559	2.002	0.580	0.304	0.916	0.000	---	---	---	---	---	---	---	---	---	---	---	---
CNSR243	1.369	2.445	1.559	1.629	1.488	1.598	1.389	1.787	0.867	1.615	0.000	---	---	---	---	---	---	---	---	---	---	---
CNSR244	0.434	2.990	**0.148**	0.892	1.341	0.450	0.328	0.558	**0.219**	1.294	0.452	0.000	---	---	---	---	---	---	---	---	---	---
CNSR260	0.648	3.367	0.676	0.940	2.324	0.988	**0.128**	0.484	0.587	0.367	0.865	0.366	0.000	---	---	---	---	---	---	---	---	---
CNSR276	**0.253**	4.215	0.377	**0.296**	1.260	0.398	0.506	**0.212**	0.331	0.730	0.856	0.878	0.369	0.000	---	---	---	---	---	---	---	---
CNSR324	3.899	5.631	1.155	1.464	1.529	0.507	0.604	**0.180**	0.311	3.839	0.872	1.287	1.697	0.924	0.000	---	---	---	---	---	---	---
CNSR327	1.626	3.984	**0.224**	2.437	1.108	1.489	0.443	0.646	1.667	4.032	0.853	1.014	1.813	0.882	1.182	0.000	---	---	---	---	---	---
CNSR361	1.689	5.235	1.174	2.340	1.664	1.897	1.163	0.988	1.909	3.186	0.847	0.863	1.690	1.376	1.924	0.664	0.000	---	---	---	---	---
CNSR406	2.016	2.881	2.706	4.175	1.658	3.482	2.589	2.825	4.132	3.838	2.114	0.768	3.993	2.552	6.523	3.326	4.067	0.000	---	---	---	---
CNSR373	1.079	3.556	0.820	1.119	1.163	1.463	1.258	0.593	0.834	1.379	0.685	1.294	0.520	0.308	6.475	2.795	1.638	2.062	0.000	---	---	---
CNSR407	0.996	3.871	1.271	0.969	1.045	1.803	0.314	**0.177**	0.556	0.970	1.204	0.480	0.656	0.618	4.304	1.281	2.274	4.007	0.541	0.000	---	---
CNSR425	0.902	3.463	0.875	**0.189**	0.772	1.206	1.152	0.664	0.854	0.444	1.212	1.024	0.695	0.469	6.065	2.731	2.566	1.299	0.881	0.349	0.000	---
CNSR458	3.766	5.358	2.040	2.053	3.172	1.541	2.494	1.613	1.667	2.802	1.083	1.810	1.265	1.257	3.204	0.462	3.052	4.388	1.706	1.600	4.617	0.000

Figure 14 (previous page). Interatomic distance population standard deviations for selected N6N, N10P, N11P, IBN, and SPN atoms.

Figure 14. Shown are the standard deviations of the interatomic distances for specific main chain oxygens of N6N, N10P, N11P, IBN, and SPN residues; these oxygens were selected if the Table 2 row for their associated residue does not contain cysteines, prolines, or missing residues. The standard deviations colored in cyan correspond to the minimal standard deviation of the distances between the CNSA118:O, CNSA224:O, and CNSA276:O atoms. The standard deviations colored in green correspond to the minimal standard deviation of the distances between the CNSA185:O atom and each of the CNSA224:O and CNSA276:O atoms. The CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms form a tetrahedron. Values under 3.00 are colored and, unless colored cyan or green as above, default to yellow.

657 **Figure 15.** Overlapping CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atom tetrahedrons.



659 Figure 15. Shown are the superposed CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms from
660 N6N, N10P, N11P, IBN, and SPN. The CNSA118:O, CNSA224:O, and CNSA276:O atoms are colored cyan. The
661 CNSA185:O atoms are colored green. The CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms form
662 a tetrahedron. Also shown are yellow lines drawn between the N6N CNSA118:O, CNSA224:O, CNSA276:O, and
663 CNSA185:O atoms. The combined superposed substrates and inhibitors from the superposed structures are
664 shown as grey sticks.

665 **Figure 16.** CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atom tetrahedrons in N6N, N10P, N11P,
666 IBN, and SPN.

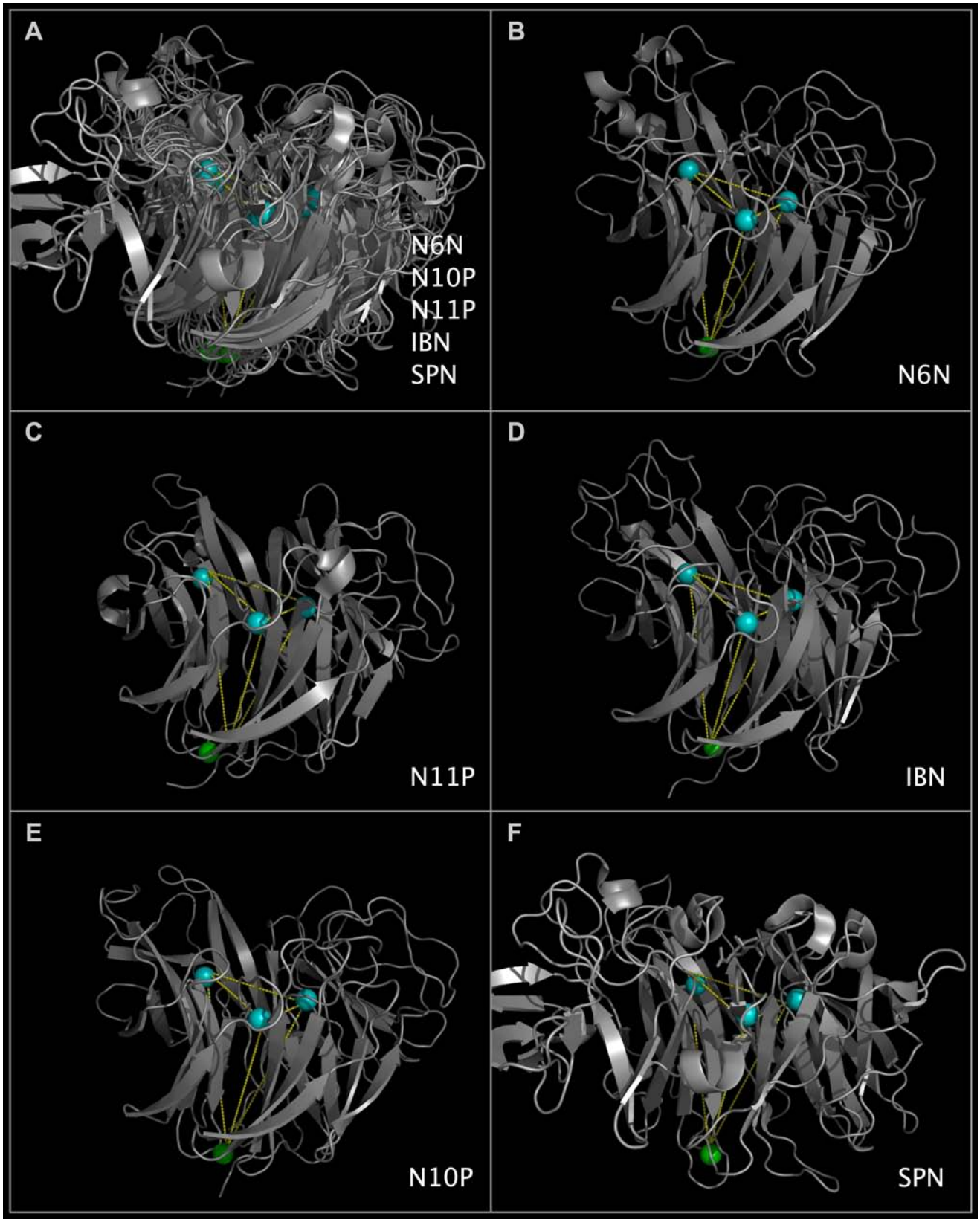


Figure 16 (previous page). CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atom tetrahedrons in N6N, N10P, N11P, IBN, and SPN.

Figure 16. Panel A shows the superposed N6N, N10P, N11P, IBN, and SPN structure ribbons colored grey, the CNSA118:O, CNSA224:O, and CNSA276:O atoms colored cyan, the CNSA185:O atoms colored green, and the lines between N6N CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms colored yellow. Panels B-E show individual N6N, N10P, N11P, IBN, and SPN structures. In Panels B-E, the structure ribbons are colored grey, the corresponding CNSA118:O, CNSA224:O, and CNSA276:O atoms are colored cyan, the corresponding CNSA185:O atoms are colored green, and actual or superposed lines between N6N CNSA118:O, CNSA224:O, CNSA276:O, and CNSA185:O atoms are colored yellow (forming a tetrahedron).

677 **Figure 17.** Interatomic distance population standard deviations for selected N10P and SEI atoms.

		N10P W106 / SEI W51													
		N	CA	C	O	CB	CG	CD1	CD2	NE1	CE2	CE3	CZ2	CZ3	CH2
N10P E109 / SEI E53	N	1.228	0.612	0.813	1.193	0.116	0.424	0.625	0.453	0.738	0.655	0.278	0.698	0.355	0.565
	CA	0.910	0.251	0.554	1.143	0.424	0.193	0.059	0.121	0.053	0.032	0.187	0.113	0.078	0.067
	C	0.284	0.240	0.187	1.017	0.978	0.722	0.621	0.542	0.452	0.385	0.485	0.212	0.266	0.152
	O	0.843	1.271	0.761	0.199	1.849	1.451	1.251	1.194	0.980	0.905	1.140	0.621	0.761	0.544
	CB	1.350	0.546	0.743	1.220	0.110	0.063	0.035	0.138	0.020	0.071	0.264	0.103	0.293	0.201
	C	1.348	0.542	0.623	1.125	0.122	0.006	0.056	0.033	0.111	0.106	0.004	0.152	0.054	0.130
	CD	3.705	0.511	0.724	1.242	0.122	0.055	0.060	0.127	0.059	0.042	0.264	0.069	0.294	0.185
	OE1	3.563	0.719	0.998	1.600	0.045	0.175	0.293	0.164	0.344	0.277	0.056	0.298	0.078	0.205
	OE2	3.913	0.262	0.488	0.931	0.295	0.325	0.174	0.531	0.268	0.468	0.773	0.594	0.911	0.799

679 Figure 17. Shown are the standard deviations of the interatomic distances for N10P residue atoms in W106 and
680 E109 and the interatomic distances SEI residue atoms in W51 and E53. The standard deviations shown in
681 magenta correspond to the minimal standard deviations less than 0.1. The standard deviations shown in yellow
682 correspond to the minimal standard deviations less than 0.2. The TRP and GLU residues in the N10P and SEI
683 proteins are in the same relative spatial orientation indicated by a large cluster of low interatomic distance
684 standard deviations despite the fact that these residues are separated by two amino acids in N10P and one
685 amino acid in SEI.

