

Computational Modelling of Protein-Protein Interactions and Peptide Drug Discovery

Akshay Chenna



**Department of Chemical Engineering
Indian Institute of Technology Delhi**

July 2026

©Indian Institute of Technology Delhi (IITD), New Delhi, 2026

Computational Modelling of Protein-Protein Interactions and Peptide Drug Discovery

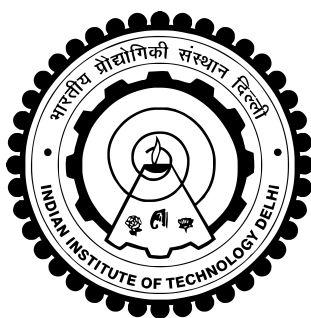
by

Akshay Chenna

submitted in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



Department of Chemical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi, India-11016

July 2026

Certificate

This is to certify that the thesis entitled “**Computational Modelling of Protein-Protein Interactions and Peptide Drug Discovery**”, submitted by **Akshay Chenna** to the Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is a record of the original, bona fide research work carried out by him under my supervision and guidance. The thesis has reached the standards fulfilling the requirements of the regulations related to the award of the degree.

The results contained in this thesis have not been submitted in part or in full to any other University or Institute for the award of any degree or diploma to the best of our knowledge.

Date: Wednesday 8th July, 2026



Prof. Gaurav Goel

Department of Chemical Engineering,
Indian Institute of Technology Delhi
New Delhi, India

Acknowledgments

First and foremost, I deeply and wholeheartedly thank my PhD supervisor, Prof. Gaurav Goel, for his unwavering guidance, selfless direction, immense patience, genuine care, and exceptional support throughout my doctoral journey. His profound technical acumen, practical approach to problem-solving, and insightful feedback constantly sharpened my thinking, and taught me to embrace challenges with ease. His openness to new ideas, constant encouragement to explore them, and remarkable ability to inspire and guide through technical difficulties created a stimulating environment ripe for learning. All of this significantly shaped my professional growth and personal development, extending well beyond the scope of this thesis. I feel truly fortunate to have had his mentorship, which is deeply reflected in the work presented here.

I am also thankful to my research committee members, Prof. A. S. Rathore, Prof. Sudip K. Pattanayek, Prof. Tapan K. Chaudhuri for their feedback and suggestions that has significantly improved the quality of my thesis.

I extend my sincere thanks to the research group leaders with whom I collaborated, namely, Prof. A.S. Rathore, Prof. Archana Chugh, and Prof. V. Haridas and research scholars from their groups, especially, Dr. Rozaleen Dash, Dr. Wajihul Khan, Anandi Chowdhury, Dr. Saurabh Saraswat, and Dr. Hanuman Singh – for their experimental investigations and invaluable support in the validation of our studies.

I gratefully acknowledge the lab and department support staff, Mr. Krishna in particular, and the IIT Delhi supercomputing facility and its staff. Their support and cooperation significantly facilitated my research.

I am thankful to my wonderful lab mates, Rajat, Rit Pratik, Ankit, and Julie, for their camaraderie,

stimulating discussions, and consistent encouragement. The friendly and collaborative atmosphere they provided greatly enhanced my PhD experience.

My heartfelt appreciation goes to my family, especially my parents, for their unwavering support, love, and belief in my abilities given the pandemic. I also extend my gratitude to my friends at IIT Delhi, whose constant encouragement kept me motivated and grounded throughout this journey.

Finally, I express my sincere thanks to the Prime Minister Research Fellowship (PMRF) scheme for providing the financial assistance that enabled me to pursue my doctoral studies smoothly.

A handwritten signature in black ink, consisting of the letters 'AR' in a stylized, cursive font, enclosed within a circular loop. A small horizontal line extends from the bottom right of the signature.

Akshay Chenna

July 2026

Abstract

Physiological protein-protein interactions (PPIs) underpin diverse biological functions, encompassing signaling, enzymatic activity, and immune responses. The human interactome, comprising over 650,000 PPIs, includes transient and weak interactions as well as stable, obligate complexes. Detailed molecular-level understanding of these interactions is crucial for designing effective modulators but remains challenging due to the limitations of conventional structural biology methods, especially when involving disordered protein regions. This thesis addresses two distinct classes of PPIs through computational approaches, complemented by rigorous experimental investigation done by our collaborators.

For the first objective, this thesis reports development of a structural bioinformatics framework for designing peptide-based therapeutics against conserved interfaces, exemplified by targeting the RNA-dependent RNA polymerase complex in SARS-CoV-2. By identifying interface hotspots using molecular dynamics simulations and *in silico* screening of a large synthetic peptide library, candidate peptides capable of effectively disrupting the critical RdRp interactions were successfully identified and validated. Further, the lead candidates obtained from this study were optimized by a Markov chain Monte Carlo procedure for improved target affinity and specificity. The lead-optimized peptides demonstrated superior *in vitro* activity compared to their predecessors and showed comparable activity to an FDA-approved treatment.

For the second objective, this thesis explores therapeutic strategies for modulating PPIs involving disordered proteins, specifically α -synuclein, an intrinsically disordered protein implicated in neurodegenerative disorders. The conformational and kinetic landscape of the aggregation-prone low molecular weight oligomers of α -synuclein were studied using structural bioinformatics and Markov state modelling. A novel method to identify conformation-specific aggregation-prone hotspots with

a high β -sheet secondary structure was developed to drive the docking of a large monomer ensemble. Key druggable intermediates with long-lived intra- and inter-oligomer contacts were identified for effective therapeutic intervention using Markov state modelling on all-atomistic explicit solvent molecular dynamics data. A graph transformer neural network guided by a variational approach for Markov processes (VAMP) was employed for the kinetic separation of the dimer ensemble into fuzzy macrostates. Additionally, this thesis proposes a protein language model (pLM)-driven methodology for the rapid identification of therapeutic antibodies targeting interactions between α -synuclein's disordered C-terminal peptide and its neuronal receptor, LAG3, crucial for preventing pathological propagation.

Finally, the thesis covers a computational study on conformational transitions of peptides based on the bispidine scaffold – a general class of small synthetic organic molecules with broad applications as molecular motors.

Overall, this thesis provides comprehensive computational strategies and experimentally validated solutions for addressing critical challenges in therapeutic modulation of diverse protein-protein interactions.

सार

फिजियोलॉजिकल प्रोटीन-प्रोटीन इंटरैक्शन (पी. पी. आई.) विभिन्न जैविक कार्यों को रेखांकित करता है, जिसमें सिग्नलिंग, एंजाइमेटिक गतिविधि और प्रतिरक्षा प्रतिक्रियाएं शामिल हैं। 650, 000 से अधिक पीपीआई वाले मानव अंतःक्रियात्मक समूह में क्षणिक और कमजोर अंतःक्रियाओं के साथ-साथ स्थिर, बाध्यकारी परिसर शामिल हैं। इन अंतःक्रियाओं की विस्तृत आणविक-स्तर की समझ प्रभावी मॉड्यूलेटर्स को डिजाइन करने के लिए महत्वपूर्ण है, लेकिन पारंपरिक संरचनात्मक जीव विज्ञान विधियों की सीमाओं के कारण चुनौतीपूर्ण बनी हुई है, विशेष रूप से जब अव्यवस्थित प्रोटीन क्षेत्रों को शामिल किया जाता है। यह थीसिस हमारे सहयोगियों द्वारा की गई कठोर प्रयोगात्मक जांच द्वारा पूरक कम्प्यूटेशनल दृष्टिकोण के माध्यम से पीपीआई के दो अलग-अलग वर्गों को संबोधित करता है।

