Structural Insight into the Role of Thrombospondin-1 Binding to Calreticulin in Calreticulin-Induced Focal Adhesion Disassembly

Running Title: Structural bases for TSP1-CRT interactions

Qi Yan, [‡] Joanne E. Murphy-Ullrich, [§] Yuhua Song [*, ‡]

[‡] Department of Biomedical Engineering and [§] Department of Pathology, The University of Alabama at Birmingham, Birmingham, Alabama 35294

* To whom correspondence should be addressed:
Department of Biomedical Engineering
The University of Alabama at Birmingham
803 Shelby Interdisciplinary Biomedical Research Building
1825 University Boulevard
Birmingham, AL 35294
Phone: (205) 996-6939 Fax: (205) 975-4919
Email: yhsong@uab.edu
Web: http://www.eng.uab.edu/yhsong

MATERIALS AND METHODS

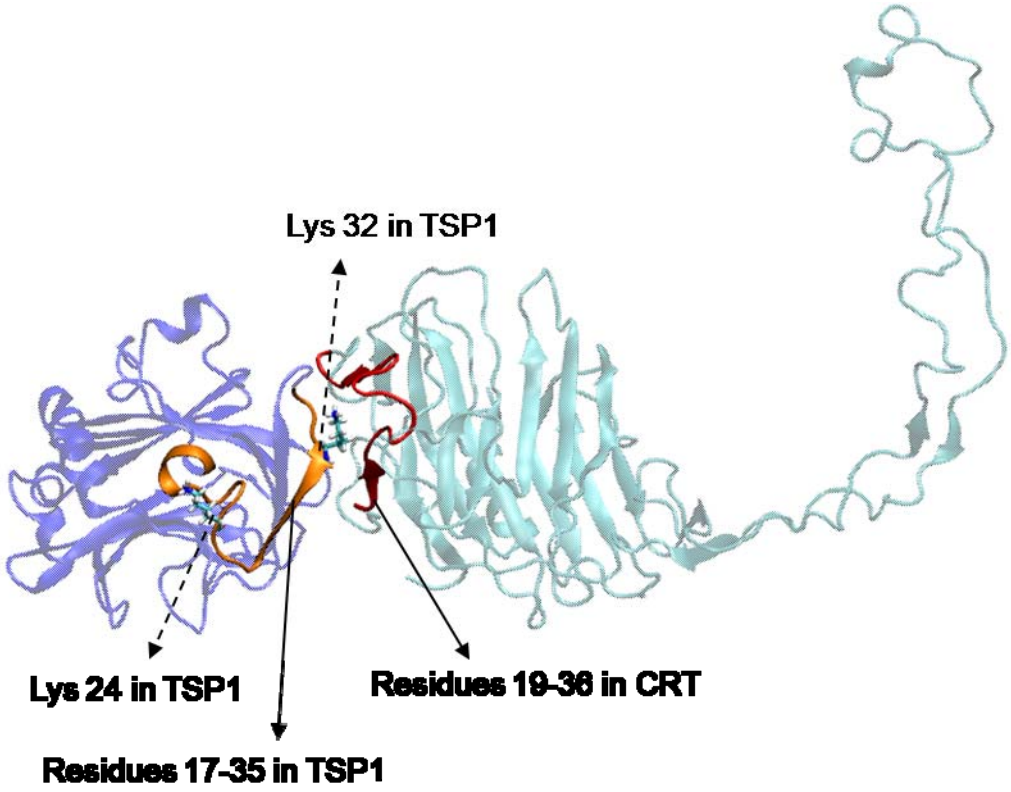


Figure 1S. Binding restrains for constructing TSP1-CRT complex. Blue: TSP1; orange: CRT binding site in TSP1 (residues 17-35 in TSP1); cyan: CRT; red: TSP1 binding site in CRT (residues 19-36 in CRT); residues represented with licorice are the key residues of Lys 32 and Lys 24 in TSP1 critical for TSP1-CRT binding. Image was made with VMD software support.