Figure Abbreviation, Reference Number, Sequence Identifier, And Sequence Descriptions
N = not included in Figure 1; I, II, III, and IV are Figure 1 sequence groupings
Numbers before sequence are: beginning sequence number - ending sequence number (as reported in PDB file) and number of amino acids in the row.

```
1A4G-IB  [9]       1A4G.pdb         neuraminidase [Influenza B virus B/Beijing/1/87] complexed with zanamivir
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
I       76-177 102 EPEWTYPRLSCQGSTFQKALLISPHRFGEARGNSAPLIIREPFIACGPKECKHFALTHYAAQPGGYYNGTREDRNKLRHLISVKLGKIPTVENSIFHMAAWS
II     178-278 101 GSACHDGREWTYIGVDGPDSNALIKIKYGEAYTDTYHSYANNILRTQESACNCIGGDCYLMITDGSASGISKCRFLKIREGRIIKEIFPTGRVEHTEECTC
III    279-373  95 GFASNKTIECACRDNSYTAKRPFVKLNVETDTAEIRLMCTETYLDTPRPDDGSITGPCESNGDKGRGGIKGGFVHQRMASKIGRWYSRTMSKTER
IV     374-465  92 MGMELYVRYDGDPWTDSDALAHSGVMVSMKEPGWYSFGFEIKDKKCDVPCIGIEMVHDGGKKTWHSAATAIYCLMGSGQLLWDTVTGVDMAL

3209-01 [11]       ADR32096.1       neuraminidase [Influenza A virus (A/Lyon/1364/2007(H1N1))]
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         1-60  60 MLQIGNIISIWASHSIQTGSQNNTGICNQRIITYENSTWVNHTYVNINNTNVVAGEDKTS
I       61-158  98 VTLAGNSSLCSISGWAIYTKDNSIRIGSKGDVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVKDRSPYRALMSCPLGEAPSPYNSKFESVAWS
II     159-259 101 ASACHDGMGWLTIGISGPDNGAVAVLKYNGIITGTIKSWKKQILRTQESECVCMNGSCFTIMTDGPSNKAASYKIFKIEKGKVTKSIELNAPNFHYEECSC
III    260-350  91 YPDTGIVMCVCRDNWHGSNRPWVSFNQNLDYQIGYICSGVFGDNPRPEDGEGSCNPVTVDGANGVKGFSYKYDNGVWIGRTKSNRLRKGFE
IV     351-435  85 MIWDPNGWTNTDSDFSVKQDVVAITDWSGYSGSFVQHPELTGLDCIRPCFWVELVRGLPRENTTIWTSGSSISFCGVNSDTANWS

4790-01 [12]       ACJ47909.1       neuraminidase [Influenza A virus (A/environment/Qinghai/1/2008(H5N1))]
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         1-62  62 MNPNQKIITIGSICMVIGIVSLMLQIGNIISIWVSHSIQTGNQHQAEPISNTNFLADKAVAS
I       63-159  98 VTLAGNSSLCPISGWAVHSKDNGKRIGSKGDVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVKDRSPHRTLMSCPVGEAPSPYNSRFESVAWS
II     160-260 101 ASACHDGTSWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFKMEKGKVVKSVELDAPNYHYEECSC
III    261-350  91 YPDAGEITCVCRDNWHGSNRPWVSFNQNLEYQIGYICSGVFGDNPRPNDGTGSCGPVSPNGAYGVKGFSFKYGNGVWIGRTKSINSRSGFE
IV     351-449  97 MIWDPNGWTGTDSSFSVKQDIVAITDWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKESTIWTSGSSISFCGVNSDTVGWSWPDGAELPFTIDK

5091-01 [13]       AF509109.2       neuraminidase [Influenza A virus (A/Chicken/Hong Kong/873.3/01 (H5N1))]
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         1-82  82 MNPNQKIITIGSICMVIGIVSLMLQIGNIISIWVSHSIQTGNQHQAEPCNQSIITYENNTWVNQTYVNISNTNFLTEKAVAS
I       83-180  98 VTLAGNSSLCPISGWAVYSKDNGIRIGSKGDVFVIREPFISCSHLECRTFFLTQGALLNDKHSNGTVKDRSPYRTLMSCPVGEAPSPYNSRFESVAWS
II     181-281 101 ASACHDGTSWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFKIEKGKVVKSVELNAPNYHYEECSC
III    282-372  91 YPDAGEITCVCRDNWHGSNRPWVSFNQNLEYQIGYICSGVFGDNPRPNDGTGSCGPVSPNGAYGIKGFSFKYGNGVWIGRTKSTNSRSGFE
IV     373-469  97 MIWDPNGWTGTDSNFSVKQDIVAITDWSGYSGSFVQHPELTGVDCIRPCFWVELIRGRPKESTIWTSGSSISFCGVNSDTVGWSWPDGAELPFTIDK

5343-01 [14]       AEV53435.1       neuraminidase [Influenza A virus (A/Fukushima/09FY004/2009(H1N1))]
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         1-92  92 MNPNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNQIETCNQSVITYENNTWVNQTYVNISNTNFAAGQSVVS
I       93-190  98 VKLAGNSSLCPVSGWAIYSKDNSIRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWS
II     191-291 101 ASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGSCFTVMTDGPSDGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSC
III    292-382  91 YPDSSEITCVCRDNWHGSNRPWVSFNQNLEYQIGYICSGIFGDNPRPNDKTGSCGPVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFE
IV     383-479  97 MIWDPNGWTGTDNNFSIKQDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTIWTSGSSISFCGVNSDTVGWSWPDGAELPFTIDK

5971-02 [15]       ADG59718.1       neuraminidase [Influenza A virus (A/El Salvador/2-Q226L/1957(H2N2))]
                   123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         1-82  82 MNPNQKIITIGSVSLTIATACLLMQIAILAATVTLHFKQHECDSPASNQVMPCEPIIIERNITEIVYLNNTTIEKEICPEVV
I       83-179  97 EYRNWSKPQCQITGFAPFSKDNSIRLSAGGDIWVTREPYVSCDPGKCYQFALGQGTTLDNKHSNGTIHDRIPHRTLLMNELGVPFHLGTKQVCVAWS
II     180-280 101 SSSCHDGKAWLHVCVTGDDRNATASFIYDGRLVDSIGSWSQNILRTQESECVCINGTCTVVMTDGSASGRADTRILFIKEGKIVHISPLSGSAQHIEECSC
III    281-375  95 YPRYPDVRCICRDNWKGSNRPVIDINMEDYSIDSSYVCSGLVGDTPRNDDSSSNSNCRDPNNERGNPGVKGWAFDNGDDVWMGRTINKDSRSGYE
IV     376-469  94 TFKVIGGWSTPNSKSQVNRQVIVDNNNWSGYSGIFSVEGKSCINRCFYVELIRGRPQETRVWWTSNSIVVFCGTSGTYGTGSWPDGANINFMPI
```

Figure Abbreviation, Reference Number, Sequence Identifier, And Sequence Descriptions
N = not included in Figure 1; I, II, III, and IV are Figure 1 sequence groupings
Numbers before sequence are: beginning sequence number - ending sequence number (as reported in PDB file) and number of amino acids in the row.

8342-02 [16] **AGW83423.1 neuraminidase [Influenza A virus (A/Djibouti/N09200/2009(H3N2))]**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          1-82   82 MNPNQKIITIGSVSLTISTICFFMQTAILITTVTLHFKQCEFNSPPNNQVMLCEPTIIERNITEIVYLTNTTIEKEICPKLA
I        83-179   97 EYRNWSKPQCDITGFAPFSKDNSIRLSAGGDIWVTREPYVSCDPDKCYQFALGQGTTLNNVHSNNTVRDRTPYRTLLMNELGVPFHLGTKQVCIAWS
II     180-280  101 SSSCHDGKAWLHVCITGDDKNATASFIYNGRLVDSVVSWSKEILRTQESECVCINGTCTVVMTDGSASGKADTKILFIEEGKIVHTSTLSGSAQHVEECSC
III    281-375   95 YPRYPGVRCVCRDNWKGSNRPIVDINIKDHSIVSSYVCSGLVGDTPRKNDSSSSSHCLDPNNEEGGHGVKGWAFDDGNDVWMGRTISEKSRFGYE
IV     376-469   94 TFKVIEGWSNPKSKLQINRQVIVDRGNRSGYSGIFSVEGKSCINRCFYVELIRGRKEETEVLWTSNSIVVFCGTSGTYGTGSWPDGADINLMPI
```

6202-03 [17] **AAO62026.1 neuraminidase [Influenza A virus (A/Goose/HonGKonG/27404/78(H5N3))]**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          1-92   92 MNPNQKIITIGVVNTTLSTIALLIGVGNLIFNTVIHEKIGDHQTVVYPTITTPVVPNCSDTIITYNNTVINNITTTIITETE
I        93-180   98 RHFKPSLPLCPFRGFFPFHKDNAIRLGENKDVIVTREPYVSCDNDNCWSFALAQGALLGTKHSNGTIKDRTPYRSLIRFPIGTAPVLGNYKEICVAWS
II     181-282  102 SSSCFDGKEWMHVCMTGNDNDASAQIIYAGKMTDSIKSWRRDILRTQESECQCIDGTCIVVVTDGPAANSADHRIYWIRRGKVIKYENVPKTKIQHLEECSC
III    283-374   92 YVDTDVYCICRDNWKGSNRPWMRINNETILETGYVCSKFHSDTPRPADPSTVSCDSPSNVNGGPGVKGFGFKAGNDVWLGRTVSTSGRSGFE
IV     375-469   95 IIKVTEGWINSPNHAKSLTQTLVSNNDWSGYSGSFIIENNGCFQPCFYIELIRGRPNKNDDVSWTSNSIVTFCGLDNEPGSGNWPDGSNIGFMPK
```

1150-09 [19] **AHA11501.1 neuraminidase [Influenza A virus (A/ZhejianG/DTID-ZJU10/2013(H7N9))]**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          1-78   78 MNPNQKILCTSATAIIIGAIAVLIGIANLGLNIGLHLKPGCNCSHSQPETTNTSQTIINNYYNETNITNIQMEERTSR
I        79-176   98 NFNNLTKGLCTINSWHIYGKDNAVRIGESSDVLVTREPYVSCDPDECRFYALSQGTTIRGKHSNGTIHDRSQYRALISWPLSSPPTVYNSRVECIGWS
II     177-277  101 STSCHDGKSRMSICISGPNNNASAVVWYNRRPVAEINTWARNILRTQESECVCHNGVCPIVFTDGSATGPADTRIYYFKEGKILEWESLTGTAKHIEECSC
III    278-371   94 YGERTGITCTCRDNWQGSNRPVIQIDPVAMTHTSQYICSPVLTDTPRPNDPNIGKCNDPYPGNDNNGVKGFSYLDGANTWLGRTISTASRSGYE
IV     372-465   94 MLKVPNALTDDRSKPIQGQTIVLNADWSGYSGSFMDYWAEGDCYRACFYVELIRGRPKEDKVWWTSNSIVSMCSSTEFLGQWNWPDGARIEYFL
```

6207-06 [18] **AAO62070.1 neuraminidase [Influenza A virus (A/quail/NanchanG/4-034/2000(H4N6))]**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          1-92   92 MNPNQKIICISATGMTLSVVSLLIGIANLGLNIGLHYKMGDTPDVNIPNMNETNSTTTIINNHTQNNFTNITNIIVNKNEEG
I        93-190   98 TFLNLTKPLCEVNSWHILSKDNAIRIGEDAHILVTREPYLSCDPQGCRMFALSQGTTLRGRHANGTIHDRSPFRALISWEMGQAPSPYNVRVECIGWS
II     191-291  101 STSCHDGISRMSICMSGANNNASAVVWYGGRPVTEIPSWAGNILRTQESECVCHKGICPVVMTDGPANNRAATKIIYFKEGKIQKIEELAGNTQHIEECSC
III    292-386   95 YGAVGVIKCICRDNWKGANRPVITIDPEMMTHTSKYLCSKILTDTSRPNDPTNGNCDAPITGGSPDPGVKGFAFLDRENSWLGRTISKDSRSGYE
IV     387-480   94 MLKVPNAETDTQSGPISHQVIVNNQNWSGYSGAFIDYWANKECFNPCFYVELIRGRPKESSVLWTSNSIVALCGSKERLGSWSWHDGAEIIYFK
```

