Supporting Information

Affinity of disordered protein complexes is modulated by entropy–energy reinforcement

Milan Kumar Hazra and Yaakov Levy*

Department of Chemical and Structural Biology
Weizmann Institute of Science
Rehovot 76100, Israel

*Corresponding author: Yaakov Levy, Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel; email: Koby.Levy@weizmann.ac.il; Tel: 972-8-9344587
Methods

Studied systems
To understand how the net charge on a sequence and its charge pattern affect the stability of fuzzy protein complexes, we prepared five net-charge variants of the wild-type H1–ProTα complex by interchanging charged residues between the two oppositely charged proteins involved (Table 1). By swapping charges between H1 and ProTα, we designed sequences that had either greater polyelectrolytic character (i.e., they were more homogenously charged with either positively or negative residues and thus had a higher absolute net charge) or greater polyampholytic character (i.e., they contained a more even balance of positively and negatively charged residues and thus had a smaller absolute net charge). We then prepared 9–12 charge-pattern variants from each of the net-charge variants by scrambling the charged residues within the sequence. We used the patterning parameter (κ) to quantify the difference between various distributions of oppositely charged residues in a linear sequence, where 0≤κ≤1 (1). In this manner, we obtained patterning parameter (κ) values ranging from nearly 0.10 (indicating strong mixing of charged residues within the linear sequence of each protein binding partner) to 0.8 (indicating strong segregation of oppositely-charged residues within the linear sequence of each protein binding partner) (see Figure S1).

Simulation details

Model: To analyze the stability of the disordered H1–ProTα complex and compare it with known ordered complexes, we applied a simple coarse-grained (CG) model for the proteins where each residue is represented as (2-5)

\[
V = \sum_{i<j\leq N} \frac{1}{2} k_b^b (d_{ij} - d_{ij}^0)^2 + \sum_{i<j<k\leq N-1} \frac{1}{2} k_b^\alpha (\Theta_{ijk} - \Theta_{ijk}^0)^2 + \sum_{i<j<k<l\leq N-2} k_b^\nu \{1 - \cos(\phi_{ijkl} - \phi_{ijkl}^0)\}
+ 0.5\{1 - \cos(3(\phi_{ijkl} - \phi_{ijkl}^0))\} + K_{\text{Coulomb}} B(\kappa) \sum_{i<j} \frac{q_i q_j e^{-\kappa r_{ij}}}{\epsilon r_{ij}}
+ \sum_{i,j|\text{native/interface}} 4 \epsilon_c \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 + \sum_{i,j|\text{non-contacts}} 4 \epsilon_c \left( \frac{\sigma}{r_{ij}} \right)^{12}
\]

The first two terms represent bonded interactions and angular interactions with uniform force constants, with the third term representing the simplest form of proper and
improper dihedrals. We used a dihedral force constant of 0.24 kcal/mol rad$^2$, which is low enough to capture the intrinsically disordered behavior of the H1–ProTα system. The next term defines the electrostatic contribution$(6)$, where $q_i$ and $q_j$ represent the charge of the two amino acids involved in the interaction, $r_{ij}$ is the inter-bead distance, $\varepsilon$ is the dielectric constant of the solvent, and $K_{\text{Coulomb}} = 4\pi\varepsilon_0 = 332$ kcal/mol. The term $B(\kappa)$ is a function of salt concentration and the radius ($a$) of ions generated by the dissociation of the salt, and is given by $B(\kappa) = \frac{e^k a}{1 + e^k a}$. According to Debye-Hückel theory, the range of electrostatic interactions of an ion is of the order of $\kappa^{-1}$, which is called the Debye screening length. The Debye screening length is related to ionic strength as follows: $\kappa^2 = \frac{8\pi N_A e^2 \rho_A I}{1000 \varepsilon k_B T}$, where $N_A$ is Avogadro’s number, $e$ is the charge of an electron, $\rho_A$ is the solvent density, $I$ denotes the ionic strength of the medium, $k_B$ is the Boltzmann constant, and $T$ denotes temperature.