Complex construction pathway I: ZDOCK+RDOCK

The procedure of using the combined ZDOCK and RDOCK programs to predict TSP1-CRT complex is shown in Fig. 1. Using the structures of TSP1 (1) and CRT (2), and using the known binding site between TSP1 and CRT as restraints (3-5), 2000 possible complexes were generated with the ZDOCK 2.3 docking software with the rotational sampling density of 15°

(6). Complexes were scored based on their geometry, desolvation, and electrostatics (7). The RDOCK program was used to perform the energy minimization for all of the 2000 predicted complexes by using the CHARMM MD simulation package (8). The electrostatic, van der Waals, and contact energies associated with each of the proposed complexes were determined and used for ranking (9, 10). Based on the binding energy of the predicted TSP1-CRT complexes, the experimentally known binding sites between TSP1 and CRT (residues 17-35 in TSP1 and residues 19-36 in CRT), and the residues in TSP1 critical for CRT binding (lysines 24 and 32 in TSP1) as restraints (Fig. 1S), one potential model of TSP-CRT complex predicted from the ZDOCK/RDOCK program was chosen for the further validation.

Complex construction pathway II: ZDOCK+ ZRANK+RosettaDock

The procedure of using the combined ZDOCK, ZRANK and RosettaDock programs to predict TSP1-CRT complex is shown in Fig. 1. ZRANK is a ranking program that adopts a weighted energy function from van der Waals, electrostatics and desolvation energies to quickly and effectively re-rank the complex predictions from ZDOCK without energy minimization (11). Based on the structures of TSP1 and CRT (1, 2) and using the experimentally known binding sites between TSP1 and CRT as restraints (3-5), we used the ZDOCK 2.3 program (6) to generate 54000 possible TSP1-CRT complexes, using the rotational sampling density of 6° for the ZDOCK run. We used ZRANK to rescore the predicted complexes from ZDOCK, and the top 20 predicted TSP1-CRT complexes were chosen for structural refinement using the RosettaDock program. RosettaDock allows for local refinements, including side chain repacking and a Monte Carlo search of the local rigid-body space of the ligand (12). Using RosettaDock for structural refinement, 300 refined models were generated for each of the top 20 predicted TSP1-CRT complexes from ZDOCK/ZRANK. These refined structures were further rescored using

ZRANK, and the top 20 predicted models were selected from each set of the 300 predicted complexes. We used a structural clustering method, based on the root mean square deviation (RMSD) between structures using the molecular modeling tools in the structural biology (MMTSB) tool set (13), to obtain the representative structures of the top 20 models of each set. Using the known binding sites between TSP1 and CRT, one model was chosen from each set of the 20 models. We then performed energy minimization for the 20 predicted TSP1-CRT complexes using the AMBER 9 MD package (14). Further screening of the 20 minimized structures was based on the key residues of TSP1 for TSP1-CRT binding (lysine residues 24 and 32) with the assumption that these two key residues were on or close to the binding interface, and three models were chosen for further validation.

Lysine residues 24 and 32 of TSP1 are critical for TSP1-CRT binding, and residue 32 is required for TSP1-CRT induced focal adhesion disassembly (4, 5). We expected that mutation of Lys residues 24 and 32 of TSP1 to Ala or mutation of residue 32 of TSP1 to Ala would significantly decrease the calculated binding free energy of the predicted TSP1-CRT complex. For the four potential TSP1-CRT complexes predicted from the two protein docking pathways, we generated two different mutant proteins using the mutagenesis function of PyMol (23) for each complex: double mutants replacing residues 24 and 32 of TSP1 with Ala and single mutants replacing residue 32 with Ala. We performed 10ns MD simulations for the four predicted wild type complexes and the eight mutant complexes. Based on the MD simulation trajectories, binding free energies of the wildtype and mutant complexes were calculated to determine the effects of the combined K24A and K32A mutations and the single K32A mutation in TSP1 on TSP1-CRT binding. The calculated change in binding free energy of the TSP1-CRT complex induced by TSP1 mutations was compared to experimental results for the four predicted TSP1-

CRT complexes and mutants, leading to the identification of a functionally validated TSP1-CRT complex.