1W1X-06 [8] **1W1X.pdb neuraminidase from english duck subtype N6 complexed with 30 mm sialic acid (NANA, NEU5AC)**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N            88    1 R
I        89-186   98 TFLNLTKPLCEVNSWHILSKDNAIRIGEDAHILVTREPYLSCDPQGCRMFALSQGTTLRGRHANGTIHDRSPFRALISWEMGQAPSPYNTRVECIGWS
II     187-287  101 STSCHDGMSRMSICMSGPNNNASAVVWYGGRPITEIPSWAGNILRTQESECVCHKGVCPVVMTDGPANNRAATKIIYFKEGKIQKIEELAGNAQHIEECSC
III    288-382   95 YGAGGVIKCICRDNWKGANRPVITIDPEMMTHTSKYLCSKVLTDTSRPNDPTNGNCDAPITGGSPDPGVKGFAFLDGENSWLGRTISKDSRSGYE
IV     383-476   94 MLKVPNAETDIQSGPISNQVIVNNQNWSGYSGAFIDYWANKECFNPCFYVELIRGRPKESSVLWTSNSIVALCGSKKRLGSWSWHDGAEIIYFE
```

4FVK-10 [5] **4FVK.pdb neuraminidase-like molecule N10 derived from bat influenza A virus fragment**

```
                  123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          74-82    9 GSPSRSPEF
I        83-179   95 FYWRAKSQMCEVKGWVPTHRGFPWGPELPGDLILSRRAYVSCDLTSCFKFFIAYGLSANQHLLNTSMEWEESLYKTPIGSASTLSTSEMILPGRS
II     180-280  101 SSACFDGLKWTVLVANGRDRNSFIMIKYGEEVTDTFSASRGGPLRLPNSECICIEGSCFVIVSDGPNVNQSVHRIYELQNGTVQRWKQLNTTGINFEYSTC
III    281-374   90 YTINNLIKCTGTNLWNDAKRPLLRFTKELNYQIVEPCNGAPTDFPRGGLTTPSCKMAQEKGEGGIQGFILDEKPAWTSKTKAESSQNGFV
IV     375-442   81 LEQIPNGIESEGTVSLSYELFSNKRTGRSGFFQPKGDLISGCQRICFWLEIEDQTVGLGMIQELSTFCGINSPVQNINWDS
```

Figure Abbreviation, Reference Number, Sequence Identifier, And Sequence Descriptions
N = not included in Figure 1; I, II, III, and IV are Figure 1 sequence groupings
Numbers before sequence are: beginning sequence number - ending sequence number (as reported in PDB file) and number of amino acids in the row.

1259-11 [20] CY125947.1 neuraminidase-like protein [Influenza A virus (A/flat-faced bat/Peru/033/2010(H18N11))]

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N          1-84   84 MSFQTSTCLLIVSLICGILTVCLQVLLPFILIWTNTEPNYSCECPAPNISLSCPNGTSVTYDSKNITENSFYSSTTNYLSPVIA
I         85-179  95 TPLVLGENLCSINGWVPTYRGEGTTGKIPDEQMLTRQNFVSCSDKECRRFFVSMGYGTTTNFADLIVSEQMNVYSVKLGDPPTPDKLKFEAVGWS
II       180-278  98 ASSCHDGFQWTVLSVAGDGFVSILYGGIITDTIHPTNGGPLRTQASSCICNDGTCYTIIADGTTYTASSHRLYRLVNGTSAGWKALDTTGFNFEFPTC
III      279-368  90 YYTSGKVKCTGTNLWNDAKRPFLEFDQSFTYTFKEPCLGFLGDTPRGIDTTNYCDKTTTEGEGGIQGFMIEGSNSWIGRIINPGSKKGFE
IV       369-448  80 IYKFLGTLFSVQTVGNRNYQLLSNSTIGRSGLYQPAYESRDCQELCFWIEIAATTKAGLSSNDLITFCGTGGSMPDVNWG
```

4K3Y-11 [7] 4K3Y.pdb neuraminidase-like protein of A/flat-faced bat/Peru/033/2010(H18N11)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N            82    1 A
I         83-179   95 TPLVLGENLCSINGWVPTYRGEGTTGKIPDEQMLTRQNFVSCSDKECRRFFVSMGYGTTTNFADLIVSEQMNVYSVKLGDPPTPDKLKFEAVGWS
II        180-280  98 ASSCHDGFQWTVLSVAGDGFVSILYGGIITDTIHPTNGGPLRTQASSCICNDGTCYTIIADGTTYTASSHRLYRLVNGTSAGWKALDTTGFNFEFPTC
III       281-375  90 YYTSGKVKCTGTNLWNDAKRPFLEFDQSFTYTFKEPCLGFLGDTPRGIDTTNYCDKTTTEGEGGIQGFMIEGSNSWIGRIINPGSKKGFE
IV        376-459  80 IYKFLGTLFSVQTVGNRNYQLLSNSTIGRSGLYQPAYESRDCQELCFWIEIAATTKAGLSSNDLITFCGTGGSMPDVNWG
```

SPN [10] 3H72.pdb streptococcus pneumoniae D39 neuraminidase A precursor (complexed with NANA)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
N         317-319   3 LPE
I         320-417  98 GAALTEKTDIFESGRNGKPNKDGIKSYRIPALLKTDKGTLIAGADERRLHSSDWGDIGMVIRRSEDNGKTWGDRVTITNLRDNPKASDPSIGSPVNID
II        418-650 248 MVLVQDPETKRIFSIYDMFPEGKGIFGMSSQKEEAYKKIDGKTYQILYREGEKGAYTIRENGTVYTPDGKATDYRVVVDPVKPAYSDKGDLYKGNQLLGN
                      IYFTTNKTSPFRIAKDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYTTNNVSHLNGSQSSRIIYSDDHGKTWH
                      AGEAVNDNRQVDGQKIHSSTMNNRRAQNTESTV
III       651-724  74 VQLNNGDVKLFMRGLTGDLQVATSKDGGVTWEKDIKRYPQVKDVYVQMSAIHTMHEGKEYIILSNAGGPKRENGM
IV        725-793  69 VHLARVEENGELTWLKHNPIQKGEFAYNSLQELGNGEYGILYEHTEKGQNAYTLSFRKFNWDFLSKDL
```

2G9H [22] 2G9H.pdb staphylococcal enterotoxin I (SEI) chain D (complexed with a human MHC class II molecule)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
          1-218 218 QGDIGVGNLRNFYTKHDYIDLKGLIDKNLPSANQLEFSTGINDLISESNNWDEISKFKGKKLDIFGIDYNGPCKSKYMYGGATLSGQYLNSARKIPINLW
                      VNGKHKTISTDKISTNKKLVTAQEIDVKLRRYLQEEYNIYGHNSTGKGKEYGYKSKFYSGFNKGKVLFHLNDEKSFSYDLFYTGDGVPVSFLKIYEDNKI
                      IESEKFHLDVEISYVDSN
```

3RKD [23] 3RKD.pdb hepatitis E virus E2S domain genotype I (complexed with a neutralizing antibody)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
       448-604 147 SRPFSVLRANDVLWLSLTAAEYDQSTYGSSTGPVYVSDSVTLVNVATGAQAVARSLDWTKVTLDGRPLSTIQQHSKTFFVLPLRGKLSFWEAGTTKAGYP
                      YNYNTTASDQLLVENAAGHRVAISTYTTSLGAGPVSISAVAVLAPP
```

3SCK [24] 3SCK.pdb spike protein receptor-binding domain from a predicted SARS coronavirus civet strain (complexed with human-civet chimeric receptor ACE2 fragment)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
       324-502 179 PFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAW
                      NTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPAPNCYWPLRGYGFYTTTGIGYQPYRVVVLSFE
```

2ABX [25] 2ABX.pdb alpha-bungarotoxin complexed to acetylcholine receptor

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
          1-74   74 IVCHTTATIPSSAVTCPPGENLCYRKMWCDAFCSSRGKVVELGCAATCPSKKPYEEVTCCSTDKCNHPPKRQPG
```

2KS9 [26] 2KS9.pdb substance P in water (complexed with NK1R, substance-P receptor tachykinin receptor 1)

```
                  1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
       365-375  11 RPKPQQFFGLM
```

Figure Abbreviation, Reference Number, Sequence Identifier, And Sequence Descriptions
N = not included in Figure 1; I, II, III, and IV are Figure 1 sequence groupings
Numbers before sequence are: beginning sequence number - ending sequence number (as reported in PDB file) and number of amino acids in the row.

1DLL [29] 1DLL.pdb THE HC FRAGMENT OF TETANUS TOXIN RECEPTOR BINDING FRAGMENT HC FROM CLOSTRIDIUM TETANI

```
                1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
     693-1315   EDIDVILKKSTILNLDINNDIISDISGFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMFNNFTVSFWLRVPKVSASHLEQYGTNEY
                SIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDSAGEVRQITFRDLPDKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIREDNNITLKL
                DRCNNNNQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDVQLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLK
                FIIKRYTPNNEIDSFVKSGDFIKLYVSYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDDKNASLGLVGTHNGQ
                IGNDPNRDILIASNWYFNHLKDKILGCDWYFVPTDEGWTND
```

3ZUQ [28] 3ZUQ.pdb STRUCTURE OF AN ENGINEERED BOTULINUM NEUROTOXIN TYPE B FROM CLOSTRIDIUM BOTULINUM

```
                1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
       1-892    MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFNKSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIK
                SKPLGEKLLEMIINGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHFASREGFGGIMQMKFCPEYV
                SVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLYGIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV
                DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNI
                SDKDMEKEYRGQNKAINKQAYEEISKEHLAVYKIQMCVDGGGGSGGGGSGGGGSAIEGRAGGGGSGGGGSGGGGSALVLQCIDVDNEDLFFIADKNSFSD
                DLSKNERIEYNTQSNYIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPAIKKIFTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLF
                SNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLIPVVGA
                FLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGIN
                QAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNSLEALASG
                HHHHHH
```