पहले उद्देश्य के लिए, थीसिस संरक्षित इंटरफेस के खिलाफ पेप्टाइड-आधारित चिकित्सीय डिजाइन के लिए एक संरचनात्मक जैव सूचना विज्ञान ढांचे के विकास की रिपोर्ट करता है, जिसका उदाहरण सार्स-कोव-2 में आरएनए-निर्भर आरएनए पोलीमरेज़ कॉम्प्लेक्स को लक्षित करना है। मॉलिक्यूलर डायनामिक्स सिमुलेशन का उपयोग करके इंटरफेस हॉटस्पॉट की पहचान करके और एक बड़ी सिंथेटिक पेप्टाइड लाइब्रेरी की स्क्रीनिंग करके, महत्वपूर्ण आरडीआरपी इंटरैक्शन को प्रभावी ढंग से बाधित करने में सक्षम कैडिडेट पेप्टाइड्स की सफलतापूर्वक पहचान की गई और उन्हें मान्य किया गया। इसके अलावा, इस अध्ययन से प्राप्त प्रमुख उम्मीदवारों को बेहतर लक्ष्य आत्मीयता और विशिष्टता के लिए मार्कोव श्रृंखला मॉटे कार्लो प्रक्रिया द्वारा अनुकूलित किया गया था। लीड-ऑप्टिमाइज़्ड पेप्टाइड्स ने अपने पूर्ववर्तियों की तुलना में बेहतर टेक्सटिट गतिविधि का प्रदर्शन किया और एफडीए-अनुमोदित उपचार के लिए तुलनीय गतिविधि दिखाई।

दूसरे उद्देश्य के लिए, थीसिस अव्यवस्थित प्रोटीन, विशेष रूप से α -सिन्यूक्लिन, न्यूरोडीजेनेरेटिव विकारों में शामिल एक आंतरिक रूप से अव्यवस्थित प्रोटीन से जुड़े पीपीआई को संशोधित करने के लिए चिकित्सीय रणनीतियों की खोज करता है। α -सिन्यूक्लिन के एकत्रीकरण-प्रवण कम आणविक भार वाले ऑलिगोमर्स के संरचनात्मक और गतिज परिदृश्य का अध्ययन संरचनात्मक जैव सूचना विज्ञान और मार्कोव अवस्था मॉडलिंग का उपयोग करके किया गया था। एक बड़े मोनोमर समूह के डॉकिंग को चलाने के लिए एक उच्च β -पत्रक माध्यमिक संरचना के साथ संरचना-विशिष्ट एकत्रीकरण-प्रवण हॉटस्पॉट की पहचान करने के लिए एक नई विधि विकसित की गई थी। ऑल-एटमिस्टिक एक्सप्लिसिट सॉल्वेंट मॉलिक्यूलर डायनामिक्स डेटा पर मार्कोव स्टेट मॉडलिंग का उपयोग करके प्रभावी चिकित्सीय हस्तक्षेप के लिए लंबे समय तक रहने वाले इंट्रा-और इंटर-ऑलिगोमर संपर्कों के

साथ प्रमुख ड्रगबल इंटरमीडिएट्स की पहचान की गई थी। मार्कोव प्रक्रियाओं (वी. ए. एम. पी.) के लिए एक परिवर्तनशील दृष्टिकोण द्वारा निर्देशित एक ग्राफ ट्रांसफॉर्मर तंत्रिका नेटवर्क को अस्पष्ट मैक्रोस्टेट में डाइमर समूह के गतिज पृथक्करण के लिए नियोजित किया गया था। इसके अतिरिक्त, थीसिस α -सिन्यूक्लिन के अव्यवस्थित सी-टर्मिनल पेप्टाइड और इसके न्यूरोनल रिसेप्टर, एलएजी3 के बीच बातचीत को लक्षित करने वाले चिकित्सीय एंटीबॉडी की तेजी से पहचान के लिए एक प्रोटीन भाषा मॉडल (पीएलएम)-संचालित पद्धति का प्रस्ताव करता है, जो पैथोलॉजिकल प्रसार को रोकने के लिए महत्वपूर्ण है।

अंत में, थीसिस में बिस्फिडाइन स्कैफोल्ड के आधार पर पेप्टाइड्स के संरचनात्मक संक्रमण पर एक कम्प्यूटेशनल अध्ययन शामिल है, जो छोटे सिंथेटिक कार्बनिक अणुओं का एक सामान्य वर्ग है जिसका आणविक मोटर के रूप में व्यापक अनुप्रयोग है।

कुल मिलाकर, यह थीसिस विविध प्रोटीन-प्रोटीन अंतःक्रियाओं के चिकित्सीय मॉड्यूलेशन में महत्वपूर्ण चुनौतियों का समाधान करने के लिए व्यापक कम्प्यूटेशनल रणनीतियाँ और प्रयोगात्मक रूप से मान्य समाधान प्रदान करता है।

Contents

Certificate	ii
Acknowledgements	ii
Abstract	v
Contents	ix
1 Introduction	2
2 Literature Review	7
2.1 Targeting intrinsically disordered protein-protein interactions	7
2.1.1 Pathogenesis of neurodegeneration by α -synuclein	7
2.1.2 Development of drugs to target oligomeric assemblies implicated in neurodegeneration	9
2.2 Targeting protein-protein interactions involving folded proteins	11
2.2.1 Design of stable peptides targeting intracellular Protein-Protein Interactions	11
2.2.2 Status of development of therapeutic strategies against SARS CoV2	13
2.3 Conformational study of molecular rotors: Isomerization of Bispidine scaffolds	14
3 An Efficient Computational Protocol for Template-Based Design of Peptides that Inhibit Interactions Involving SARS-CoV-2 Proteins	16
3.1 Summary	16
3.2 Introduction	17
3.3 Results	19

3.3.1	Determination of binding hotspot residues for the nsp12-nsp7 complex	22
3.3.2	Peptide library design and lead candidate(s) identification	22
3.3.3	Structure and interaction map for peptide-nsp7 complexes	23
3.3.4	Lead candidate evaluation	26
3.4	Discussion	28
3.5	Conclusion	32
3.6	Methods and materials	33
3.6.1	Computational design	33
3.6.2	Experimental methods	39
4	Intracellular inhibition of SARS-CoV2 RNA-dependent RNA polymerase com- plexation by lead optimization of peptide binders	41
4.1	Summary	41
4.2	Introduction	42
4.3	Results	43
4.3.1	Lead optimization of peptides and evaluation of <i>in silico</i> efficacy	43
4.3.2	<i>In vitro</i> efficacy of optimized peptide candidates	46
4.4	Discussion	47
4.5	Methods	51
4.5.1	Computational protocol for the optimization of the affinity of known binders	51
4.5.2	<i>In silico</i> evaluation of the designed peptides and calculation of interface metrics	51
4.5.3	Cytotoxicity assay	53
4.5.4	Antiviral screening assay	54
4.5.5	Cellular Uptake assay	54
5	Protein Language Model Based Structure-guided Antibody Screening for Dis- ordered Protein Targets	55
5.1	Summary	55
5.2	Introduction	56
5.3	Results	57