In addition to the electrostatics, a short-range Lennard-Jones (LJ) potential was used for the folded domains and for interface contacts in the ordered protein complexes. For the H1–ProTα system, this short-range interaction was applied only to the folded domain of histone and not to the interfaces. For the ordered protein complex formed between deoxyribonuclease 2 and the immunity protein for colicin E2 (DNase2–Im2), a uniform strength of $\varepsilon_C=0.50$ kcal/mol was used for the interfacial pairs, but for folded domain pairs the contact strength was set to $\varepsilon_C=0.70$ kcal/mol. For the ordered barnase–barstar protein complex, a uniform strength of $\varepsilon_C=0.55$ kcal/mol was used for the interfacial pairs but for folded domain pairs the contact strength was set to $\varepsilon_C=0.70$ kcal/mol. The higher contact strength of the folded domain ensured that it always remains folded during equilibration. For ordered protein complexes, the $\varepsilon_C$ of the interfacial contacts is a free parameter that can vary in order to achieve the experimental dissociation constant depending on the geometry of the complex and the exact number of contacts, which is only partially captured by the reduced model. H1-ProTα interactions in the simulation are purely electrostatic and we calibrated the DH salt concentration which resembles the experimental dissociation constant [Figure 1C]. The real salt concentration is actually 1.67 times that of DH salt concentration. The charge of the counter-ions in CG model was calibrated by comparing the ion-protein radial distribution functions obtained from our CG model with all-atom simulation of poly-Asp 12-mer’s radial distribution function in SPC/E water with Amber-99SB-ILDN forcefield [Figure S12]. For the ordered protein complex formed between deoxyribonuclease 2 and the immunity protein for colicin E2 (DNase2–Im2) and barnase–barstar, a uniform strength of $\varepsilon_C=0.50$ kcal/mol and $\varepsilon_C=0.55$ kcal/mol was used respectively.
for the interfacial pairs. $\varepsilon_C$ was calibrated to replicate the experimental dissociation constant as shown in Figure 1C. Alteration of $\varepsilon_C$ will lead to change in dissociation constant of the folded complexes. For folded domain pairs the contact strength was chosen such that the folded domain remains intact always. If folded domain pairs the contact strength is decreased, we will lose the folded structure of the two proteins and essentially that will affect dissociation constant, interaction energy and configurational entropy. In addition to the LJ potential for interfacial pairs and contact pairs in folds, all the non-native beads interact with a purely repulsive potential with residue independent $\sigma=6\text{Å}$ and $\varepsilon=0.017 \text{kcal/mol}$ to avoid any overlap. The results are independent of the choice of $\varepsilon$ but alteration of the contact diameter for repulsion can change the energetics. Although the essential trends of the results will remain same as reported. GROMACS 5.1.5 (7) was used to run Langevin dynamics simulations with a friction coefficient of 0.1 ps$^{-1}$ and a time-step of 10 fs for a total of 2 μs for equilibration of the complexes.

**Steered molecular dynamics, umbrella sampling, and conformational entropy:** The dissociation constants for the binding of the three selected systems (H1–ProTα, barnase–barnstar, and DNase2–Im2) were estimated from potential of mean force (PMF) profiles. The PMF was plotted as free energy (F) as a function of the distance between the centers of mass (COM) of the two interacting proteins ($R_{\text{COM}}$), with this distance used as the reaction coordinate. After the 2 μs equilibration, a pulling force was applied to the COM of ProTα at a pulling rate of 0.01 nm/ps and a pull force constant of 2.4 kcal/mol nm$^2$ to achieve a separation of 30 nm from the histone. Using steered MD to separate the two complexed proteins to a dissociation distance of 30 nm creates a vector between their COMs. The PMF was then estimated using the umbrella sampling technique by spacing 200 harmonic umbrellas at distances of 0–30 nm along the vector. Each umbrella sampling windows was equilibrated for 2 μs. The PMF profile was reconstructed from the equilibrated umbrella sampling windows using the weighted histogram analysis method (WHAM) (8). The convergence of the PMF profiles was confirmed for three selected systems (Figure S1).