1YQY [27] 1YQY.pdb STRUCTURE OF B. ANTHRAX LETHAL FACTOR

```
                1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
     257-780    GSKDPGMLSRYEKWEKIKQHYQHWSDSLSEEGRGLLKKLQIPIEPKKDDIIHSLSQEEKELLKRIQIDSSDFLSTEEKEFLKKLQIDIRDSLSEEEKELL
                NRIQVDSSNPLSEKEKEFLKKLKLDIQPYDINQRLQDTGGLIDSPSINLDVRKQYKRDIQNIDALLHQSIGSTLYNKIYLYENMNINNLTATLGAADLVD
                STDNTKINRGIFNEFKKNFKYSISSNYMIVDINERPALDNERLKWRIQLSPDTRAGYLENGKLILQRNIGLEIKDVQIIKQSEKEYIRIDAKVVPKSKID
                TKIQEAQLNINQEWNKALGLPKYTKLITFNVHNRYASNIVESAYLILNEWKNNIQSDLIKKVTNYLVDGNGRFVFTDITLPNIAEQYTHQDEIYEQVHSK
                GLYVPESRSILLHGPSKGVELRNDSEGFIHEFGHAVDDYAGYLLDKNQSDLVTNSKKFIDIFKEEGSNLTSYGRTNEAEFFAEAFRLMHSTDHAERLKVQ
                KNAPKTFQFINDQIKFIINSLVPR
```

```
ATOM   5868  N   ARG B 365      25.143  14.015 105.317  1.00  1.00           N
ATOM   5869  CA  ARG B 365      24.317  14.121 106.554  1.00  1.00           C
ATOM   5870  C   ARG B 365      24.505  12.854 107.384  1.00  1.00           C
ATOM   5871  O   ARG B 365      24.737  11.715 106.988  1.00  1.00           O
ATOM   5872  CB  ARG B 365      22.839  14.317 106.150  1.00  1.00           C
ATOM   5873  CG  ARG B 365      22.833  15.678 105.407  1.00  1.00           C
ATOM   5874  CD  ARG B 365      21.465  16.371 105.483  1.00  1.00           C
ATOM   5875  NE  ARG B 365      21.606  17.800 105.226  1.00  1.00           N
ATOM   5876  CZ  ARG B 365      20.705  18.638 105.726  1.00  1.00           C
ATOM   5877  NH1 ARG B 365      19.771  18.193 106.522  1.00  1.00           N
ATOM   5878  NH2 ARG B 365      20.754  19.907 105.422  1.00  1.00           N
ATOM   5894  N   PRO B 366      24.341  13.112 108.653  1.00  1.00           N
ATOM   5895  CA  PRO B 366      24.251  14.327 109.512  1.00  1.00           C
ATOM   5896  C   PRO B 366      25.622  14.975 109.649  1.00  1.00           C
ATOM   5897  O   PRO B 366      26.634  14.281 109.744  1.00  1.00           O
ATOM   5898  CB  PRO B 366      23.736  13.786 110.855  1.00  1.00           C
ATOM   5899  CG  PRO B 366      24.164  12.355 110.874  1.00  1.00           C
ATOM   5900  CD  PRO B 366      24.090  11.881 109.421  1.00  1.00           C
ATOM   5908  N   LYS B 367      26.770  14.329 109.791  1.00  1.00           N
ATOM   5909  CA  LYS B 367      27.972  15.062 110.152  1.00  1.00           C
ATOM   5910  C   LYS B 367      28.333  16.041 109.030  1.00  1.00           C
ATOM   5911  O   LYS B 367      29.021  16.981 109.430  1.00  1.00           O
ATOM   5912  CB  LYS B 367      29.133  14.091 110.387  1.00  1.00           C
ATOM   5913  CG  LYS B 367      28.996  13.083 111.533  1.00  1.00           C
ATOM   5914  CD  LYS B 367      29.066  13.913 112.822  1.00  1.00           C
ATOM   5915  CE  LYS B 367      28.871  12.899 113.953  1.00  1.00           C
ATOM   5916  NZ  LYS B 367      27.414  12.719 114.207  1.00  1.00           N
ATOM   5930  N   PRO B 368      28.565  16.453 107.807  1.00  1.00           N
ATOM   5931  CA  PRO B 368      29.636  15.856 106.952  1.00  1.00           C
ATOM   5932  C   PRO B 368      31.029  16.188 107.479  1.00  1.00           C
ATOM   5933  O   PRO B 368      31.238  17.237 108.089  1.00  1.00           O
ATOM   5934  CB  PRO B 368      29.386  16.488 105.576  1.00  1.00           C
ATOM   5935  CG  PRO B 368      28.763  17.802 105.891  1.00  1.00           C
ATOM   5936  CD  PRO B 368      27.861  17.539 107.096  1.00  1.00           C
ATOM   5944  N   GLN B 369      31.983  15.298 107.233  1.00  1.00           N
ATOM   5945  CA  GLN B 369      33.347  15.527 107.682  1.00  1.00           C
ATOM   5946  C   GLN B 369      33.990  16.644 106.867  1.00  1.00           C
ATOM   5947  O   GLN B 369      33.918  16.647 105.638  1.00  1.00           O
ATOM   5948  CB  GLN B 369      34.167  14.241 107.548  1.00  1.00           C
ATOM   5949  CG  GLN B 369      35.453  14.526 106.766  1.00  1.00           C
ATOM   5950  CD  GLN B 369      36.657  14.456 107.700  1.00  1.00           C
ATOM   5951  NE2 GLN B 369      36.864  13.377 108.402  1.00  1.00           N
ATOM   5952  OE1 GLN B 369      37.428  15.411 107.791  1.00  1.00           O
ATOM   5961  N   GLN B 370      34.616  17.595 107.559  1.00  1.00           N
ATOM   5962  CA  GLN B 370      35.270  18.725 106.891  1.00  1.00           C
ATOM   5963  C   GLN B 370      36.780  18.519 106.838  1.00  1.00           C
ATOM   5964  O   GLN B 370      37.447  18.482 107.872  1.00  1.00           O
ATOM   5965  CB  GLN B 370      34.962  20.019 107.646  1.00  1.00           C
ATOM   5966  CG  GLN B 370      33.548  19.949 108.227  1.00  1.00           C
ATOM   5967  CD  GLN B 370      32.635  20.926 107.495  1.00  1.00           C
ATOM   5968  NE2 GLN B 370      32.123  21.938 108.142  1.00  1.00           N
ATOM   5969  OE1 GLN B 370      32.381  20.764 106.302  1.00  1.00           O
ATOM   5978  N   PHE B 371      37.311  18.394 105.627  1.00  1.00           N
ATOM   5979  CA  PHE B 371      38.746  18.202 105.448  1.00  1.00           C
ATOM   5980  C   PHE B 371      39.512  19.419 105.954  1.00  1.00           C
```

```
ATOM   5981  O   PHE B 371      40.545  19.290 106.610  1.00  1.00           O
ATOM   5982  CB  PHE B 371      39.062  17.972 103.969  1.00  1.00           C
ATOM   5983  CG  PHE B 371      40.390  17.267 103.840  1.00  1.00           C
ATOM   5984  CD1 PHE B 371      41.575  18.010 103.794  1.00  1.00           C
ATOM   5985  CD2 PHE B 371      40.436  15.868 103.766  1.00  1.00           C
ATOM   5986  CE1 PHE B 371      42.807  17.356 103.674  1.00  1.00           C
ATOM   5987  CE2 PHE B 371      41.668  15.215 103.647  1.00  1.00           C
ATOM   5988  CZ  PHE B 371      42.854  15.959 103.600  1.00  1.00           C
ATOM   5998  N   PHE B 372      38.996  20.603 105.638  1.00  1.00           N
ATOM   5999  CA  PHE B 372      39.630  21.846 106.059  1.00  1.00           C
ATOM   6000  C   PHE B 372      39.602  21.976 107.578  1.00  1.00           C
ATOM   6001  O   PHE B 372      40.541  22.491 108.186  1.00  1.00           O
ATOM   6002  CB  PHE B 372      38.914  23.040 105.426  1.00  1.00           C
ATOM   6003  CG  PHE B 372      39.928  24.107 105.081  1.00  1.00           C
ATOM   6004  CD1 PHE B 372      41.290  23.873 105.293  1.00  1.00           C
ATOM   6005  CD2 PHE B 372      39.501  25.330 104.550  1.00  1.00           C
ATOM   6006  CE1 PHE B 372      42.229  24.861 104.973  1.00  1.00           C
ATOM   6007  CE2 PHE B 372      40.439  26.319 104.231  1.00  1.00           C
ATOM   6008  CZ  PHE B 372      41.803  26.084 104.443  1.00  1.00           C
ATOM   6018  N   GLY B 373      38.516  21.511 108.183  1.00  1.00           N
ATOM   6019  CA  GLY B 373      38.371  21.583 109.632  1.00  1.00           C
ATOM   6020  C   GLY B 373      39.491  20.822 110.327  1.00  1.00           C
ATOM   6021  O   GLY B 373      39.932  21.201 111.412  1.00  1.00           O
ATOM   6025  N   LEU B 374      39.952  19.745 109.698  1.00  1.00           N
ATOM   6026  CA  LEU B 374      41.026  18.942 110.272  1.00  1.00           C
ATOM   6027  C   LEU B 374      42.328  19.734 110.276  1.00  1.00           C
ATOM   6028  O   LEU B 374      42.713  20.322 109.265  1.00  1.00           O
ATOM   6029  CB  LEU B 374      41.195  17.644 109.465  1.00  1.00           C
ATOM   6030  CG  LEU B 374      40.934  16.401 110.367  1.00  1.00           C
ATOM   6031  CD1 LEU B 374      40.336  15.260 109.529  1.00  1.00           C
ATOM   6032  CD2 LEU B 374      42.250  15.912 111.003  1.00  1.00           C
ATOM   6044  N   MET B 375      43.004  19.745 111.423  1.00  1.00           N
ATOM   6045  CA  MET B 375      44.267  20.471 111.557  1.00  1.00           C
ATOM   6046  C   MET B 375      43.554  21.819 111.540  1.00  1.00           C
ATOM   6047  O   MET B 375      42.458  21.893 112.070  1.00  1.00           O
ATOM   6048  CB  MET B 375      45.685  20.651 112.104  1.00  1.00           C
ATOM   6049  CG  MET B 375      46.204  19.312 112.635  1.00  1.00           C
ATOM   6050  SD  MET B 375      47.630  18.787 111.649  1.00  1.00           S
ATOM   6051  CE  MET B 375      48.891  19.772 112.497  1.00  1.00           C
ATOM   6052  OXT MET B 375      44.116  22.755 110.997  1.00  1.00           O
END
```

>> N6N, N10P, N11P, IBN, and SPN Example of Common Distributed Relative Spatial Occupancy <<

Original Data from 1W1X.pdb (22 points):

```
ATOM    293  O   ARG A 124      32.542   4.910  59.543  1.00 14.80           O
ATOM    770  O   TRP A 185      39.672   4.897  56.458  1.00 13.17           O
ATOM    784  O   SER A 186      40.878   0.283  56.594  1.00 15.41           O
ATOM    825  O   ASP A 192      34.555 -19.986  56.865  1.00 18.02           O
ATOM    904  O   GLY A 203      48.222   7.443  56.825  1.00 18.27           O
ATOM    983  O   TYR A 214      43.328 -13.981  64.025  1.00 16.85           O
ATOM   1105  O   LEU A 230      44.646   2.786  51.864  1.00 18.46           O
ATOM   1113  O   ARG A 231      41.929  -0.134  49.208  1.00 16.63           O
ATOM   1149  O   SER A 235      33.584  -4.759  53.629  1.00 14.85           O
ATOM   1202  O   GLY A 242      40.087 -18.835  53.462  1.00 18.72           O
ATOM   1255  O   ASP A 250      46.557   2.834  43.371  1.00 19.94           O
ATOM   1263  O   GLY A 251      44.703   6.760  44.397  1.00 20.15           O
ATOM   1389  O   GLY A 267      46.987 -13.160  56.350  1.00 17.71           O
ATOM   1511  O   GLU A 283      34.396   1.036  45.104  1.00 15.09           O
ATOM   1877  O   ASP A 331      24.511   8.068  41.321  1.00 16.01           O
ATOM   1898  O   ARG A 334      22.914  13.063  37.830  1.00 16.82           O
ATOM   2137  O   TRP A 368      19.911  -4.598  43.297  1.00 19.40           O
ATOM   2218  O   ARG A 378      22.228  12.223  51.392  1.00 17.66           O
ATOM   2235  O   GLY A 380      18.788   8.473  49.691  1.00 17.40           O
ATOM   2492  O   SER A 413      26.956   0.641  50.022  1.00 16.30           O
ATOM   2659  O   GLU A 433      23.619   6.960  55.288  1.00 16.40           O
ATOM   2914  O   TRP A 466      16.074   5.722  58.706  1.00 14.95           O
```

Distance Matrix for above 22 points from 1W1X.pdb:

	ARG124	TRP185	SER186	ASP192	GLY203	TYR214	LEU230	ARG231	SER235	GLY242	ASP250	GLY251	GLY267	GLU283	ASP331	ARG334	TRP368	ARG378	GLY380	SER413	GLU433	TRP466
ARG124	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TRP185	7.769	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER186	9.980	4.771	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ASP192	25.120	25.407	21.234	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY203	16.114	8.929	10.259	30.645	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TYR214	22.210	20.664	16.269	12.818	23.125	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
LEU230	14.491	7.092	6.545	25.405	7.687	20.755	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ARG231	14.845	9.109	7.472	22.519	12.451	20.328	4.792	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER235	11.382	11.760	9.350	15.597	19.323	16.973	13.506	10.515	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY242	25.646	23.924	19.389	6.596	27.713	12.068	22.154	19.267	15.506	0.000	------	------	------	------	------	------	------	------	------	------	------	------
ASP250	21.500	14.931	14.615	29.101	14.319	26.828	8.705	8.019	18.198	24.764	0.000	------	------	------	------	------	------	------	------	------	------	------
GLY251	19.512	13.200	14.330	31.205	12.935	28.589	8.459	8.853	18.481	27.542	4.461	0.000	------	------	------	------	------	------	------	------	------	------
GLY267	23.353	19.483	14.768	14.192	20.645	8.542	16.730	15.693	16.051	9.389	20.602	23.343	0.000	------	------	------	------	------	------	------	------	------
GLU283	15.064	13.102	13.214	24.089	19.225	25.755	12.403	8.658	10.340	22.296	12.415	11.811	22.057	0.000	------	------	------	------	------	------	------	------
ASP331	20.162	21.657	23.701	33.608	28.337	36.820	23.334	20.805	19.958	33.373	22.751	20.467	34.375	12.707	0.000	------	------	------	------	------	------	------
ARG334	25.112	26.354	28.950	39.876	32.139	42.829	27.836	25.791	26.098	39.456	26.350	23.614	40.127	18.149	6.300	0.000	------	------	------	------	------	------
TRP368	22.669	25.571	25.303	25.206	33.608	32.650	27.198	23.231	17.138	26.704	27.663	27.292	31.254	15.647	13.620	18.730	0.000	------	------	------	------	------
ARG378	15.043	19.586	22.747	34.919	26.982	35.937	24.328	23.358	20.551	35.886	27.284	24.164	35.803	17.685	11.131	13.605	18.811	0.000	------	------	------	------
GLY380	17.290	22.242	24.550	33.316	30.304	36.220	26.565	24.695	20.236	34.837	29.032	26.506	36.160	17.887	10.148	13.371	14.594	5.366	0.000	------	------	------
SER413	11.835	14.874	15.399	23.023	23.341	26.037	17.915	15.015	9.279	23.740	20.815	19.597	25.135	8.927	11.698	17.869	11.059	12.585	11.321	0.000	------	------
GLU433	10.096	16.227	18.552	29.123	24.656	30.055	21.709	20.556	15.472	30.658	26.176	23.732	30.855	15.967	14.039	18.507	17.062	6.694	7.547	8.877	0.000	------
TRP466	16.509	23.719	25.481	31.715	32.249	34.048	29.526	28.160	21.029	34.744	34.245	32.023	36.300	23.295	19.466	23.162	18.938	11.560	9.808	14.820	8.375	0.000

Original Data from 4FVK.pdb (22 points):

```
ATOM    293  O   ARG A 118      -7.217 -39.973  23.523  1.00 19.36           O
ATOM    557  O   TRP A 154       4.466 -35.386  28.327  1.00 33.21           O
ATOM    756  O   SER A 179      -2.336 -34.410  18.061  1.00 21.63           O
ATOM    796  O   ASP A 185      -7.099 -42.180  -1.165  1.00 23.11           O
ATOM    881  O   GLY A 196       1.976 -27.673  22.767  1.00 41.18           O
ATOM    973  O   TYR A 207       4.146 -41.648   4.669  1.00 25.84           O
ATOM   1090  O   LEU A 223      -2.078 -27.428  19.099  1.00 26.36           O
ATOM   1098  O   ARG A 224      -5.797 -27.972  16.023  1.00 23.11           O
ATOM   1132  O   SER A 228      -9.715 -36.426  13.059  1.00 20.90           O
ATOM   1184  O   GLY A 235      -4.573 -36.000  -1.891  1.00 26.30           O
ATOM   1239  O   ASP A 243      -6.321 -20.388  18.236  1.00 29.31           O
ATOM   1247  O   GLY A 244      -4.992 -20.515  21.506  1.00 31.41           O
ATOM   1378  O   GLY A 260       2.514 -33.745   4.489  1.00 25.11           O
ATOM   1509  O   GLU A 276     -14.628 -27.882  17.679  1.00 21.34           O
ATOM   1887  O   ASP A 324     -24.904 -29.692  25.184  1.00 29.53           O
```

```
ATOM   1913  O   ARG A 327     -27.588 -26.475  29.864  1.00 39.08           O
ATOM   2129  O   TRP A 361     -27.898 -34.298  13.349  1.00 24.80           O
ATOM   2454  O   ARG A 406     -17.968 -39.082  22.230  1.00 17.66           O
ATOM   2224  O   GLY A 373     -26.151 -39.431  26.618  1.00 27.35           O
ATOM   2465  O   SER A 407     -17.954 -36.270  18.433  1.00 16.93           O
ATOM   2633  O   GLU A 425     -19.239 -41.968  25.634  1.00 24.31           O
ATOM   2851  O   TRP A 458     -26.179 -50.268  26.304  1.00 34.09           O
```

Distance Matrix for above 22 points from 4FVK.pdb:

	ARG118	TRP154	SER179	ASP185	GLY196	TYR207	LEU223	ARG224	SER228	GLY235	ASP243	GLY244	GLY260	GLU276	ASP324	ARG327	TRP361	ARG406	GLY373	SER407	GLU425	TRP458
ARG118	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TRP154	13.439	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER179	9.198	12.354	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ASP185	24.787	32.399	21.277	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY196	15.374	9.829	9.280	29.420	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TYR207	22.077	24.475	16.545	12.679	22.968	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
LEU223	14.260	13.831	7.063	25.563	5.473	21.194	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ARG224	14.223	17.655	7.588	22.338	10.295	20.367	4.857	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER228	11.328	20.864	9.140	15.565	17.537	17.023	13.258	9.778	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY235	25.858	31.547	20.140	6.716	26.837	12.286	22.810	19.669	15.815	0.000	------	------	------	------	------	------	------	------	------	------	------	------
ASP243	20.306	21.051	14.578	29.187	11.935	27.306	8.265	7.918	17.191	25.532	0.000	------	------	------	------	------	------	------	------	------	------	------
GLY244	19.688	18.898	14.560	31.429	10.069	28.524	7.879	9.291	18.623	28.060	3.532	0.000	------	------	------	------	------	------	------	------	------	------
GLY260	22.266	23.974	14.428	13.983	19.268	8.072	16.566	15.344	15.172	9.799	21.106	22.824	0.000	------	------	------	------	------	------	------	------	------
GLU276	15.338	23.114	13.923	24.824	17.367	26.669	12.638	8.985	10.885	23.452	11.202	12.719	22.410	0.000	------	------	------	------	------	------	------	------
ASP324	20.525	30.082	24.131	34.165	27.064	37.520	23.731	21.259	20.569	34.441	21.913	22.231	34.590	12.853	0.000	------	------	------	------	------	------	------
ARG327	25.246	33.305	28.982	40.364	30.427	43.267	27.705	25.858	26.474	40.358	24.991	24.818	40.036	17.844	6.281	0.000	------	------	------	------	------	------
TRP361	23.736	35.678	25.993	26.559	32.016	34.003	27.330	23.144	18.309	27.914	26.133	27.950	31.681	15.363	13.048	18.277	0.000	------	------	------	------	------
ARG406	10.865	23.540	16.839	25.982	22.983	28.355	19.953	17.609	12.620	27.762	22.385	22.664	27.618	12.542	12.042	17.600	14.155	0.000	------	------	------	------
GLY373	19.193	30.930	25.799	33.800	30.728	37.478	27.931	25.649	21.518	35.919	28.742	28.838	36.657	18.603	9.923	13.434	14.334	9.292	0.000	------	------	------
SER407	12.446	24.522	15.733	23.170	22.134	26.585	18.184	14.915	9.838	24.335	19.688	20.632	24.895	9.055	11.711	17.872	11.341	4.725	12.007	0.000	------	------
GLU425	12.368	24.749	20.005	29.421	25.742	31.408	23.423	21.655	16.720	31.754	26.216	26.082	31.431	16.821	13.528	18.101	16.874	4.640	7.428	9.272	0.000	------
TRP458	21.755	34.127	29.798	34.409	36.273	38.236	33.977	31.910	25.261	38.280	36.773	36.840	39.651	26.626	20.646	24.099	20.636	14.462	10.842	18.043	10.840	0.000

Original Data from 4K3Y.pdb (22 points):

```
ATOM    273  O   ARG A 118      86.013 -31.444   7.522  1.00 56.67           O
ATOM    656  O   TRP A 178      88.849 -34.460   2.688  1.00 49.42           O
ATOM    670  O   SER A 179      93.277 -35.859   4.171  1.00 53.94           O
ATOM    709  O   ASP A 185     105.225 -37.767  21.429  1.00 52.27           O
ATOM    795  O   GLY A 196      94.498 -34.397  -6.670  1.00 50.02           O
ATOM    851  O   TYR A 207     106.946 -30.939  10.936  1.00 63.94           O
ATOM    964  O   LEU A 223      92.738 -41.170  -0.317  1.00 52.48           O
ATOM    972  O   ARG A 224      92.961 -43.636   3.851  1.00 62.04           O
ATOM   1004  O   SER A 228      92.319 -39.210  12.455  1.00 45.89           O
ATOM   1052  O   GLY A 235     107.336 -41.621  17.743  1.00 70.71           O
ATOM   1109  O   ASP A 243      91.692 -49.165  -1.492  1.00 50.57           O
ATOM   1117  O   GLY A 244      89.925 -47.451  -4.112  1.00 55.92           O
ATOM   1246  O   GLY A 260     106.478 -38.506   8.356  1.00 57.83           O
ATOM   1364  O   GLU A 276      86.341 -47.175   8.114  1.00 41.55           O
ATOM   1748  O   ASP A 324      73.896 -48.836  10.981  1.00 51.88           O
ATOM   1770  O   ARG A 327      68.828 -50.633   8.330  1.00 63.20           O
ATOM   1990  O   TRP A 361      80.624 -50.235  21.761  1.00 60.69           O
ATOM   2331  O   ARG A 406      79.890 -38.238  13.499  1.00 52.56           O
ATOM   2086  O   GLY A 373      70.977 -40.796  16.164  1.00 52.43           O
ATOM   2342  O   SER A 407      82.932 -41.872  14.302  1.00 45.15           O
ATOM   2510  O   GLU A 425      77.347 -35.609  12.904  1.00 60.39           O
ATOM   2708  O   TRP A 458      71.677 -31.201  23.071  1.00 65.69           O
```

Distance Matrix for above 22 points from 4K3Y.pdb:

	ARG118	TRP178	SER179	ASP185	GLY196	TYR207	LEU223	ARG224	SER228	GLY235	ASP243	GLY244	GLY260	GLU276	ASP324	ARG327	TRP361	ARG406	GLY373	SER407	GLU425	TRP458
ARG118	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TRP178	6.364	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER179	9.137	4.875	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ASP185	24.546	25.106	21.077	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY196	16.797	10.931	11.007	30.265	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TYR207	21.216	20.197	16.025	12.637	21.838	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
LEU223	14.187	8.317	6.974	25.306	9.452	20.813	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ARG224	14.505	10.122	7.790	22.222	14.086	20.174	4.848	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------