5.3.1	Selection of structurally diverse and sequentially distinct Fv antibodies	57
5.3.2	Generation of alternate conformations of the Fv antibodies	58
5.3.3	Resolving the bound conformation of the disordered C-terminus of α S	59
5.3.4	Generation of synthetic training data of Fv- α S bound poses	60
5.3.5	Evaluating model performance at predicting the binding affinities from Fv sequence	62
5.4	Discussion	63
5.5	Methods	67
5.5.1	Selection of sequentially and structurally diverse variable (Fv) regions from the observed antibody space (OAS) [148]	67
5.5.2	Generation of an ensemble of solution-state conformations via explicit solvent MD at elevated temperatures	68
5.5.3	Docking of C-terminus peptide of α S with the Domain 1 of LAG3 (L3D1) using experimentally derived contact restraints to obtain α S conformations as natively bound to L3D1	69
5.5.4	Generation of Fv- α S docked complexes and Fv paratope re-design via affinity maturation	69
5.5.5	Training neural networks for prediction of binding affinities	71
6	Markov State Model and Structural Bioinformatics aided discovery of Metastable α-synuclein dimers	73
6.1	Summary	73
6.2	Introduction	74
6.3	Results and Discussion	75
6.3.1	Identification of aggregation-prone hotspots on the monomer of α -Synuclein .	75
6.3.2	Generation of an ensemble of dimers of α -Synuclein using information driven docking	78
6.3.3	Conformational sampling of the α -Synuclein dimeric landscape and Markov state modelling	79
6.4	Conclusion	83

6.4.1	Identification of aggregation hotspots	84
6.4.2	Data driven docking	85
6.4.3	Molecular Dynamics simulations	86
6.4.4	Markov State Modelling	87
7	Conformational study of molecular rotors: Isomerization of Bispidine scaffolds	89
7.1	Summary	89
7.2	Introduction	90
7.3	Results and Discussion	92
7.4	Conclusion	101
8	Scope for Future Work	103
8.1	Preventing cell-to-cell transmission of disordered proto-fibrils of α -Synuclein	103
8.2	Peptide drug design for SARS-CoV-2	106
8.3	Therapeutic strategies for Parkinson's Disease	106
A	Supplementary Information: An Efficient Computational Protocol for Template-Based Design of Peptides that Inhibit Interactions Involving SARS-CoV-2 Proteins	108
A.1	RdRp clusters from a molecular dynamics trajectory	108
A.2	Hot-segments on the binding interface of nsp12	109
A.3	Contact maps	109
A.4	Peptide library	110
A.5	Comparison of peptide structures with conformations of the origin domain	110
A.6	Rigid body docking	110
A.7	MMPBSA-WSAS parameters	111
A.8	Molecular dynamics simulation	111
A.9	Estimating IC ₅₀ from ELISA data	112
A.10	Estimation of inhibition constant (K_i)	112
A.11	Cell penetration prediction	113
A.12	Cargo delivery assay	114

A.13 Scripts for peptide design and analysis	115
B Supplementary Information: Protein Language Model Based Structure-guided Antibody Screening for Disordered Protein Targets	125
B.1 Code Execution, packages and dependencies	125
B.2 Datasets	126
C Supplementary Information: Markov State Model and Structural Bioinformatics aided discovery of Metastable α-synuclein dimers	134
C.1 Code Execution, packages and dependencies	134
C.2 Datasets	135
D Supplementary Information: Conformational study of molecular rotors: Isomer- ization of Bispidine scaffolds	147
D.1 Generation of starting structure of molecules for MD simulations	147
D.2 Simulation details	148
D.3 Analysis and comparison of properties from simulations	151
D.4 Computational Modelling	165
D.4.1 ω torsional angle scan using DFT	165
D.4.2 Weighted Ensemble simulations	165
D.5 Well-Tempered Metadynamics	168
Bibliography	179
E List of Publications	212
E.1 Journals	212
E.2 Conference presentations	212
F Bio Data	214

List of Figures

- 3.1 **Generation of peptide library from holo-RdRp subunit, viz. nsp12-nsp7, interactions.** **A.** Crystal structure of the tetrameric holo-RdRp (PDB ID: 6M71) [36] in surf (transparent) and ribbon view. Individual proteins, viz. nsp7 (blue), nsp8 (green), and nsp12 (grey), are labelled. The fingertip loop motif of nsp12 in contact with nsp7 is highlighted (orange). **B** nsp12 residues involved in dominant (hotspot) interactions with nsp7 are shown in licorice view. The residue color denotes its contribution to the binding hotspot area, ($\langle \Delta A_{i,1} \rangle$, eq 3.3), in the central structure of the highest population cluster obtained from a 10 μ s MD simulation of holo-RdRp (Figure A.1). **C** Three longest stretches of sequentially contiguous binding interface residues of nsp12 in the nsp12-nsp7 subunit, referred to as (binding interface) segments (Table A.2), are shown in ball-stick (opaque) and surf (translucent) representation: residues 374-390 (segment 1), 399-418 (segment 2), and 520-522 (segment 3). The inhibitory peptide library consists either of intrasegment residues only or intersegment residues connected by a linker amino acid sequence. 20

3.2 Framework for peptide design and screening. **A** The template protein (nsp12) hotspot area ($\langle \Delta A_{i,c} \rangle$) in the nsp12-nsp7 complex captured by a peptide ($A_{S,c}$) made from nsp12 binding interface segment(s), is plotted against designed peptide length (number of amino acid residues). The data is reported corresponding to three largest population clusters (Table A.1) of the dimeric nsp12-nsp7 complex in 10 μ s MD trajectory, with the mean interface area values ($\langle \Delta A_i \rangle$) weighted by respective cluster population fraction P_c , to give the hotspot score of residue i corresponding to a cluster c ($\langle \Delta A_{i,c} \rangle$) for peptides of length 5–20. **B** Nodes in top row represent 13 peptides, S_i , screened on basis of and shape complementarity (Figure A.2B) to the nsp12 origin domain. Nodes in bottom row represent set of five most stable conformations, C_j^i , of each peptide i . Node S_i and node C_j^k are connected if the TM-score for any conformer of peptide i and conformer C_j is above 0.5 (Figure A.3). The minimum number of unique peptides, i.e., a minimum number of top row nodes sufficient to have at least one connection to each of the bottom row nodes, was determined using integer linear programming (nodes and edges shown in blue). The four peptides at the nodes are, **P1**:LKHFFFAQDG, **P2**:HFFFAQDGN, **P3**:LKHFFFAQDGN, and **P4**:LKHFFFAQ. **C** Binding interface similarity measured by the cosine distance ($\cos \theta$) of the rigid-body docked peptides in comparison to the interface of cluster centre of the most populated cluster of nsp12-nsp7 is plotted for peptides **P1** to **P5**. The peptide **P5**:HFFFAQDGAFQ is an intersegment peptide (see text for details). **D** The binding poses with the highest dot products for peptides **P2** to **P5** from the rigid-body docking stage were subjected to flexible refinement followed by molecular dynamics (MD) simulations. 21

- 3.3 Binding interface interaction map for nsp7-nsp12 and nsp7-peptide complexes. A-C** Dominant PPIs of the nsp7-nsp12, nsp7:**P4**, and nsp7:**P5** complexes are shown, respectively. The binding interface hydrophobic contacts and the hydrogen bonds shown here were identified using the PLIP web tool [83] for the central structure of the most populated cluster obtained from the MD trajectory of holo-RdRp, nsp7:**P4**, and nsp7:**P5** complexes, respectively. The residues involved in these interactions are shown in stick: nsp7 (orange), nsp12 (yellow), **P4** (magenta), and **P5** (cyan). Nsp7 backbone is rendered as ribbon in silver (N terminal) to black (C terminal) gradient. **D** The heatmap for fraction of frames, i.e., a fraction of simulation time, in which select nsp7 residues, those involved in dominant binding interface interactions, are in contact with nsp12, **P4**, and **P5** during the MD trajectory of respective complexes. 24
- 3.4 SPR sensorgrams for serial diluted analyte concentrations. A-C** The association and dissociation phases obtained for five injections for each of the three analytes, viz. nsp12, (**R**)₂**P4**, and (**R**)₃**P5** respectively. In each case, nsp7 was immobilized on the sensor surface and the analyte concentration in the flow buffer was increased from 0.19 μM to 3.12 μM 25