The dissociation constant is related to the PMF by the following relation,

$$K_D = \frac{1}{4\pi N_A \int_0^{R_{\text{COM}}^b} e^{\beta E(R_{\text{COM}})} R_{\text{COM}}^2 dR_{\text{COM}}}$$
where $\beta = \frac{1}{k_B T}$ and $R_{COM}^b$ defines the maximum extent of the bound state. For the folded DNase2–Im2 and barnase–barstar protein complexes, steered MD was used to apply a force constant of 70 kcal/mol nm$^2$ with the same pull rate of 0.01 nm/ps to separate the two proteins to a dissociation distance of 10 nm. For the DNase2–Im2 and barnase–barstar complexes, 370 umbrella windows were spaced at a distance of 0–6 nm distance while a pulling force constant of 200 kcal/mol nm$^2$ was applied to separate the COMs of each of the two protein units in the complex.

To determine the conformational entropy, we employed two methods that have been widely used in many studies: quasi-harmonic analysis (9) and Maximum Information Spanning Tree (MIST) (10, 11), which produced very well-matched results. However, the MIST approach was found to be less sensitive to the $\kappa$ parameter, most likely because obtaining the configurational entropy using the MIST approach involves a directly correlated movement of beads of up to a maximum of four atoms (a dihedral angle), however charged patches demonstrate correlated movement over a longer range. The current paper discusses the results in light of the conformational entropy obtained from quasi-harmonic analysis, the details of which can be found in the supporting material.

**Configurational entropy from Quasi-harmonic analysis**

We have used Quasi-harmonic analysis in order to obtain absolute entropy of biomolecules from cartesian coordinate covariance matrix in light of the work by Andricioaei et al. and Karplus et al(9). The quasi-harmonic analysis assumes that the fluctuations in a system can be described by a multi-variate Gaussian distribution with covariance $\sigma$

$$p(X) \propto e^{-X^T \sigma^{-1} X}$$

The quadratic effective potential can be written as

$$V_{eff} = \frac{1}{2} X^T F X$$

Where $(F)_{ij} = k_B T (\sigma^{-1})_{ij}$

The normal modes for the effective QH potential $V_{eff}$ is obtained by solving the secular equation

$$\det(F - \omega^2 M) = 0$$
Where $M$ is the mass matrix.

Incorporating the expression of $F$, one obtains

$$\det \left( M^2 \sigma M^2 - \frac{k_B T}{\omega^2} I \right) = 0$$

Where $\sigma' = M^2 \sigma M^2$ is the mass-weighted covariance matrix obtained from molecular dynamics simulations. $\sigma'$ is diagonalized in order to obtain the QH frequencies $\omega_i$. Such $3N-6$ non-zero frequencies are substituted in the harmonic oscillator entropy,

$$S = k \sum_{i=1}^{3N-6} \frac{\hbar \omega}{k_B T} \ln \left( 1 - e^{-\frac{\hbar \omega}{k_B T}} \right)$$

Entropy obtained is exact in the harmonic oscillator limit.

References

1. Das RK & Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences of the United States of America 110(33):13392-13397.
2. Noel JK, et al. (2016) SMOG 2: A Versatile Software Package for Generating Structure-Based Models. PLoS Comput Biol 12(3):e1004794.
3. Noel JK, Whitford PC, & Onuchic JN (2012) The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. J Phys Chem B 116(29):8692-8702.
4. Whitford PC, et al. (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins 75(2):430-441.
5. Clementi C, Nymeyer H, & Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. Journal of Molecular Biology 298(5):937-953.
6. Givaty O & Levy Y (2009) Protein Sliding along DNA: Dynamics and Structural Characterization. Journal of the molecular biology 385:1087-1097.
7. Pronk S, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845-854.
8. Kumar S, Rosenberg, J.M., Bouzida D., Swendsen R.H., Kollman P.A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of Computational Chemistry 13(8):1011-1021.
9. Andricioaei I & Karplus M (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. Journal of Chemical Physics 115(14):6289-6292.
10. Killian BJ, Yundenfreund Kravitz J, & Gilson MK (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *Journal of Chemical Physics* 127(2).

11. Fleck M, Polyansky AA, & Zagrovic B (2016) PARENT: A Parallel Software Suite for the Calculation of Configurational Entropy in Biomolecular Systems. *Journal of Chemical Theory and Computation* 12(4):2055-2065.
Figure S1: Representative convergence of PMF profiles for (A) wild type histone-proTα (B) Colicine IM2-DNASE2 (C) Barnase-Barstar at 200Mm salt concentration and at 300K temperature. In order to test the convergence, the PMF profiles are calculated in three equal segments of 400ns from the 2 μs trajectory after discarding initial 800ns for equilibration.