MM-PBSA method

As described in previous studies (15-19), the MM-PBSA method combines molecular mechanics, Poisson-Boltzmann electrostatics for polar solvation free energy, apolar solvation energy based on solvent-accessible surface area, and normal mode analyses for entropy to calculate the binding free energy for the protein complexes. The MM-PBSA method has been implemented in the AMBER 9.0 MD program. The binding free energy is expressed as

$$\Delta G_{binding} = \langle G_{complex} \rangle - \langle G_{protein} \rangle - \langle G_{ligand} \rangle \quad (1S)$$

Where $\langle G \rangle$ is the average value over the simulation trajectories.

The free energy of each molecule is defined as

$$G_{molecule} = E_{MM} + G_{polar,solvation} + G_{apolar,solvation} - TS \quad (2S)$$

Where E_{MM} is the free energy from the molecular mechanics contribution, $G_{polar,solvation}$ is the polar solvation energy, $G_{apolar,solvation}$ is the nonpolar solvation energy, T is the temperature, and S is entropy of the system. The free energy of the molecule from the contributions of molecular mechanics and entropy is also known as the free energy in the gas phase.

The free energy of the molecule from molecular mechanics contribution E_{MM} is the sum of internal energy E_{int}, which includes the bond, angle, and torsion energies, electrostatic energy E_{elec} and van der Waals energy E_{vdw}, and it is expressed as

$$E_{MM} = E_{\text{int}} + E_{vdw} + E_{elec} \quad (3S)$$

For the calculation of polar solvation energy with the Poisson-Boltzmann equation using the PBSA solver in the AMBER 9 MD program, values of 1 and 78.54 were used for the dielectrics of the solute and solvent, respectively (15). A 150mM ionic strength and a temperature of 300K were used as in MD simulations, and a spacing grid of 0.5 Å was used for the calculations. The calculation of the apolar solvation energy was based on the solvent-accessible surface area using the Molsurf solver in the AMBER 9 MD program. For the MD simulation trajectories after initial equilibration, frames at an interval of 20ps were used for the calculation of all of the components of binding free energy except entropy. Frames at an interval of 100ps were used for the normal mode analysis to calculate the entropy due to the high computational cost for entropy calculations. We calculated the mean and standard deviations of the binding free energy using the trajectories of MD simulations after the system reached initial equilibration using the bootstrap method (20). Significant differences in the TSP1-CRT binding free energy caused by mutations in TSP1 or CRT were assessed using the Student's t-test with 95% confidence.

Protocol for Anisotropic network model (ANM) restrained MD simulation

The protocol for the ANM restrained MD simulation was comprised of two steps (21). The first step was to generate a succession of conformations using the deformations derived from ANM analysis as restraints in MD simulations, succeeded after each mode by a short energy minimization to make the structure settle into a local energy minimum (21). The structure that had a lower energy was used as the starting structure for the next ANM restrained MD simulations. The number of modes in one loop was determined by the eigenvalue distribution and degree of collectivity. Each mode corresponded to a fluctuation between two opposite

directions, two sets of deformations were considered for each mode, and the corresponding conformations R_k^+ and R_k^- were calculated as shown in equation (4S) (21).

$$R_k^{\pm} = R^0 \pm s\lambda_k^{-1/2} u_k^{ANM} \quad (4S)$$

Where R^0 was the initial conformation before the application of restraint, s is a scaling factor to retain the RMSD of the protein after reconfiguration along mode 1 to be close to 1.5 Å and $1.5(\lambda_i/\lambda_1)^{1/2}$ Å for mode i. λ_k was the eigenvalue for mode k and u_k was the eigenvector for each atom in current mode.