- 3.5 Inhibition of nsp7-nsp12 interaction by designed peptides and their cytotoxicity.** **A,B** Inhibition of nsp12-nsp7 complexation in the presence of designed peptides **(R)₂P4** and **(R)₃P5**, respectively, is measured by a competitive ELISA. The absorbance signal at 450 nm (reported as OD) is reported as a function of the designed peptide concentration, which provides a measure of nsp7 that remains bound to immobilized nsp12. Designed peptide:nsp12 molar concentration ratio and the mass ratio (latter in bracket) used in each well is mentioned. *Positive* control refers to a signal with no designed peptide added to the well. Data is reported as mean +SD from three technical replicates. **C** The inhibition of nsp12-nsp7 complexation is reported as a ratio of the measured inhibitory response (R) at a given peptide concentration to the maximum observed response (R_{max}) (calculated using OD data in A, B. See SI Section A.9). Inhibitory response in presence of **(R)₂P4** is fit to Hill's equation using a Hill's coefficient (*n*) of 2 (solid line), yielding an IC₅₀ of 25 μM (SI eq A.1). **D** The cellular viability of the Vero E6 cell lines was performed using an MTT assay in triplicates to measure the cytotoxicity after 48 h up to a maximum peptide concentration of 100 μM. The cells were cultured in DMEM with 10 % FBS medium. *Control* refers to case with no added peptide. 30
- 4.1 Computational optimization of existing binders to nonstructural protein 7 (nsp7) of SARS-CoV-2 RNA-dependent RNA polymerase (RdRp).** **A** *In silico* peptide binder optimization protocol based on a Markov chain Monte Carlo procedure starting with a docked peptide-protein docked pose. The peptide undergoes several iterations of alternating small-scale backbone perturbations and side-chain packing guided by the binding affinity. **B** The docked pose of nsp7 (in green) with the peptide P4 (in gray) was used for *in silico* affinity maturation to give high affinity binders. Peptide P4O1 (highlighted in orange), a P4 optimized peptide is overlaid on the P4-nsp7 complex. A transparent surface representation of the proteins is overlaid on the respective backbone cartoons. The side-chains of the peptides are displayed in a bond-stick representation. 44

- 4.2 ***In silico* evaluation of the Tat-2 optimized peptide binders.** **A** Top five predictions of the AF2 multimer predictions of the TAT-2 conjugated optimized peptide (P4O1) with nsp7. TAT-2 is coloured in silver. **B.** Nsp7 residues in contact with the optimized peptides are compared to the contacts made by nsp12, quantified by $\cos \theta$. AF2 multimer predictions with the highest $\cos \theta$ were simulated in explicit solvent and the $\cos \theta$ is calculated over a 1000 ns trajectory. **C,D** Visual depiction of key nsp7 (chain A) and peptide (chain B) residues at the interface. The thickness of the bands represents the long-lived contacts over the 1 μ s molecular dynamics trajectory for peptides P4O1 (EDHFLQFY) and P4O2 (EDHFIKYY) respectively. 47
- 4.3 **Evaluation of the inhibitory activity of the optimized peptides.** **A** The antiviral activity of the optimized peptides was measured by the % of nucleocapsid-positive cells compared to the original nuclei using confocal microscopy. Staining was performed with Anti-SARS-CoV-2 antibodies and Hoechst. **B** Analysis of the uptake of the FITC labelled optimized peptides by HeLa cells by fluorescence microscopy at two concentrations (5 μ M and 10 μ M) after 30 min. **C.** Cytotoxic effects of the optimized peptides was measured by the viability of the % VeroE6 cells using Hoechst 33342 and Sytox orange dyes 50

- 5.1 Sampling of alternate backbone conformations and comparison of CDR backbone RMSDs of short *apo* trajectories simulated at different temperatures to the microstates derived for G6 Fab from Fernández-Quintero et al. [145].** **A** The heatmap depicts the minimum backbone RMSD of CDR L2 of *apo* G6 (binds to VEGF) simulated at different temperatures for 50 ns to microstates derived from a Markov-state model (MSM) analysis on a 18.9 μ s long aggregate simulation. Plots for the remaining CDRs are shown in Figure B.2 **B** The CDR L2 RMSDs of the cluster centers derived from clustering a 50 ns trajectory using a cutoff of 2.5 on the backbone with respect to the MSM microstates are compared to a 10 ns trajectory but using a stringent cutoff of 1.5. Both trajectories were simulated at 450 K. **C** A solution state ensemble of the *apo* Fv antibody clustered from a 10 ns trajectory simulated at 450 K used for docking to α -Synuclein (α S) and subsequent affinity maturation. 58
- 5.2 Contact similarity of the docked LAG3- α S complexes to NMR determined contacts.** Docking simulations were performed for LAG3 Domain 1 and the C-terminal residues of α S using the NMR derived contacts as restraints from Zhang et. al. [26, 27]. The fraction of contacts common between NMR and the docked poses is plotted for LAG3 and α S. Each point represents a docked pose whose size is indicative of the docking score, larger sizes implying better HADDOCK scores. Darker shades indicate more populated states. 60
- 5.3 Ensembles of α S and Fv region of the antibody.** **A.** A matrix of the backbone RMSDs (in nm) of the C-terminal peptides (residues 118-140) of α S with the highest NMR determined contact similarity to LAG3 and binding affinities. **B** A visual depiction of the starting ensembles of α S obtained after backbone clustering. 61

- 5.4 Paratope redesign of the Fv antibodies in the *holo* state.** **A** The docked poses of the Fv antibody and α S obtained after clustering using a fraction of common contacts (FCC) criteria of 0.6 are used as an input to RosettaAntibodyDesign (RAbD) for improving the binding energy by modification of the paratope residues. The rosetta energies of the complex (G_{complex}) plotted against the binding energies (ΔG) are shown. The starting docked structures are depicted with a $\star$ and the complexes derived from each docked pose are colored accordingly. **B** A depiction of the conformational changes of the CDRs upon paratope redesign. The docked complex is colored in green and the affinity matured complex is colored in blue. The antigen (α S) is shown in lighter shades with side chains. 65
- 5.5 Model architectures and evaluation of the model predictions for the ΔG of Fv- α S binding.** **A** Fixed-length embeddings from the ESM-2 (150 M) language model are used as an input to a feed forward network. This model is abbreviated with *ESM2 MLP*. **B** For training the *ESM2 FT* model, the ESM-2 (150M) architecture is fine-tuned with the sequences of antibodies generated via the RAbD pipeline to generate contextual embeddings. **C** The spearman correlation coefficients (ρ) for the binding affinity predictions on the test sets are shown for the *ESM2 MLP* and the *ESM2 FT* models. **D** The distribution (in blue) of the best 100 ΔG predictions from the *ESM2 FT* model compared against all the scores of all their sister sequences in the training dataset (in orange). 67

5.6 Model architectures and evaluation of the predictions of ΔG of Fv- α S binding for out-of-training-distribution antibody sequence (<20% CDR identity).