SER228	11.154	11.402	8.987	15.785	19.841	16.872	12.928	9.697	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY235	25.743	24.894	20.373	5.736	28.513	12.673	23.226	20.092	16.102	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------
ASP243	20.677	15.550	14.548	28.956	15.899	26.820	8.148	7.793	17.147	25.916	0.000	------	------	------	------	------	------	------	------	------	------	------	------
GLY244	20.171	14.703	14.636	31.308	14.066	28.086	7.859	9.337	18.658	28.544	3.595	0.000	------	------	------	------	------	------	------	------	------	------	------
GLY260	21.665	18.955	14.099	13.154	19.652	8.008	16.465	15.143	14.757	9.927	20.718	22.571	0.000	------	------	------	------	------	------	------	------	------	------
GLU276	15.746	14.050	13.846	24.948	21.175	26.384	12.168	8.633	10.864	23.756	11.174	12.743	21.925	0.000	------	------	------	------	------	------	------	------	------
ASP324	21.477	22.339	24.298	34.831	30.732	37.585	23.269	21.008	20.838	34.871	21.734	22.060	34.281	12.879	0.000	------	------	------	------	------	------	------	------
ARG327	25.772	26.348	28.867	40.766	33.876	42.984	27.129	25.523	26.445	40.653	24.928	24.698	39.555	17.852	5.995	0.000	------	------	------	------	------	------	------
TRP361	24.185	26.082	26.003	27.582	35.379	34.386	26.765	22.727	18.572	28.353	25.775	27.635	31.396	15.109	12.784	17.880	0.000	------	------	------	------	------	------
ARG406	10.926	14.540	16.489	26.551	25.198	28.140	19.093	17.119	12.511	27.977	21.987	22.265	27.082	12.267	12.433	17.399	14.585	0.000	------	------	------	------	------
GLY373	19.703	23.263	25.797	34.782	33.400	37.660	27.300	25.357	21.720	36.403	28.476	28.538	36.422	18.481	10.001	12.758	14.611	9.648	0.000	------	------	------	------
SER407	12.814	14.994	15.678	23.762	25.089	26.599	17.617	14.592	9.930	24.647	19.478	20.472	24.517	8.834	11.882	17.645	11.441	4.807	12.147	0.000	------	------	------
GLU425	11.019	15.427	18.168	29.232	26.053	30.030	21.038	19.753	15.406	30.966	24.429	24.248	29.626	15.415	13.804	17.867	17.410	3.706	8.838	8.507	0.000	------	------
TRP458	21.151	26.851	29.077	34.224	37.624	37.299	33.014	31.258	24.555	37.530	36.423	36.551	38.484	26.342	21.496	24.556	21.073	14.443	11.843	17.817	12.448	0.000	------

Original Data from 1A4G.pdb (22 points):

ATOM	312	O	ARG A 115	4.385	50.615	-4.433	1.00	5.48	O
ATOM	796	O	TRP A 176	0.827	51.334	-11.511	1.00	6.60	O
ATOM	810	O	SER A 177	-0.271	46.725	-11.977	1.00	5.12	O
ATOM	847	O	ASP A 183	-6.159	27.896	-4.274	1.00	8.37	O
ATOM	939	O	GLY A 194	1.200	52.406	-20.699	1.00	10.76	O
ATOM	1019	O	TYR A 205	2.374	30.784	-13.885	1.00	13.88	O
ATOM	1152	O	LEU A 221	-4.704	50.750	-16.197	1.00	9.50	O
ATOM	1160	O	ARG A 222	-7.500	48.630	-13.066	1.00	12.18	O
ATOM	1196	O	SER A 226	-4.636	43.422	-4.848	1.00	10.68	O
ATOM	1239	O	GLY A 233	-9.130	29.475	-9.672	1.00	9.53	O
ATOM	1300	O	ASP A 241	-12.761	53.386	-16.618	1.00	11.87	O
ATOM	1308	O	GLY A 242	-10.068	56.852	-15.900	1.00	11.10	O
ATOM	1429	O	GLY A 258	-4.430	33.958	-16.864	1.00	7.85	O
ATOM	1559	O	GLU A 274	-11.159	52.010	-6.042	1.00	9.27	O
ATOM	1943	O	ASP A 323	-11.543	60.383	3.557	1.00	13.23	O
ATOM	1965	O	ARG A 326	-13.300	66.002	4.585	1.00	12.54	O
ATOM	2232	O	TRP A 363	-11.991	46.912	7.597	1.00	9.84	O
ATOM	2321	O	ARG A 373	-0.214	61.260	5.138	1.00	6.97	O
ATOM	2340	O	GLY A 375	-3.198	58.362	8.633	1.00	11.08	O
ATOM	2610	O	SER A 409	-5.482	50.040	1.766	1.00	7.86	O
ATOM	2748	O	GLU A 427	1.210	54.622	4.031	1.00	9.86	O
ATOM	2956	O	TRP A 455	-0.035	50.826	18.009	1.00	11.17	O

Distance Matrix for above 22 points from 1A4G.pdb:

	ARG115	TRP176	SER177	ASP183	GLY194	TYR205	LEU221	ARG222	SER226	GLY233	ASP241	GLY242	GLY258	GLU274	ASP323	ARG326	TRP363	ARG373	GLY375	SER409	GLU427	TRP455
ARG115	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TRP176	7.955	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER177	9.681	4.761	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ASP183	25.047	25.505	21.179	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY194	16.671	9.258	10.512	30.408	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
TYR205	22.060	20.744	16.271	13.173	22.701	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
LEU221	14.867	7.273	7.325	25.818	7.607	21.309	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------
ARG222	14.823	8.892	7.555	22.561	12.174	20.412	4.703	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------	------
SER226	11.545	11.698	8.988	15.611	19.132	17.045	13.509	10.142	0.000	------	------	------	------	------	------	------	------	------	------	------	------	------
GLY233	25.632	24.090	19.528	6.361	27.462	12.321	22.689	19.522	15.427	0.000	------	------	------	------	------	------	------	------	------	------	------	------
ASP241	21.216	14.660	14.897	29.081	14.578	27.338	8.488	7.932	17.431	25.163	0.000	------	------	------	------	------	------	------	------	------	------	------
GLY242	19.475	12.977	14.626	31.447	13.029	28.955	8.130	9.068	18.221	28.092	4.448	0.000	------	------	------	------	------	------	------	------	------	------
GLY258	22.576	18.927	14.289	14.080	19.666	8.077	16.807	15.463	15.297	9.691	21.140	23.598	0.000	------	------	------	------	------	------	------	------	------
GLU274	15.689	13.192	13.480	24.690	19.176	26.367	12.099	8.611	10.850	22.915	10.785	11.037	22.097	0.000	------	------	------	------	------	------	------	------
ASP323	20.321	21.493	23.556	33.848	28.537	37.068	23.017	20.756	20.150	33.707	21.389	19.830	34.145	12.743	0.000	------	------	------	------	------	------	------
ARG326	25.117	25.958	28.560	39.769	32.162	42.745	27.174	25.436	25.960	39.432	24.678	22.667	39.567	17.700	5.976	0.000	------	------	------	------	------	------
TRP363	20.654	23.430	22.815	23.163	31.699	30.462	25.179	21.215	14.871	24.707	25.077	25.585	28.693	14.584	14.071	19.370	0.000	------	------	------	------	------
ARG373	15.036	19.411	22.454	35.172	27.349	36.019	24.203	23.324	20.916	36.182	26.320	23.646	35.317	18.175	11.472	13.930	18.725	0.000	------	------	------	------
GLY375	16.978	21.711	23.849	33.220	30.252	36.037	26.014	24.167	20.174	34.709	27.456	25.521	35.315	17.863	9.974	13.297	14.474	5.433	0.000	------	------	------
SER409	11.667	14.757	15.067	22.963	23.557	26.028	17.994	15.035	9.395	23.813	20.054	19.481	24.634	9.853	12.121	17.996	9.282	12.846	11.029	0.000	------	------
GLU427	9.888	15.891	17.911	28.941	24.829	29.843	21.428	20.102	15.442	30.448	24.962	23.009	29.923	16.164	14.002	18.449	15.698	6.879	7.389	8.421	0.000	------
TRP455	22.874	29.537	30.266	32.555	38.760	37.745	34.523	32.034	24.463	36.122	36.980	35.872	38.987	26.525	20.800	24.217	16.330	16.570	12.438	17.150	14.538	0.000

Original Data from 3H72.pdb (22 points):