After screening the selected modes in the first loop, the second step was to initiate a new outer loop with the updated ANM modes corresponding to the final structure of the first cycle. This procedure was repeated until the RMSD with the starting structure as reference reached 6 Å (chosen due to the flexible characteristics of CRT P-domain).

RESULTS AND DISCUSSION

Table 1S Calculated binding free energy results for four candidate TSP1-CRT complex models and their mutants (all energies are in kcal/mol)

		ΔE^{elec}	ΔE^{vdw}	ΔG^{SA}	ΔG^{PB}	ΔG^{SA}	ΔG^{calc}	$\Delta\Delta G^{calc}$
mode1	Wild type	-886.50 ±14.47	-81.40 ±0.62	-10.81 ±0.11	920.29 ±15.62	-42.89 ±1.92	**-15.53 ±1.46**	
	TSP1 K32A mutant	-522.43 ±6.03	-67.60 ±2.45	-11.37 ±0.35	559.96 ±6.99	-37.48 ±4.82	**-3.96 ±1.82**	**11.57***
	TSP1 K24A&K32A mutant	-383.57 ±18.82	-69.88 ±3.63	-10.80 ±0.33	417.37 ±10.36	-46.37 ±3.35	**-0.51 ±3.48**	**15.02***
mode2	Wild type	-881.35 ±33.67	-60.15 ±1.34	-10.16 ±0.53	911.54 ±26.57	-36.78 ±4.22	**-3.34 ±2.67**	
	TSP1 K32A mutant	-704.22 ±4.34	-35.99 ±3.36	-7.46 ±0.29	710.98 ±4.41	-47.11 ±4.89	**10.42 ±3.32**	**13.76***
	TSP1 K24A&K32A mutant	-759.52 ±19.17	-68.06 ±0.73	-10.82 ±0.79	770.24 ±18.42	-43.68 ±4.06	**-24.48 ±2.88**	**-21.14***
mode3	Wild type	-964.41 ±27.64	-109.21 ±3.12	-13.66 ±0.58	1020.05 ±26.63	-43.81 ±4.17	**-23.42 ±3.93**	
	TSP1 K32A mutant	-583.36 ±20.97	-82.83 ±1.25	-11.23 ±0.70	617.04 ±17.22	-38.00 ±3.87	**-22.38 ±3.95**	**1.04**
	TSP1 K24A&K32A mutant	-234.42 ±13.92	-68.11 ±0.75	-11.32 ±0.49	269.37 ±13.25	-38.05 ±4.91	**-6.43 ±2.13**	**16.99***
mode4	Wild type	-721.36 ±28.25	-76.79 ±1.74	-12.88 ±0.89	772.41 ±32.21	-35.66 ±3.42	**-2.96 ±1.64**	
	TSP1 K32A mutant	-286.04 ±6.51	-39.02 ±2.06	-6.98 ±0.20	309.89 ±6.88	-51.70 ±2.69	**29.55 ±1.88**	**32.51***
	TSP1 K24A&K32A mutant	-192.24 ±8.20	-55.62 ±2.83	-10.18 ±0.16	225.93 ±9.56	-42.84 ±3.87	**10.73 ±1.12**	**13.69***
Experimental results (4, 5)	Wild type						**strong**	
	TSP1 K32A mutant							**abolished binding**
	TSP1 K24A&K32A mutant							**abolished binding**

All values in this table are expressed in terms of kcal/mol. ΔE^{elec} is the electrostatic energy, ΔE^{vdw} is the van der Waals energy, ΔG^{SA} is the non-polar solvation energy, ΔG^{PB} is the polar solvation energy, $T\Delta S$ is the solute entropy, ΔG^{calc} is the binding free energy for the complex, $\Delta\Delta G^{calc}$ is relative binding free energies with respect to the wild type TSP1-CRT complex.