A The neural network architecture for the the Model Extension (*ESM2 ME*) network is depicted. The sequence embeddings generated per-residue are re-learned by the transformer encoder architecture before generating a fixed-length embedding to predict the binding energies (ΔG). **B** For the training-test split with low CDR sequence identity (<20%), the spearman ρ is compared for different model embeddings and architectures. **C** The model-wise precision of the top 100 predicted sequences compared against different train-test split ratios based on the % CDR sequence identity evaluated for all the models depicted in Figure 5.5C and 5.6B. . . . 70

6.1 Identification of aggregation-prone hotspots on the monomers of α -synuclein.

A A heatmap showing the contact surface areas (CSA, in $\text{\AA}^2$) between the Voronoi-derived tessellation of solvent-exposed 4-mer β -like residue fragments of the monomer classified by the structural alphabet approach. The fragments were clustered together using a Markov Clustering algorithm to identify continuous patches with a propensity to nucleate self assembly. **B** A histogram of residues belonging to aggregation-prone patches with a solvent exposure greater than $600 \text{\AA}^2$ are identified from an ensemble of 576 monomer conformations of α -synuclein. Residue fragments with a β secondary structure are plotted in blue, while those also with a tendency to form a β secondary structure are depicted in orange. Residues with a nucleating tendency identified from literature are overlapped with gray bands. **C** An illustration of the identified aggregation-prone patches (in red) on the monomeric conformations of α S (in cyan are green). These patches are used to drive the docking of monomers into dimers. 76

6.2	Data driven docking of an ensemble of α synuclein monomers. A The monomers of α S are rigidly docked using the patch information as restraints. We modify the pre-refinement docking score function (HADDOCK) to recapitulate the scores of the explicit solvent refined docked poses to alleviate the need to perform a full scale docking run for every pair of monomers. We obtain a spearman-rank correlation coefficient (ρ) of 0.73 averaged over 60 docked ensembles; shown here are the rescaled scores from one such representative ensemble. B We generate over 50 million docked poses and rescore them based on the modified score function. The distribution of the scores is shown here. We select the top 2500 ranked dimeric docked poses for initiating molecular dynamics simulations. C An visual representation of two top scoring dimers of α S.	78
6.3	Markov State Model analysis of the dimeric landscape of αS. A. A network plot of a six state Markov state model obtained from probabilistic state assignments of a Graph Transformer based VAMPNet model trained on 200 μ s of MD trajectory using a lag time (τ) of 10 ns. The state size is indicative of the stationary populations. The thickness of the inter connecting arrows indicate the rate of transitions (in sec^{-1}) between the states. B High probability (> 0.9) inter-residue contacts (in orange, $\text{C}\alpha$ - $\text{C}\alpha$ distances under 0.8 nm) are identified for State 2 from the Markov model. Residues 1-140 belong to the first chain while residues 141-280 form a part of the second chain. C,D The MSM derived state-wise β secondary structure content and the dimer interface contacts are depicted. The dashed line are the values obtained on simulating a proto-fibrillar assembly of α S (PDB ID: 2N0A).	80
7.1	(a) Possible paths for <i>syn-anti</i> interconversion. (b) Bispidine diamide conformations representation.	91
7.2	The conformations of bispidine diamides, showing the equatorial hydrogen atoms that are proximal to the carbonyl oxygen atoms. The <i>anti</i> (<i>a1</i>) and <i>syn</i> (<i>s1</i>) forms are shown.	92

7.3	Chemical structures of foldamers of bispidine-cored (F1- F4) in which bispidine scaffold is at the middle. Control compound F5 with 1,3 diaminopropane spacer instead of bispidine and F6 with bispidine placed at the terminal.	93
7.4	(a) Solution structure of F3 in chloroform obtained from MD simulation (intramolecular C—H · · · O bond between Leu carbonyl and bispidine equatorial hydrogen is shown by a dashed red line). (b) Density plot of the population (fraction f reported in units of free energy, $-RT \ln f$) of main-chain torsion angles (ϕ , ψ) obtained from MD simulation of a single molecule of F3 (consisting of a bispidine linker) or F7 (consisting of a 1,3 diaminopropane linker) in chloroform. A significant shift to the characteristic β -strand region ($\phi \sim -140^\circ$, $\psi \sim 130^\circ$) is seen for all residues of F3 . (c) Comparison of the distribution of two torsion angles (shown in inset) capturing the linker conformational space: C22-N5-C6-C1 (solid green) and N5-C6-C1-C9 (dashed green) in F3 , C54-N31-C29-C34 (solid brown) and N31-C29-C34-C39 (dashed brown) in F6 . (d) A representative snapshot showing the main chain ψ dihedral of leucine, and the C—H · · · O bond between bispidine equatorial hydrogen (atom 77) and leucine carbonyl oxygen (atom 24) in F3 . (e and f) Overlay of central structures of five most populated clusters of F7 and F3 , respectively (cluster population in Table S9†) obtained from equilibrium portion of MD trajectory using <i>gromos</i> algorithm with a backbone root mean squared deviation (RMSD) cut-off of 0.1 nm.	95
7.5	Chemical structures of bispidine-based diamides D1-D6	97
7.6	Free energy surface for the <i>syn-anti</i> transition from weighted ensemble (WE) simulations. The Leu–bispidine peptide bond (in D3 and D5) torsional angles from each arm, ω_1 (<i>O30-C2-N31-C19</i>) and ω_2 (<i>O29-C3-N23-C1</i>), were chosen as the progress coordinates (nomenclature in Figure D.10). (a and b) Free energy profiles for D3 and D5 in chloroform at 328K after a molecular time of 2.5 ns (total simulation time: 6 μ s for D3 and 5.8 μ s for D5), respectively. The two <i>syn</i> (<i>syn</i> , <i>syn'</i>) and <i>anti</i> (<i>anti</i> , <i>anti'</i>) states are shown in Figure D.12. WE simulation was run in an equilibrium set-up and a steady state is reached after a molecular time of 1 ns.	98

7.7	The <i>syn</i>→<i>anti</i>-transition in D5 in the ω-ψ torsional angle space of the L-Leu arm. (a) The ψ torsion angle distribution of parent trajectories (trajectory up to but excluding the transition) of walkers that successfully transition to the <i>anti</i> -state from the <i>syn</i> state as obtained from the WE set 2. (b) The ψ torsion angle distribution for trajectories that start from $-100^\circ < \psi < 10^\circ$ (blue) and $60^\circ < \psi < 200^\circ$ (red) regions. (c) The free energy surface (FES) in the ω - ψ (of L-Leu) space as obtained from well-tempered metadynamics simulations. The colour scale and contours represent increments of $5k_B T$ and $10 k_B T$, respectively. (d) A schematic of proposed pathway for the <i>syn</i> - β → <i>anti</i> - β isomerization in D5 involving ω and ψ as two reaction coordinates. All simulations were run at 328 K with D5 solvated in chloroform.	99
8.1	Computational Design and <i>in vitro</i> screening of LAG3 binder antagonists. A Visual representation of the disordered nature of the C-terminal peptide of αS (residues 118-140) from the full-length proto-fibrillar assembly (PDB ID: 2N0A). B The disordered C-terminus of α -Synuclein (αS) is docked to its neuronal-cell surface receptor (LAG3) using NMR-derived contact residues as restraints. The fraction of contacts made by the docked poses are compared to the experimental values. The size denotes the binding free energy (larger is better) and the shade represents the cluster population (darker is populated). C Partial diffusion of 30 steps was performed to design the backbone of the inhibitors starting from the top docked poses identified in the previous step. D 49 anti-LAG3 inhibitors passing AF2 derived metrics were validated by molecular dynamics. The ΔG of binding is calculated by the MM/PBSA-IE method on a 100 ns trajectory. The binding affinity of the C-terminus αS (MM/PBSA-IE) with L3D1 is highlighted as red band. E Visualization of one of the best performing designed anti-LAG3 binders (in orange) bound to LAG3 (in gray). F BLI binding curves for a subset of the experimentally validated binders show affinity in the pM-nM range.	105