ATOM	207	O	ARG A 347	111.803	1.632	39.130	1.00	7.51	O
ATOM	405	O	TRP A 373	110.167	1.568	28.102	1.00	6.74	O

```
ATOM    744  O   ASP A 417     103.433  -1.127  38.543  1.00  8.64           O
ATOM    791  O   ASP A 423     100.196 -11.521  55.174  1.00  9.70           O
ATOM    943  O   GLY A 441     100.598   3.718  26.430  1.00  8.28           O
ATOM   1737  O   TYR A 540     104.043 -15.109  44.030  1.00  7.28           O
ATOM   1949  O   LEU A 566      95.771   0.538  34.798  1.00  7.59           O
ATOM   1957  O   GLY A 567      96.819   2.819  38.767  1.00  5.20           O
ATOM   1983  O   THR A 572     102.133   1.211  50.147  1.00  9.30           O
ATOM   2066  O   GLY A 583      92.715  -7.377  54.605  1.00  8.89           O
ATOM   2119  O   TYR A 590      94.546   4.781  37.063  1.00  6.33           O
ATOM   2131  O   THR A 591      92.778   8.671  35.535  1.00  5.40           O
ATOM   2302  O   GLY A 613      93.961 -12.098  45.305  1.00 11.08           O
ATOM   2569  O   GLU A 647      99.490  10.112  43.558  1.00  9.05           O
ATOM   2848  O   ASP A 684      89.051  16.817  52.165  1.00 29.08           O
ATOM   2873  O   ARG A 687      96.219  22.335  52.871  1.00 16.33           O
ATOM   3064  O   TYR A 710     101.786   8.428  61.840  1.00 11.15           O
ATOM   3143  O   ARG A 721     114.937  16.069  41.419  1.00  9.88           O
ATOM   3171  O   GLY A 724     115.210  15.718  51.246  1.00 12.19           O
ATOM   3408  O   ASN A 753     108.332   6.063  47.305  1.00  7.41           O
ATOM   3525  O   GLU A 768     115.916   9.876  43.732  1.00  7.92           O
ATOM   3678  O   TRP A 786     113.666  11.708  65.474  1.00 26.65           O

Distance Matrix for above 22 points from 3H72.pdb:

          ARG347  TRP373  ASP417  ASP423  GLY441  TYR540  LEU566  GLY567  THR572  GLY583  TYR590  THR591  GLY613  GLU647  ASP684  ARG687  TYR710  ARG721  GLY724  ASN753  GLU768  TRP786
ARG347     0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
TRP373    11.149   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
ASP417     8.833  12.713   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
ASP423    23.773  31.680  19.877   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
GLY441    17.064   9.949  13.351  32.536   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
TYR540    19.092  23.861  15.032  12.323  26.002   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
LEU566    16.643  15.910   8.689  24.087  10.170  19.962   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
GLY567    15.035  17.131   7.705  22.051  12.934  20.032   4.696   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
THR572    14.665  23.466  11.908  13.825  23.898  17.533  16.629  12.662   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
GLY583    26.172  32.970  20.296   8.571  31.290  17.319  21.548  19.278  13.503   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
TYR590    17.663  18.293  10.774  25.014  12.281  23.116   4.963   3.452  15.540  21.422   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
THR591    20.602  20.201  14.784  29.128  12.984  27.651   8.698   7.812  18.886  24.924   4.538   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
GLY613    23.345  27.301  15.994  11.688  25.504  10.599  16.533  16.536  16.351  10.504  18.793  22.983   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------
GLU647    15.593  20.637  12.923  24.565  18.316  25.633  13.499   9.126  11.385  21.767   9.749  10.559  22.954   0.000  ------  ------  ------  ------  ------  ------  ------  ------
ASP684    30.301  35.460  26.728  30.599  31.100  36.197  24.734  20.876  20.464  24.591  20.078  18.889  30.121  15.100   0.000  ------  ------  ------  ------  ------  ------  ------
ARG687    29.331  35.204  28.422  34.166  32.633  39.261  28.319  24.086  22.105  29.968  23.682  22.340  35.327  15.711   9.073   0.000  ------  ------  ------  ------  ------  ------
TYR710    25.735  35.434  25.234  21.093  35.742  29.602  28.805  24.259  13.745  19.607  26.069  27.806  27.495  18.502  18.060  17.460   0.000  ------  ------  ------  ------  ------
ARG721    14.950  20.258  20.888  34.172  24.142  33.130  25.542  22.602  21.468  34.891  23.710  24.091  35.334  16.693  28.038  22.820  25.463   0.000  ------  ------  ------  ------
GLY724    18.890  27.592  24.162  31.350  31.198  33.572  29.645  25.697  19.562  32.414  27.345  28.279  35.504  18.375  26.198  20.176  18.590   9.837   0.000  ------  ------  ------
ASN753     9.925  19.807  12.348  20.912  22.385  21.849  18.567  14.696   8.369  21.859  17.222  19.679  23.245  10.422  22.606  21.035  16.116  13.356  12.492   0.000  ------  ------
GLU768    10.298  18.611  17.430  28.911  23.915  27.664  23.934  20.956  17.499  30.890  22.959  24.577  31.102  16.429  29.000  25.034  23.014   6.683   9.544   9.210   0.000  ------
TRP786    28.267  38.881  31.539  28.760  41.941  35.659  37.229  32.804  21.866  30.353  34.939  36.632  36.903  26.150  28.445  24.003  12.849  24.480  14.863  19.759  21.935   0.000

Population Standard Deviation of five sets of interatomic distances for 22 atoms from structures 1W1X.pdb (N6N) 4FVK.pdb (N10P) 4K3Y.pdb (N11P) 1A4G.pdb (IBN) and 3H72.pdb (SPN):

1W1X-->  ARG124  TRP185  SER186  ASP192  GLY203  TYR214  LEU230  ARG231  SER235  GLY242  ASP250  GLY251  GLY267  GLU283  ASP331  ARG334  TRP368  ARG378  GLY380  SER413  GLU433  TRP466
4FVK-->  ARG118  TRP154  SER179  ASP185  GLY196  TYR207  LEU223  ARG224  SER228  GLY235  ASP243  GLY244  GLY260  GLU276  ASP324  ARG327  TRP361  ARG406  GLY373  SER407  GLU425  TRP458
4K3Y-->  ARG118  TRP178  SER179  ASP185  GLY196  TYR207  LEU223  ARG224  SER228  GLY235  ASP243  GLY244  GLY260  GLU276  ASP324  ARG327  TRP361  ARG406  GLY373  SER407  GLU425  TRP458
1A4G-->  ARG115  TRP176  SER177  ASP183  GLY194  TYR205  LEU221  ARG222  SER226  GLY233  ASP241  GLY242  GLY258  GLU274  ASP323  ARG326  TRP363  ARG373  GLY375  SER409  GLU427  TRP455
3H72-->  ARG347  TRP373  ASP417  ASP423  GLY441  TYR540  LEU566  GLY567  THR572  GLY583  TYR590  THR591  GLY613  GLU647  ASP684  ARG687  TYR710  ARG721  GLY724  ASN753  GLU768  TRP786
CSNR#--> CNSR118 CNSR178 CNSR179 CNSR185 CNSR196 CNSR207 CNSR223 CNSR224 CNSR228 CNSR235 CNSR243 CNSR244 CNSR260 CNSR276 CNSR324 CNSR327 CNSR361 CNSR406 CNSR373 CNSR407 CNSR425 CNSR458
CNSR118   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR178   2.581   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR179   0.410   3.789   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR185   0.485   3.293   0.530   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR196   0.601   0.686   1.357   1.027   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR207   1.174   1.800   0.525   0.276   1.410   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR223   0.908   3.665   0.729   0.600   1.639   0.473   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR224   0.287   3.954   0.112   0.189   1.234   0.140   0.069   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR228   1.331   5.232   1.125   0.730   2.121   0.230   1.348   1.091   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR235   0.198   3.937   0.407   0.950   1.559   2.002   0.580   0.304   0.916   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR243   1.369   2.445   1.559   1.629   1.488   1.598   1.389   1.787   0.867   1.615   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR244   0.434   2.990   0.148   0.892   1.341   0.450   0.328   0.558   0.219   1.294   0.452   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR260   0.648   3.367   0.676   0.940   2.324   0.988   0.128   0.484   0.587   0.367   0.865   0.366   0.000  ------  ------  ------  ------  ------  ------  ------  ------  ------
CNSR276   0.253   4.215   0.377   0.296   1.260   0.398   0.506   0.212   0.331   0.730   0.856   0.878   0.369   0.000  ------  ------  ------  ------  ------  ------  ------  ------
CNSR324   3.899   5.631   1.155   1.464   1.529   0.507   0.604   0.180   0.311   3.839   0.872   1.287   1.697   0.924   0.000  ------  ------  ------  ------  ------  ------  ------
```

CNSR327	1.626	3.984	0.224	2.437	1.108	1.489	0.443	0.646	1.667	4.032	0.853	1.014	1.813	0.882	1.182	0.000	------	------	------	------	------	------
CNSR361	1.689	5.235	1.174	2.340	1.664	1.897	1.163	0.988	1.909	3.186	0.847	0.863	1.690	1.376	1.924	0.664	0.000	------	------	------	------	------
CNSR406	2.016	2.881	2.706	4.175	1.658	3.482	2.589	2.825	4.132	3.838	2.114	0.768	3.993	2.552	6.523	3.326	4.067	0.000	------	------	------	------
CNSR373	1.079	3.556	0.820	1.119	1.163	1.463	1.258	0.593	0.834	1.379	0.685	1.294	0.520	0.308	6.475	2.795	1.638	2.062	0.000	------	------	------
CNSR407	0.996	3.871	1.271	0.969	1.045	1.803	0.314	0.177	0.556	0.970	1.204	0.480	0.656	0.618	4.304	1.281	2.274	4.007	0.541	0.000	------	------
CNSR425	0.902	3.463	0.875	0.189	0.772	1.206	1.152	0.664	0.854	0.444	1.212	1.024	0.695	0.469	6.065	2.731	2.566	1.299	0.881	0.349	0.000	------
CNSR458	3.766	5.358	2.040	2.053	3.172	1.541	2.494	1.613	1.667	2.802	1.083	1.810	1.265	1.257	3.204	0.462	3.052	4.388	1.706	1.600	4.617	0.000

```
>> N10P and SEI Example of Common Localized Relative Spatial Occupancy <<

Original Data from 4VFK.pdb (23 points):

ATOM    199  N   TRP A 106     -15.310 -52.464  30.222  1.00 23.48           N
ATOM    200  CA  TRP A 106     -14.530 -52.396  31.461  1.00 24.83           C
ATOM    201  C   TRP A 106     -15.260 -51.623  32.560  1.00 26.50           C
ATOM    202  O   TRP A 106     -16.159 -50.841  32.281  1.00 23.87           O
ATOM    203  CB  TRP A 106     -13.156 -51.768  31.225  1.00 21.96           C
ATOM    204  CG  TRP A 106     -12.403 -52.299  30.014  1.00 22.29           C
ATOM    205  CD1 TRP A 106     -11.745 -51.552  29.070  1.00 21.06           C
ATOM    206  CD2 TRP A 106     -12.243 -53.672  29.618  1.00 21.43           C
ATOM    207  NE1 TRP A 106     -11.190 -52.368  28.127  1.00 21.34           N
ATOM    208  CE2 TRP A 106     -11.475 -53.674  28.432  1.00 19.57           C
ATOM    209  CE3 TRP A 106     -12.686 -54.904  30.145  1.00 22.36           C
ATOM    210  CZ2 TRP A 106     -11.130 -54.840  27.764  1.00 21.36           C
ATOM    211  CZ3 TRP A 106     -12.336 -56.077  29.485  1.00 22.95           C
ATOM    212  CH2 TRP A 106     -11.569 -56.040  28.300  1.00 21.37           C
ATOM    224  N   GLU A 109     -13.800 -53.081  36.977  1.00 26.44           N
ATOM    225  CA  GLU A 109     -12.483 -53.666  36.826  1.00 25.81           C
ATOM    226  C   GLU A 109     -12.097 -53.794  35.365  1.00 24.77           C
ATOM    227  O   GLU A 109     -12.945 -53.748  34.489  1.00 23.84           O
ATOM    228  CB  GLU A 109     -12.417 -55.022  37.537  1.00 27.69           C
ATOM    229  CG  GLU A 109     -12.427 -54.901  39.052  1.00 29.04           C
ATOM    230  CD  GLU A 109     -12.111 -56.216  39.757  1.00 32.96           C
ATOM    231  OE1 GLU A 109     -11.281 -56.190  40.693  1.00 34.42           O
ATOM    232  OE2 GLU A 109     -12.680 -57.268  39.378  1.00 30.39           O

Distance Matrix for above 23 points from 4VFK.pdb:

               N___TRP CA__TRP C___TRP O___TRP CB__TRP CG__TRP CD1_TRP CD2_TRP NE1_TRP CE2_TRP CE3_TRP CZ2_TRP CZ3_TRP CH2_TRP N___GLU CA__GLU C___GLU O___GLU CB__GLU CG__GLU CD__GLU OE1_GLU OE2_GLU
N___TRP_A_106   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CA__TRP_A_106   1.466   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
C___TRP_A_106   2.485   1.529   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
O___TRP_A_106   2.756   2.397   1.224   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CB__TRP_A_106   2.476   1.529   2.496   3.315   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CG__TRP_A_106   2.919   2.574   3.886   4.623   1.522   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CD1_TRP_A_106   3.856   3.766   4.954   5.504   2.585   1.372   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CD2_TRP_A_106   3.351   3.202   4.686   5.517   2.654   1.438   2.246   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
NE1_TRP_A_106   4.623   4.719   6.064   6.654   3.718   2.244   1.365   2.243   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CE2_TRP_A_106   4.402   4.488   5.964   6.692   3.776   2.292   2.232   1.413   1.371   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CE3_TRP_A_106   3.584   3.380   4.819   5.756   3.350   2.624   3.644   1.411   3.570   2.432   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CZ2_TRP_A_106   5.400   5.586   7.100   7.854   5.052   3.625   3.591   2.458   2.499   1.387   2.845   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CZ3_TRP_A_106   4.737   4.719   6.152   7.060   4.719   3.815   4.582   2.410   4.113   2.761   1.391   2.439   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----   -----
CH2_TRP_A_106   5.521   5.660   7.161   7.997   5.415   4.199   4.557   2.793   3.696   2.372   2.438   1.386   1.412   0.000   -----   -----   -----   -----   -----   -----   -----   -----   -----
N___GLU_A_109   6.949   5.606   4.875   5.713   5.935   7.145   8.312   7.545   9.254   8.875   7.158   9.752   8.201   9.435   0.000   -----   -----   -----   -----   -----   -----   -----   -----
CA__GLU_A_109   7.284   5.881   5.485   6.492   5.952   6.948   8.073   7.212   8.890   8.454   6.798   9.237   7.728   8.897   1.449   0.000   -----   -----   -----   -----   -----   -----   -----
C___GLU_A_109   6.208   4.808   4.752   5.893   4.729   5.564   6.692   5.750   7.433   6.962   5.369   7.733   6.312   7.432   2.451   1.517   0.000   -----   -----   -----   -----   -----   -----
O___GLU_A_109   5.045   3.675   3.687   4.864   3.823   4.735   5.969   4.922   6.742   6.233   4.503   7.051   5.553   6.742   2.714   2.384   1.220   0.000   -----   -----   -----   -----   -----
CB__GLU_A_109   8.272   6.948   6.664   7.688   7.140   8.001   9.175   8.035   9.854   9.252   7.398   9.859   8.121   9.332   2.448   1.533   2.516   3.345   0.000   -----   -----   -----   -----
CG__GLU_A_109   9.603   8.266   7.805   8.733   8.462   9.405  10.551   9.515  11.283  10.733   8.911  11.362   9.639  10.846   3.083   2.546   3.864   4.735   1.520   0.000   -----   -----   -----
CD__GLU_A_109  10.734   9.448   9.100  10.058   9.678  10.505  11.666  10.454  12.285  11.624   9.718  12.111  10.275  11.471   4.518   3.903   5.016   5.877   2.539   1.525   0.000   -----   -----
OE1_GLU_A_109  11.822  10.497  10.141  11.098  10.617  11.421  12.523  11.398  13.135  12.518  10.719  13.000  11.258  12.397   5.461   4.772   5.899   6.872   3.552   2.381   1.251   0.000   -----
OE2_GLU_A_109  10.669   9.478   9.220  10.187   9.846  10.604  11.824  10.411  12.362  11.584   9.531  11.966   9.970  11.201   4.955   4.419   5.340   6.030   2.916   2.403   1.255   2.202   0.000

Original Data from 2G9H.pdb (23 points):

ATOM   3501  N   TRP D  51      14.410  15.929  30.312  1.00 45.46           N
ATOM   3502  CA  TRP D  51      14.451  17.223  30.991  1.00 45.60           C
ATOM   3503  C   TRP D  51      15.869  17.755  31.250  1.00 46.25           C
ATOM   3504  O   TRP D  51      16.095  18.960  31.127  1.00 46.36           O
ATOM   3505  CB  TRP D  51      13.611  17.179  32.277  1.00 44.91           C
ATOM   3506  CG  TRP D  51      13.495  18.478  33.052  1.00 44.10           C
ATOM   3507  CD1 TRP D  51      13.674  18.637  34.400  1.00 43.25           C
ATOM   3508  CD2 TRP D  51      13.148  19.784  32.540  1.00 43.72           C
ATOM   3509  NE1 TRP D  51      13.459  19.951  34.760  1.00 43.83           N
ATOM   3510  CE2 TRP D  51      13.137  20.676  33.642  1.00 43.51           C
ATOM   3511  CE3 TRP D  51      12.846  20.287  31.261  1.00 43.40           C
ATOM   3512  CZ2 TRP D  51      12.848  22.039  33.504  1.00 43.51           C
ATOM   3513  CZ3 TRP D  51      12.546  21.649  31.131  1.00 42.91           C
ATOM   3514  CH2 TRP D  51      12.552  22.502  32.247  1.00 43.17           C
ATOM   3523  N   GLU D  53      18.525  17.482  29.399  1.00 47.78           N
ATOM   3524  CA  GLU D  53      18.919  18.108  28.130  1.00 47.66           C
ATOM   3525  C   GLU D  53      18.133  19.397  27.880  1.00 46.50           C
ATOM   3526  O   GLU D  53      18.697  20.487  27.833  1.00 46.76           O
ATOM   3527  CB  GLU D  53      18.703  17.115  26.964  1.00 48.12           C
```

```
ATOM   3528  CG  GLU D  53      19.142  17.604  25.567  1.00 50.40           C
ATOM   3529  CD  GLU D  53      20.647  17.386  25.283  1.00 53.26           C
ATOM   3530  OE1 GLU D  53      21.426  18.376  25.327  1.00 53.29           O
ATOM   3531  OE2 GLU D  53      21.050  16.226  25.022  1.00 51.95           O
```