Figure S2: Illustration of snapshots of wild-type H1-ProTα with different separations distance between their COM illustrating the long-range attractions between the two binding components.
Figure S3: Configurational entropy obtained from two different approaches (A) Quasi-harmonic analysis, (B) Maximum information spanning tree\(^2\,^3\) (MIST) method for all the variants along charge pattern parameter \(\kappa\). A linear equation is fitted to all the charge pattern variants of each net-charge variant for clarity. Color schemes represent the net-charge variants. MIST is not sensitive to variation of charge-pattern parameter as such probably due to the fact that correlated movement up to dihedral angle is not sufficient to capture the connected movement of long charged patches developing with increase in \(\kappa\).
Figure S4: Dissociation constant ($K_D$) of all the net-charge and charge-pattern variants of wild type H1- proTα as a function of charge-pattern parameter at 66mM salt concentration and temperature 300K. Color schemes represent the net-charge variants. $K_D$ of Charge-pattern variants for each of the net-charge variant is fitted to a linear equation for clarity.
Figure S5: Estimated absolute energies and quasi-harmonic configurational entropies of net-charge and charge-pattern variants as a function of charge pattern order parameter of the variants in bound [(A)-(B)] and unbound states [(C)-(D)]. Color schemes represent net-charge variants. Along the κ coordinate, as the sequences develop long charged patches, the bound state is significantly energetically stabilized for all the net-charge variants equally. Tighter interactions in the bound state shrinks the configurational space hence the entropy decreases. While sequences with relatively higher poly-ampholytic behavior inherits energetic stabilization with an entropy compensation along the κ coordinate in the unbound state whereas stability of highly charged sequences remain unaltered due to repulsion.
Figure S6: (A) ΔE as a function of charge pattern parameter (κ) for various net-charge variants of the histone-ProTα complex. A linear dependence of ΔE is observed on the charge patterning in sequences. As the sequences evolve to have elongated charged patches, the energetic stabilization tends to increase. (B) TΔSQH as a function charge-pattern parameter for all the net-charge variants of wild-type histone-ProTα system. (C) Configuration entropy deference (-TΔSQH) plotted against energy difference (ΔE) between bound and isolated states for all the net-charge and charge-pattern variants of histone-ProTα complex. A linear correlation (solid black line) is evident between TΔSQH and ΔE; illustrating an enthalpy-entropy reinforcement. (D) Energy and configurational entropy contributions from bound and isolated proteins to the stability of the complex for all the variants at salt 66mM and 300K (room temperature).
**Figure S7:** Simulated bound state trajectory projected on the first two principal components for bound state of the complex. Panel (A) to (E) denotes each of the net-charge variants in low (black) intermediate (dark grey) and high $\kappa$ (light grey) regime.
**Figure S8:** Simulated unbound state trajectory projected on the first two principal components for unbound histone. Panels (A) to (E) denotes each of the net-charge variants in low (black) intermediate (dark grey) and high $\kappa$ (light grey) regime.
Figure S9: Simulated unbound state trajectory projected on the first two principal components for unbound proTα. Panels (A) to (E) denotes each of the net-charge variants in low (black) intermediate (dark grey) and high κ (light grey) regime.
**Figure S10:** Dimension of the complexes of H1-ProTa. A) Rg of the 52 simulated variants of the complex shown as a matrix with the 16 variants of H1 shown in the X-axis and the 16 variants of ProTα shown in the Y-axis. B). Sum of radius of gyration (Rg) of the ProTα and H1 variants in the bound state as a function of charge-pattern parameter. Color code represents the net-charge variants. All the κ-variants for a specific net-charge variants are fitted with a straight line for clarity. C). Similar to B but shows the Rg of the complex.
**Figure S11:** Net-charge and charge-pattern variants used in this study.