- A.1 Residue-wise hotspot scores of nsp12.** Hotspot scores ($\langle A_{i,c} \rangle$) of nsp12 residues averaged over the most populated cluster ($c = 1$) of the MD trajectory are shown in the figure. The per-residue hotspot score is calculated by weighting the fractional population of the cluster P_c with the residue area participating in binding with nsp7 (ΔA) as given in eq 3 (main text). The hotspot scores are visualized on the nsp12's interface with nsp7 in Figure 1B (main text). 116
- A.2 Peptide sequence scores and conformational assessment.** **A** Cumulative hotspot scores of intrasegment peptide sequences captured from the origin domain's (nsp12) hotspot ($A_{S,c}$) is plotted for the best 50 peptides sorted descendingly. The hotspot score of a peptide is the sum of the hotspot scores of individual residues that make up a sequence as given in eq 4 (main text). The coloured sequences possess a length of 8-12 amino acids and were shortlisted for further calculations. The conformations of the sequences coloured in green show a similar backbone fold (backbone RMSD < 0.35 nm) as their origin domain (in Figure A.2B). These sequences are renamed as **S1-S13** before integer linear programming (ILP) optimisation. Four intrasegment peptides **P1-P4** obtained from ILP and labelled. **B** Backbone conformational similarity (RMSD) of the five lowest energy conformations of the peptide intrasegment sequences (**I1-I23**) and two intersegment sequences (**M1-M2**) with the origin domain (nsp12). The peptide conformations for each of the sequences coloured in Figure A.2A were generated from PEPFOLD server. These conformations were tested if they folded into the backbone conformation of the origin domain (nsp12). We applied an RMSD check of 0.35 nm to evaluate the backbone similarity. 117

A.3 TM-score relationships between peptide sequences and generated conformations.	
Binary relationships between a peptide sequence and all conformations are shown. TM-scores were computed between all five conformations of peptides S1-S13 (Figure A.2). A sequence to conformation relationship is encoded with 1 if any of the five conformations generated from the sequence have a TM-score greater than 0.5 with a given conformation. This matrix was used as an input for integer linear optimisation of the peptide sequence and structural similarity (Figure 2B, Methods ILP (main text))	118
A.4 Summary of peptides investigated for docking and MD simulations.	
Five sequences are shortlisted from a combined set of 50 intra-segment sequences with the highest sequence scores ($A_{S,c}$) and two inter-segment peptides using the bioinformatics protocols developed in this study. They are renamed before docking and experimental validation. Five most stable conformations generated from PEPFOLD are represented.	119
A.5 Comparison of the ensemble averaged contacts of the peptide residues and their corresponding origin domain with nsp7.	
The frequency of the residue-wise contacts (fraction of simulation time) determined from the molecular dynamics trajectory formed by the peptides P4 (A) and P5 (B) with nsp7 is compared to the contacts made by origin domain (nsp12) residues. The residues are arranged from the N- to C- terminal of the designed peptide sequences.	120

A.6	Interface similarity of nsp7 in peptide bound complex compared to its interface in RdRp. The similarity is measured with respect to the 16 cluster centers (Cl 1-16) of nsp12-nsp7 complex derived from clustering the MD trajectory. This similarity is expressed as a cosine distance ($\cos \theta$) between two nsp7 interfaces (eq 8 (main text)). A The heatmap illustrates the cosine distance for rigidly docked peptides computed as a Boltzmann average of the binding energy over all docked poses. B The interface similarity of the best scoring poses from rigidly docked peptides when subjected to local flexible refinement is presented. C Ensemble averaged cosine distance of the peptide-nsp7 complex from nsp12-nsp7 complex is analyzed from a 20 ns simulation trajectory for the highest $\cos \theta$ scored poses from flexible refinement.	121
A.7	Prediction of peptide-protein binding from AlphaFold Multimer. The nsp7-peptide poses with the highest dot products for P4 and P5 are compared to the predictions from ColabFold [101]. The P4 (A , in red) and P5 (B , in blue) peptide poses with the highest dot products are shown bound to nsp7 (in white). The surface shaded in yellow is the interface made by nsp7 with nsp12. The peptides coloured in magenta and teal represent the binding poses predicted from AlphaFold Multimer.	122
A.8	X-gal staining assay. HeLa cells were treated with different cargo to peptide ratios. Tat-treated cells were taken as a positive control. Untreated and β -galactosidase were taken to eliminate false positives.	123
A.9	A flowchart summarising the peptide design protocol developed in this study.	124
B.1	A flowchart summarising the computational pipeline developed in this study.	127
B.2	Comparison of CDR backbone RMSDs to microstates from a Markov-state model. A-F The minimum CDR backbone RMSDs (in) of <i>apo</i> murine VEGF antibody, G6 (PDB ID: 2FJG) simulated at different temperatures is compared to the microstates dervied from a Markov-state model analysis. Subplots A-C depict the RMSDs corresponding to CDR H1-H3 and (D-F) correspond to CDR L1-L3 respectively. A 50ns simulation was performed at each temperature starting from the <i>apo</i> state of the G6 Fv antibody.	128

B.3 Similarly of trajectories generated at physiological temperature (310 K) to those generated at a higher temperature.	
From the antibody set assembled by Guest et al. [152], the heatmaps highlight the fraction of MD trajectory frames generated at 310 K that are structurally similar to at least one frame from trajectories generated at higher temperatures (340 K, 370 K, 450 K). Two structures are similar if the combined CDR RMSDs are under 2. MD trajectories of length 0.5 μ s were generated at each temperature for comparison.	129
B.4 The predicted aligned errors (PAE) of the LAG3-αS C-terminal peptide as obtained from the structure prediction from AlphaFold2 [101].	130
B.5 Visualization of structures used for docking. A The complex of LAG3 Domain 1 (L3D1) with the C-terminus of α S corresponding to the highest fraction of NMR contacts as determined by Zhang et al. [27]. B An ensemble of the C-terminus of α S, residues 118-140 (blue to red) generated via HADDOCK [187].	131
B.6 Model performance evaluation by measuring the correlation between the predicted and true binding affinity values. A The correlation calculated using a feedforward network evaluated on a test set that shares > 80 % CDR sequence identity to the training set. B Correlation evaluation using a feedforward network on a test set that shares < 20 % CDR sequence identity to the training dataset. C Correlation plots using per-residue sequence embeddings generated from ESM2 (150M parameters) fed to a transformer layer on test dataset sharing < 20 % CDR sequence identity to the training dataset. The model architectures are depicted in Figure 5 (main text)	132
B.7 Enhancements in structural and sequence diversity. A Post molecular dynamics, the number of backbones (after clustering) results in an average of 7 distinct conformations per Fv sequence. B Unique backbone folds as clustered by FoldSeek is plotted. MD yields larger number of unique folds. C The distribution of the Fv- α S binding affinities (Δ G) on taking a single Fv sequence (in green) compared to taking multiple sequences (in blue).	133