Distance Matrix for above 23 points from 2G9H.pdb:

	N___TRP	CA__TRP	C___TRP	O___TRP	CB__TRP	CG__TRP	CD1_TRP	CD2_TRP	NE1_TRP	CE2_TRP	CE3_TRP	CZ2_TRP	CZ3_TRP	CH2_TRP	N___GLU	CA__GLU	C___GLU	O___GLU	CB__GLU	CG__GLU	CD__GLU	OE1_GLU	OE2_GLU
N___TRP_D__51	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CA__TRP_D__51	1.462	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
C___TRP_D__51	2.518	1.536	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
O___TRP_D__51	3.562	2.395	1.232	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CB__TRP_D__51	2.462	1.537	2.547	3.266	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CG__TRP_D__51	3.853	2.596	3.067	3.271	1.517	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CD1_TRP_D__51	4.958	3.772	3.939	4.084	2.576	1.369	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CD2_TRP_D__51	4.628	3.264	3.631	3.371	2.659	1.445	2.248	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
NE1_TRP_D__51	6.072	4.757	4.791	4.597	3.725	2.256	1.379	2.248	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CE2_TRP_D__51	5.937	4.547	4.660	4.245	3.784	2.304	2.241	1.418	1.371	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CE3_TRP_D__51	4.726	3.469	3.943	3.512	3.358	2.627	3.642	1.407	3.568	2.430	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CZ2_TRP_D__51	7.068	5.664	5.706	5.067	5.070	3.647	3.614	2.471	2.512	1.400	2.846	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CZ3_TRP_D__51	6.072	4.821	5.121	4.453	4.736	3.827	4.586	2.414	4.109	2.757	1.401	2.424	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CH2_TRP_D__51	7.099	5.749	5.876	5.134	5.427	4.211	4.564	2.798	3.694	2.371	2.442	1.372	1.405	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----
N___GLU_D__53	4.492	4.382	3.249	3.328	5.703	6.296	7.062	6.639	7.778	7.565	6.602	8.357	7.491	8.306	0.000	-----	-----	-----	-----	-----	-----	-----	-----
CA__GLU_D__53	5.463	5.379	4.377	4.205	6.800	7.334	8.192	7.454	8.784	8.391	7.172	9.011	7.884	8.763	1.469	0.000	-----	-----	-----	-----	-----	-----	-----
C___GLU_D__53	5.639	5.288	4.379	3.858	6.686	7.008	7.935	6.835	8.336	7.733	6.338	8.157	6.845	7.737	2.476	1.530	0.000	-----	-----	-----	-----	-----	-----
O___GLU_D__53	6.730	6.217	5.209	4.467	7.521	7.638	8.472	7.310	8.701	8.043	6.784	8.293	7.075	7.830	3.393	2.408	1.228	0.000	-----	-----	-----	-----	-----
CB__GLU_D__53	5.572	5.857	5.178	5.248	7.359	8.127	9.105	8.311	9.814	9.395	7.927	10.065	8.708	9.735	2.469	1.547	2.524	3.482	0.000	-----	-----	-----	-----
CG__GLU_D__53	6.907	7.181	6.560	6.484	8.706	9.417	10.440	9.450	11.060	10.522	8.903	11.058	9.530	10.585	3.883	2.622	3.096	3.694	1.544	0.000	-----	-----	-----
CD__GLU_D__53	8.143	8.426	7.653	7.573	9.923	10.616	11.546	10.707	12.168	11.709	10.247	12.250	10.863	11.841	4.632	3.408	4.136	4.463	2.584	1.547	0.000	-----	-----
OE1_GLU_D__53	8.948	9.059	8.145	7.899	10.527	11.072	11.937	11.070	12.447	11.964	10.606	12.404	11.102	11.986	5.079	3.770	4.290	4.264	3.418	2.423	1.261	0.000	-----
OE2_GLU_D__53	8.495	8.954	8.244	8.325	10.435	11.253	12.172	11.473	12.897	12.519	11.078	13.153	11.792	12.799	5.207	4.212	5.170	5.621	3.173	2.416	1.255	2.204	0.000

Population Standard Deviation of two sets of interatomic distances for 23 atoms from structures 4VFK.pdb (N10P) and 2G9H.pdb (SEI):

	N___TRP	CA__TRP	C___TRP	O___TRP	CB__TRP	CG__TRP	CD1_TRP	CD2_TRP	NE1_TRP	CE2_TRP	CE3_TRP	CZ2_TRP	CZ3_TRP	CH2_TRP	N___GLU	CA__GLU	C___GLU	O___GLU	CB__GLU	CG__GLU	CD__GLU	OE1_GLU	OE2_GLU
N___TRP	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CA__TRP	0.002	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
C___TRP	0.016	0.004	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
O___TRP	0.403	0.001	0.004	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CB__TRP	0.007	0.004	0.026	0.024	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CG__TRP	0.467	0.011	0.409	0.676	0.003	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CD1_TRP	0.551	0.003	0.507	0.710	0.004	0.002	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CD2_TRP	0.639	0.031	0.528	1.073	0.002	0.004	0.001	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
NE1_TRP	0.724	0.019	0.636	1.028	0.004	0.006	0.007	0.003	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CE2_TRP	0.768	0.029	0.652	1.224	0.004	0.006	0.004	0.002	0.000	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CE3_TRP	0.571	0.044	0.438	1.122	0.004	0.001	0.001	0.002	0.001	0.001	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CZ2_TRP	0.834	0.039	0.697	1.393	0.009	0.011	0.011	0.006	0.006	0.006	0.000	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CZ3_TRP	0.667	0.051	0.515	1.303	0.008	0.006	0.002	0.002	0.002	0.002	0.005	0.008	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
CH2_TRP	0.789	0.044	0.642	1.431	0.006	0.006	0.003	0.002	0.001	0.000	0.002	0.007	0.003	0.000	-----	-----	-----	-----	-----	-----	-----	-----	-----
N___GLU	1.228	0.612	0.813	1.193	0.116	0.424	0.625	0.453	0.738	0.655	0.278	0.698	0.355	0.565	0.000	-----	-----	-----	-----	-----	-----	-----	-----
CA__GLU	0.910	0.251	0.554	1.143	0.424	0.193	0.059	0.121	0.053	0.032	0.187	0.113	0.078	0.067	0.010	0.000	-----	-----	-----	-----	-----	-----	-----
C___GLU	0.284	0.240	0.187	1.017	0.978	0.722	0.621	0.542	0.452	0.385	0.485	0.212	0.266	0.152	0.012	0.000	-----	-----	-----	-----	-----	-----	-----
O___GLU	0.843	1.271	0.761	0.199	1.849	1.451	1.251	1.194	0.980	0.905	1.140	0.621	0.761	0.544	0.339	0.012	0.004	0.000	-----	-----	-----	-----	-----
CB__GLU	1.350	0.546	0.743	1.220	0.110	0.063	0.035	0.138	0.020	0.071	0.264	0.103	0.293	0.201	0.010	0.007	0.004	0.069	0.000	-----	-----	-----	-----
CG__GLU	1.348	0.542	0.623	1.125	0.122	0.006	0.056	0.033	0.111	0.106	0.004	0.152	0.054	0.130	0.400	0.038	0.384	0.521	0.012	0.000	-----	-----	-----
CD__GLU	3.705	0.511	0.724	1.242	0.122	0.055	0.060	0.127	0.059	0.042	0.264	0.069	0.294	0.185	0.057	0.248	0.440	0.707	0.022	0.011	0.000	-----	-----
OE1_GLU	3.563	0.719	0.998	1.600	0.045	0.175	0.293	0.164	0.344	0.277	0.056	0.298	0.078	0.205	0.191	0.501	0.804	1.304	0.067	0.021	0.005	0.000	-----
OE2_GLU	3.913	0.262	0.488	0.931	0.295	0.325	0.174	0.531	0.268	0.468	0.773	0.594	0.911	0.799	0.126	0.103	0.085	0.204	0.129	0.006	0.000	0.001	0.000