Variants of wild type H1-ProTα used in the projects

**VARIANT A**

H1  
K=0.14

```
TENSTSAPAA KPKRAKASEK ST DHPKYSDMIVAAIOAEKNR AGSSROSIOK YIKSH YKVGE NADSI QKL SI
KRIVVTGVYK QTKGVGAGSFSRFA KSDEPKKS VAEKETKKIE KKVA TPKKAS KPK EAASKAP TKP KPATPYE
KAKKKLAATP EKAKKPETVK AKPK KASEPK KAKPVMPKAE SSAA KAGEKK
```

K=0.38

```
TENSTSAPAA PDKRRKK SAA ST DHPKYSDMIVAAIOAEKNR AGSSROSIOK YIKSH YKVGENADSOIKL SIK
KRIVVTGVK QTKGVGAGSFSRFA PSVAKKKK KEEFTSVAPK KKKKTPVAPK DKKKKKSPAP TAAAPATPVS
AKKKKLAAATP AKKKK ESTVP AKKKKKKV ASPPDEEEEEE DDAAQKRRKK
```

K=0.617

```
TENSTSAPAA PKKRKK SAA ST DHPKYSDMIVAAIOAEKNR AGSSROSIOK YIKSH YKVGENADSOIKL SIK
KRIVVTGVK QTKGVGAGSFSRFA PSVATSVAPK PIPPKKKKK KKKKKKKK TKPVAASSPAP TAAAPATPVS
AKKKKKKKK ALAATPSTVP AKKKKKKKK EDDDEEEEEE DDAAQVASPP
```

**ProTα**

K=0.11

```
SDAAVDTSSSE IETDLKKEKK EVEVEAENGR DAPENENAEEN ENGEQQEADN
EKVEKEKEKE KEAKKEKKEGD EGKEKGDKAE EAESETGERD EKADTDADV
TKEQKEKT KD
```

K=0.39

```
SKAAVDTSSK IITDLKKEKK KVVEEANGR KAPANGNAEN
EKNGEQEADN KVDEEEEEE EGGK GDDEDE
EATQFTGSAA AEDDEDDDDD KK KKKKRD DEDD
```

K=0.68

```
SAAAVNTSSG KRK KRRKKKK KEEE EEDDEL GAPANGNANS KEEE DENG AQ
NVGGGKEEDE EEEEEEEEE DEE EDEDEDK KK KATGITA AGK DEDDDD DE
DDVQTVTAVT T
```
VARIANT B

H1

K=0.17511

**TE**N**ST**SAPAA **K**P**K**RA**K**AS**K**K **ST** **DHPK**Y**S**DMIV **AAIAQAEKNRA** **GSSRSIQKY** **IKSHYKV**GEN **ADSQIKL** **SIK** **RLVTTGVLKQ** **TKGV**GASGSFRLA **KSDEPK**KS **VA**FE**EKT**KKEI **KKVATPK**KAS **KPKEAASKAP** **TKKP**KATPV **D**KAKKKLAA**T**P **KKKKP**ETVK AKPVKASKPK KAKPVKPKAD SSARAGKKK

K=0.418

**TE**N**ST**SAPAA **PKKRKK**SAA **ST** **DHPK**Y**SD**MI**V**AAIAQAEKNR**AGSSRSIQKY**I**KSHYKV**GEN**ADSQIKL**SIK** **RLVTTGVLKQ** **KOTKGV**GASGSFRLA **PSVA**KKKK **KK**FPTSVAPI KKKKTPVAAAS KKKKKS**P**AP TAAAPATP**V**S KKKKLAA**T**P AKKKKAKSSPPK KKKKDEEEEEDDAAGKRKKK

K=0.60

**TE**N**ST**SAPAA **PKKRKK**SAA **ST** **DHPK**Y**SD**MI**V**AAIAQAEKNR**AGSSRSIQKY**I**KSHYKV**GEN**ADSQIKL**SIK** **RLVTTGVLKQ** **KOTKGV**GASGSFRLA **PSVAT**SA **PIFPPK**KKKK KKKKKKKKKK TPVA**SS**PAP TAAAPATP**V**S KKKKKKKKKK ALAAPSTVP AKKKKKKKKKK RKKKDEEEEEDDAAGVASPP

**ProTα**

K=0.153

SDAAVDT**T**SE **I**ETKDL**KE**KK **EVERVEAE**NGR **DAPENENA**N **ENEGEQEADN** EDVEEE**GE**KEG KEAEKEGKEDG EGEKDGDKAE EAESETGERD EDADTDADVD TKEQKEDTDK