C.1	Structure Alphabet determination and clustering of spatially close fragments	
	A A schematic of the structural alphabet representation from a 4-mer residue fragment (adapted from [214]). B A voronoi tessellation based contact surface are representation of a conformation of α S calculated using VoroContacts [281]. C The per-residue counts of structure alphabets adopting a β secondary structure (in blue) or additionally possessing a tendency to transition to a β secondary sheet (in orange) calculated over the ensemble of 576 conformations of α S. D The aggregation prone clusters are scored based on the weighted average of the solvent-exposure of the individual residues comprising the cluster.	6.4.1 136
C.2	Docking metrics. A The fraction of residues from the ensemble of conformations of α S that are aggregation prone and thus used for driving docking (<i>active residues</i>) using HADDOCK. B The backbone RMSDs (in Å) of the top 10 structures scored after rigid docking are compared to the top 10 structures obtained after flexible refinement and solvent MD. C The rigidly docked poses are rescored using a custom scoring function and compared to the top 10 post flexible refinement and solvent MD using backbone RMSDs (in Å)	 137
C.3	Comparison and selection of appropriate atomistic force fields for molecular dynamics simulations of intrinsically disordered protein assemblies. A,B,C The radius of gyration and the inter-chain interface contacts are plotted for trajectories simulated for 10 ns with CHARMM-36m (TIP3P), AMBER-99sb disp (TIP4P), and CHARMM-36mW (modified TIP3P) respectively starting from a representative docked pose. D,E A heatmap plotting the populations observed during the trajectory in the dimensions of the dimer radius of gyration and interface contacts for CHARMM-36m (TIP3P) and AMBER-99sb disp (TIP4P).	 138
C.4	Pair-wise interface backbone RMSDs (in Å) for six 50 ns long trajectories simulated with CHARMM-36mW. The y-axis is the <i>reference</i> frame for comparison while the x-axis represents the <i>mobile</i> frame.	 139

C.5	The folding landscape of a representative 1 μ s long MD trajectory of α S dimer is plotted with respect to the strand polarity ($P1$) and its degree of orderness (represented by $P2$). The heatmap depicts the population counts (number of frames) as a function of $P1$ and $P2$	140
C.6	Time-dependent observables from 65 500 ns long trajectories of αS dimers. A The radius of gyration (in nm) of the α S dimers observed from explicit solvent trajectories simulated using CHARMM-36mW forcefield at 310 K under physiological conditions. Similarly B represents the number of α S residues in the dimer adopting a β secondary structure (calculated using dssp assignments) and C plots the inter-chain hydrogen bonds formed between the backbone-only atoms.	141
C.7	A flow chart of the procedure for data preprocessing and constructing deep markov state models (VAMPNets) using the graph transformer architecture and the VAMP loss functions for probabilistic state assignments.	142
C.8	Training and validation of the markov state model. A The training loss using the VAMP score is plotted over 1000 epochs in blue while the orange curve is the validation loss on a held out dataset (10% of the total training dataset size) for a six-state assignment objective and using a lag time τ of 10 ns. B The Chapman-Kolmogorov equality tests plotted over lag times ranging from 10 ns to 100 ns. The predicted values are shown in solid black lines while the model estimated values are shown in broken blue lines. C The inter-state transition probabilities from the markov model are plotted. The y-axis is the source state and the x-axis is target state.	143
C.9	The state averaged observables from the markov state model. A The distribution of the probabilities of the state assignments for the MD frames. B The stationary distribution of the state populations from the markov state model. C,D The state-wise radius of gyration and the inter-chain backbone hydrogen bonds are plotted. The dashed line indicates the observed properties from simulating a fibrillar assembly of α S (PDBID: 2N0A). The properties are averaged using the state assignment probabilities output from the markov state model (equation 6.6).	144

C.10	Contact maps showing the probability of inter-residue contacts averaged for each state using the probabilistic state assignments from the markov state model. Equation 6.6 was used to calculate state averages.	145
C.11	Highly likely inter-residue contacts (probability > 0.9) are plotted for each state.	146
D.1	Convergence of MD trajectories. Averages using 25 ns blocks are plotted for ϕ, ψ dihedrals and number of C-H $\cdots$ O bonds for single molecule MD trajectories of F3 , F6 and F7 . The error bars reflect one standard deviation from the mean.	150
D.2	Hydrogen bond analysis. The presence of C-H $\cdots$ O bonds is depicted between the bisipdine equatorial hydrogens (donor) and the carbonyl oxygen of the adjacent Leu residues (acceptor) in the central structure of the most populated cluster in the anti-form of F3 simulated in CHCl ₃ solvent. A distance cutoff of 2.7 Å and a minimum angle criterion of 90° between the donor and acceptor atoms is considered for the formation of a CH $\cdots$ O bonds.	154
D.3	Hydrogen bond analysis. The presence of C-H $\cdots$ O bonds is depicted between the bisipdine equatorial hydrogens (donor) and the carbonyl oxygen of the adjacent Leu residues (acceptor) in the central structure of the most populated cluster in the anti-form of F3 simulated in CHCl ₃ solvent. A distance cutoff of 2.7 Å and a minimum angle criterion of 90° between the donor and acceptor atoms is considered for the formation of a CH $\cdots$ O bonds.	155
D.4	Hydrogen bond analysis. C-H $\cdots$ O bonds in the syn form of F3 obtained from simulations in CHCl ₃ solvent are shown by dashed lines between the bisipdine equatorial hydrogens (donor) and the carbonyl oxygen of the adjacent Leu (acceptor).	156
D.5	Hydrogen bond analysis The central structure of the most populated cluster of F7 simulated in CHCl ₃ solvent.	157
D.6	NOE correlations in F3 he distance matrix represents the r^{-3} and r^{-6} averaged interproton distances in the upper and lower triangular matrices respectively for the hydrogen atoms labelled in F3	160

D.7	NOE correlations in F6 Hydrogen atoms on bispidine linker and backbone residues are labelled on F6 . The distance matrix represents the r^{-3} and r^{-6} averaged inter-proton distances in the upper and lower triangular matrices respectively.	161
D.8	Cartoon representation of β -strand – β -turn – β -strand (β -hairpin) showing the faces of two isolated β -strands in opposite side (<i>anti</i> form) and in same direction (<i>syn</i> form).162	162
D.9	Potential Energy Surface. Two dimensional potential energy surface of F1 in different strand orientations is plotted relative to the most stable structure (<i>anti</i>). The <i>syn</i> and <i>anti</i> forms of the bispidine linker represent its most stable states. Four lowest energy structures are displayed adjacent to their strand configurations in the potential energy map. The structures show C-H—O bonds (displayed in red) on both sides of the linker. Consequently the adjacent leucine residues adopt a β secondary structure with ϕ greater than 150° (Table D.5).	164
D.10	Structures of D1 , D2 , D3 and D5 respectively. D2 and D3 are X-ray resolved crystal structures while D1 and D5 were modified in-silico. The atom numbering follows the topologies generated from CGENFF. The same index scheme was used for defining the reaction coordinates in DFT, WE and metaD simulations.	169
D.11	A pictorial representation of the energy profile of D1 and D2 based on DFT calculation. (a, b) Potential energy surface of D1 and D2 . The ω dihedral angle of the arms are chosen as the reaction coordinates for studying the isomerization. The contours are plotted for every 12.5 kJ/mol increment in the potential energy. Details of DFT scans and geometry optimization are elaborated earlier.	170
D.12	Structures of D3 and D5 sampled from WE simulations. Figures (a,e) (in green) represent the <i>anti</i> -conformation of D3 and D5 respectively while the structures in (c,g) are the corresponding <i>anti'</i> states. The structures in cyan (b,f) denote the <i>syn</i> state and the yellow colored poses (d,h) belong to the <i>syn'</i> state of corresponding D3 and D5 molecules. The state definitions are mentioned earlier.	171

- D.13 Free energy surface (FES) from WE set-2 simulations. (a) is the FES of **D5** obtained on studying the *syn*→*anti* transition by finer binning along ω_1 and restricting sampling along ω_2 by coarser bins. The same is shown for **D3** in figure (b). (c) depicts the FES of **D3** for the *syn*→*anti*' transition, with finer bins along ω_2 and coarser along ω_1 . The colour represents the free energy (in kJ/mol) of the walkers after 300 iterations of WE (3ns of molecular time). 172
- D.14 (a) The figure illustrates the relative weight of the walkers along ω_2 for a given ω_1 (**D5**). An interval of 10° was used as spacing between the bins for estimation of the relative weights. The curve highlighted in red represents the minimum free energy path (highest relative weight) for the *syn*→*anti* isomerization. (b) The rate of the *syn*→*anti* isomerization directly calculated from successful transition events from WE set 2 simulation of **D5**. The blue dashed line represents the estimate rate from NMR relaxation experiments (data not shown). Rate calculation from WE is given earlier. Transition events only after 1ns of molecular time were considered for calculation of the minimum energy path and rates. 173
- D.15 The figure plots the time evolution of the weights of the sub-ensembles of the states from WE set 2 simulations of **D5**. The green curve denotes the *syn* state in β -region and the blue curve denotes the *syn* state in α -region. The red curve is the population evolution of the anti- β state and the orange curve represents the anti- α state. The average populations over the last 200 iterations are 0.623, 0.016, 0.355, 6.876e-13 respectively and 0.638, 0.355, 1.2e-17 and 7.5e-20 respectively when averaged over the last 50 iterations. 174

D.16 Histograms of all possible heavy-atom torsional angles of **D5** from WE simulations.

The colors denote three sub-ensembles viz, blue: walkers in the *syn* state initialized from the *syn* state (i.e. *syn* states with a history in the *syn* state at iteration #0), red: walkers in the intermediate region (between *syn* and *anti* ($-50^\circ < \omega_1 < 50^\circ$) initialized from the intermediate region and green: *syn* states initialized from the intermediate region. The atom quadruples denoting the 62 dihedrals are provided as an attachment. Dihedral #6 denotes the $\psi_1^{(L)}$ of the L-leu arm of **D5**. The green dihedral shows two peaks, one belonging to the α -helical ($-1.75 < \psi < 0.17$) region and the other belonging to β region ($1.04 < \psi < 3.49$). The sub-ensemble belonging to the α region was observed in successful isomerization events to the *anti*-state. 175

D.17 The projection of the free energy surface from metaD simulation of **D5** for the *syn* state. It is observed that a larger barrier exists for the β region ($1.04 < \psi < 3.49$) to the α region ($-1.75 < \psi < 0.17$) conformation (~ 164 kJ/mol) than the reverse transition ($\alpha \rightarrow \beta$, ~ 122 kJ/mol). 176

D.18 A 2D distribution of count of walkers in the WE ensemble simulation of **D5** projected along the ω_1 and the χ_1 torsion. The ω_1 torsion separates the *syn* and *anti* states and the χ_1 torsion (C15-C19-N31-C2) separates the chair-chair and the chair-boat conformations. (a) and (b) correspond to the data for the α and β sub-ensembles (ψ_1 L-leu), respectively. The continuous distribution was obtained from the discrete counts in 2D bins, with total count equal to 118905 and 126102 in the α and β sub-ensembles, respectively. 177

List of Tables

3.1	Peptide-nsp7 equilibrium dissociation constant, K_D . The two computational estimates correspond to the interface score, I_{sc} , obtained from Rosetta FlexPepDock and the MMPBSA-WSAS binding energy calculated from the equilibrium portion (last 10 ns) of the MD trajectory (see eq 3.10). For the SPR assay, the association (on), k_a , and the dissociation (off), k_d , rate constants were obtained by fitting respective temporal profiles (Figure 3.4) into a 1:1 Langmuir model. χ^2 is the average of the squared differences between the SPR measurement and the fitting curve, reported in RU. U-value (uniqueness) is a measure of the correlation between the estimated rate constants and the analyte binding capacity.	27
4.1	Structural and energetic metrics of top three optimized peptides with the best Rosetta binding affinities obtained after optimization are shown. P4 is the starting peptide docked to nsp7 while peptides P4O1, P4O2, and P4O3 are the final peptides designed and selected based on the application of an optimization algorithm using the Rosetta energy function.	45
4.2	<i>In silico</i> metrics of top optimized designs based on the Rosetta energy function. The optimized peptides are evaluated for binding using the docked poses from Rosetta by molecular dynamics simulations in explicit solvent under physiological conditions. The binding energies calculated using the MM/PBSA approximation and the interface similarity of the peptide-nsp7 complex compared to the native nsp12-nsp7 complex is shown.	46
4.3	Viability and viral inhibitory effects of the TAT-2 conjugated peptides on Vero E6 cells. The peptides were evaluated at a concentration of 70 μ M	47

A.1		109
A.2	The table lists the indices of nsp12 residues from all five <i>hot-segments</i> constructed by identifying hotspot residues from the most populated cluster. A segment terminates if three or more consecutive residues do not contribute to binding with nsp7 ($\Delta A_{i-(i+2)} = 0$). The three longest segments (#1, #2, #3) are depicted in Figure 1C (main text).	109
D.1	$^3J^{\text{HNH}\alpha}$ values for Leu, Val and Ile residues present in F3 , F6 and F7 are shown. Karplus equation was used to calculate the J-coupling between the H of backbone Nitrogen and H of C-alpha carbon of a particular residue from the ϕ -dihedrals using the default parameters of <i>gmx chi</i> : A=6.51, B=-1.76, C=1.6. Values of F3 closely resemble the experimentally determined values listed	151
D.2	The fraction of population of the total trajectory binned in the β region ($\varphi : \{-170^\circ, -40^\circ\}$ and $\psi : \{90^\circ, 180^\circ\}$) of the Ramachandran plot is compared for F3 , F6 and F7 . The ϕ , ψ dihedral angles for each residue were plotted on the Ramachandran plots and the fractions were obtained using the rama function in Bioinformatics Toolbox of MATLAB2020b.	152
D.3	The population of the most populated clusters as obtained from simulations. <i>gmx cluster</i> command line tool was used to cluster the trajectories after periodic boundary corrections. Gromos method was applied to cluster on the basis of backbone atoms of the residues and heavy atoms of the linker. A cutoff of 0.1 nm was used to separate the clusters.	153

D.4	C—H · · · O hydrogen bonds between the linker and the carbonyl oxygen of Leu are compared for the bispidine linker (F3 , F6) and diaminopropane linker (F7). The effective number of C—H · · · O bonds formed per tripeptide arm is listed in the top row. The last row is the trajectory average for the number of intramolecular hydrogen bonds considering N, O atoms as acceptors and NH, OH atoms as donors. The presence of C—H · · · O bonds in F3 and F6 is hypothesized in conferring β strand secondary structure stability. The absence of C—H · · · O bonds as observed in F7 and a “hydrophilic collapse” arising from intramolecular hydrogen bonds in a nonpolar solvent (CHCl ₃) leads to a poor β character (Table D.3).	158
D.5	Properties of F1 determined from DFT based geometry optimization.	165
D.6		165