K=0.36

SKAAVDT**T**SE **IT**KD**L**KE**KK **KVVEE**AENG**R **DAPANGNA**N **EN**KNGEQEADN EVDVEEE**EEEE**GG **EKEE**EEEE**GD GEEEDGEDEDE EA**K**SA**T**GKRAEDDEDDD**V**D TKKQKTDEDD

K=0.77

SAAAVNT**SS**G **KRKK**KRKKK **KEEE**EE**DEL** GAPANG**N**S **EEEEDEN**GAQ NVGGGDDEDEDEEEEEEEEE DEEEDEDEDE **K**KKKATGITA AGGDDEDDDE DDVVVTAVTT
VARIANT C
K=0.19 (wild type)

K=0.17

K=0.42

K=0.64

ProTa
K=0.194

K=0.42 (wild-type)

K=0.65

K=0.76

19
VARIANT D
K=0.19
TVNSTSAPA AKPKRAKASK KST DHPKYSMDIVAAIQAEKNRAGSSRSQISOQYIKSHYKVGENADSQI
KLSIKRLVTGVVLTQTKGVSASGSFRLA KSAAPKS VAFKKTKKAI KKVATPKKAS KPKKAASKAP
TK KPKATPVIK KAKKKLAAAP KKAAPKPTVK AKPVKASKPK KAKPKPKAK SSAKRAKAK

K=0.424
TVNSTSAPA PKKRRKSSAA ST DHPKYSMDIVAAIQAEKNRAGSSRSQISOQYIKSHYKVGENADSQI
KLSIKRLVTGVVLTQTKGVSASGSFRLA TAAAPAFS VKKKKKKKK KKKKTPVAAAS PKKKASAAP TP
KKKATPVIK LAATPKKKKK KKKKKKPKTAAP APVPAKKKK KKKKKKSSAKK SSAAGKRRKK

K=0.67
TVNSTSAPA PKKRRKSSAA ST DHPKYSMDIVAAIQAEKNRAGSSRSQISOQYIKSHYKVGENADSQI
KLSIKRLVTGVVLTQTKGVSASGSFRLA TAAAPAFS VKKKKKKKK KKKKKKTPVAAAS PKKKASAAP
TPATPATPVIK LAATPVAAASP VAAAPVAKKK KKKKKKKKK KKKKKKKKK KKKKSSAAG

ProTa
K=0.2015
SDAEVTSSE IETKDLKEKK EDEVEEEENGR DEPENENAEN ENEGEOEADN
EDVEEVEKEEG EEAAEOEEEGD EGEDKGGDDAE EAENSEGGERD EADADVDDVD
TKEQKEDTDD

K=0.403
SDAEDDSSTE ITTKDLKEKK EVVEEAENGR DAPENGNAEN EENEGEOEADN
EVDEEEEEEGG EEEEEEEDG GEEEDGDEDE EAENSEKTRA AEDDDEDDVD
TKQKTDEDD

K=0.770
SAAAVNTSSSG KRKRRKRRKK DEEEEDDEI GNPNGEDEEE EEEDEENGAQ
NVGGGDDEDE EEEEEEEDE DEEEDEDEDE EGASATGITA AEDDDEDDDE
TTTVQTDDEDD
VARIANT E

H1

K=0.21

TVNSTSPPR AKPKRAKASK KST DHPKYSMDIVAAIQAEKNRAGSSRQSIQKYIKSHYKVENADSOI KLSIKRLVTGGVLQTKGVGASGFSRK KKKKKKKK VKFKT KKA KKVATTPKAS KPKKAKSKAP TK

K=0.37

TVNSTSAPP RRKRKKKSAA ST DHPKYSMDIVAAIQAEKNRAGSSRQSIQKYIKSHYKVENADSOI KLSIKRLVTGGVLQTKGVGASGFSRK LA TKKPAPFS VKKKKKKKK KKKTPKKS PKKRASKAP TP KKKATPVA LKATPKKKK KKKKKKPTVA ASPVPAKKK APAFPVKKKK SSAAGKRKKK

K=0.69

TVNSTSAPP RRKRKKKSAA ST DHPKYSMDIVAAIQAEKNRAGSSRQSIQKYIKSHYKVENADSOI KLSIKRLVTGGVLQTKGVGASGFSRLA TFSVPAKK KKKKKKKKK KKKKKKKKK VSIITPASSEP TPATPATPVA LAPTVAASPV PAPAPVKKK KKKKKKKKK KKKKKKKKK KRKKKSAAG

ProTα

K=0.079

SDAEVDSSE IETADLAEEA EDEVEEEEANGA DEPENEENAEN ENEGEEQEADN EDVEEDEEAEG EEAEEEGGDD EGEADGDDEAE EASETGEAD EADTDADVD TAEQAEDTDD

K=0.285

SDAEDDTSSS ITTADLAEEA EVVREEAENGA DAPENGAEN ENEGEEQEADN EVDEEEEDEEG EEEEEEEDGDEDE EAESETGAAA AEDDEDDDDVD TAAQTADDD

K=0.72

SAAAVNTSSG AAAAAAAA DEEEEEDEDEL GNPNNGDEEE EEEEDEENGAQ NVGGGDDEDE EEEEEEEEE DEEEDEDEDE EGAQATGITA AEDDEDDDEDE TTQVQDDEDD
Table S1: Analysis of fuzzy complexes from FuzDB with measured affinity
| Complex                                                                 | K of fuzzy region | FCR of fuzzy region | f+ of fuzzy region | f- of fuzzy region | K_D(M)   |
|------------------------------------------------------------------------|-------------------|---------------------|-------------------|-------------------|----------|
| Splicing factor 1 (SF1) and U2 small nuclear RNA auxiliary factor (U2AF65) | 0.1343            | 0.3148              | 0.1574            | 0.1574            | 11.8E-09 |
| T cell receptor ζ chain and T cell receptor ζ chain                   | 0.1788            | 0.34343             | 0.1818            | 0.1616            | 10E-03   |
| Methyl CpG binding protein 2 (MeCP2) and DNA                         | 0.1571            | 0.3866              | 0.1866            | 0.2               | 0.8E-09  |
| Ets-1 transcription factor-DNA                                        | 0.3010            | 0.2280              | 0.0701            | 0.1578            | 1E-08    |
| GCN4 and Med15                                                        | 0.1602            | 0.2388              | 0.0522            | 0.1865            | 10.1E-06 |
| Vesicle associated membrane protein C (VAPC) and Hepatitis C virus non-structural protein NS5B | 0.2226            | 0.2222              | 0.1388            | 0.0833            | 49.13E-03 |
| c-Myc and Bin1 SH3 domain                                              | 0.2960            | 0.2159              | 0.0681            | 0.1477            | 33E-06   |
| Prothymosin α (ProT α)-                                               | 0.5424            | 0.6923              | 0.0961            | 0.5961            | 2.6E-06  |
| Kelch-like ECH and associated protein 1 (Keap1), Kelch domain | 0.2793 | 0.3604 | 0.2209 | 0.1395 | 2.0E-4 |
|------------------|--------|--------|--------|--------|-------|
| AP180 and Clathrin | -1.0  | 0      | 0      | 0      | 7.3E-9 |
| Nup153 and Importinβ | 0.1580 | 0.1057 | 0.0192 | 0.0865 | 10E-06 |
| Gelatinase-associated lipocalin receptor (LCN2-R, SLC22A17) and Lipocalin (NGAL) | 0.2009 | 0.1515 | 0.0909 | 0.0606 | 100E-09 |
| Kappa-bungarotoxin and Kappa-bungarotoxin | 0.2424 | 0.2027 | 0.1216 | 0.0810 | 0.4E-09 |
Figure S12: Scaling of counter-ion charges for CG model. As CG model lacks explicit solvent screening, counter-ion charges have been calibrated (scaled) in order to reproduce the all-atom protein-counterion radial distribution function. Radial distribution function (A) and cumulative coordination number of counter ion Na (B) around model 12-mer poly-Asp peptide. While black curve shows the results from explicit atomistic simulation with AMBER99SB-ILDN force field and SPC/E water, red, green, blue and magenta denote the radial distribution and cumulative coordination number for different partial charges of Na ion in a CG model. A choice of partial charge 0.55 for Na ions in CG model seems to be a reasonable choice as it reproduces the exact RDF obtained from all atom simulations. The charge of Cl ions has been set as -0.55.