* denotes differences that are statistically significant (Student's *t*-test, $p < 0.05$).

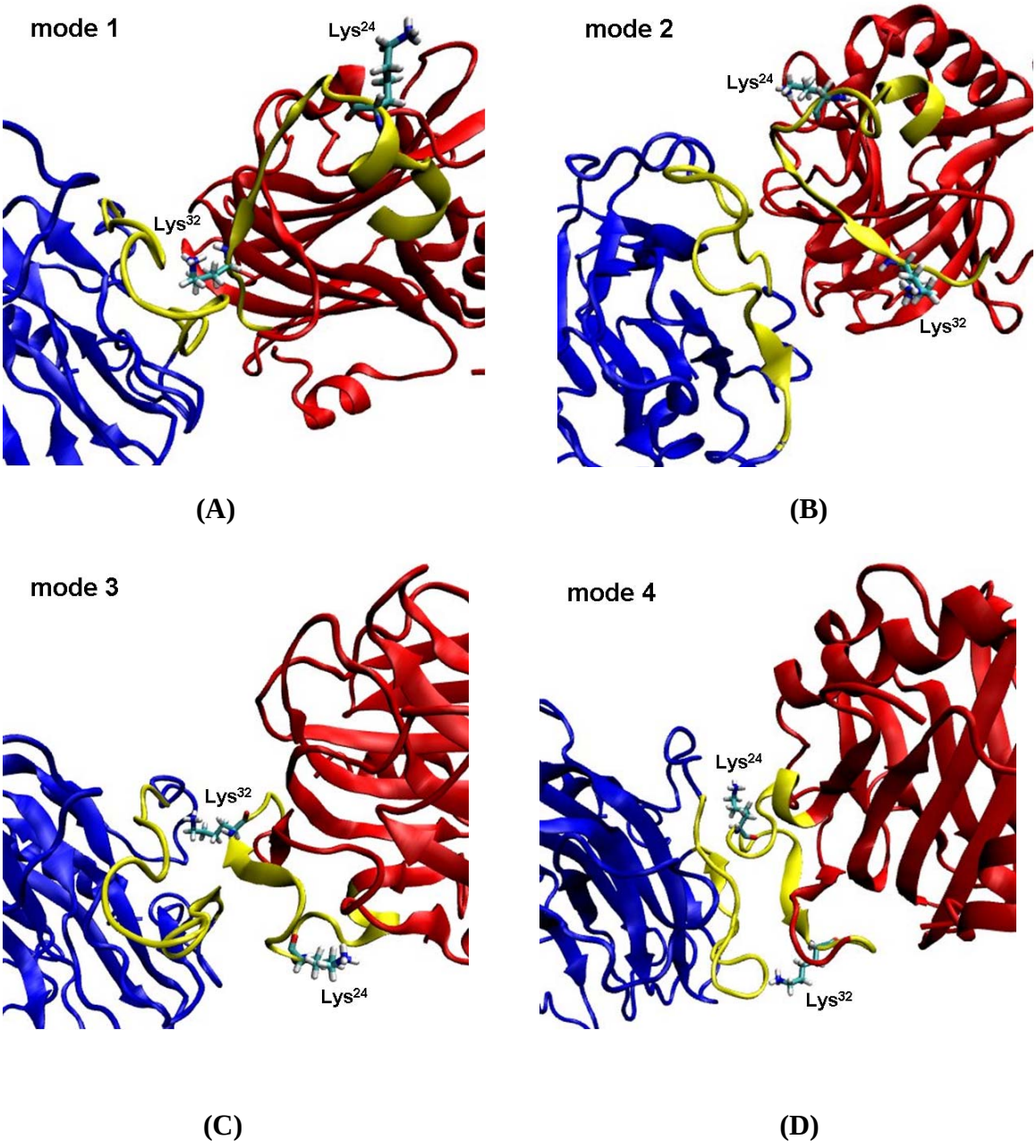
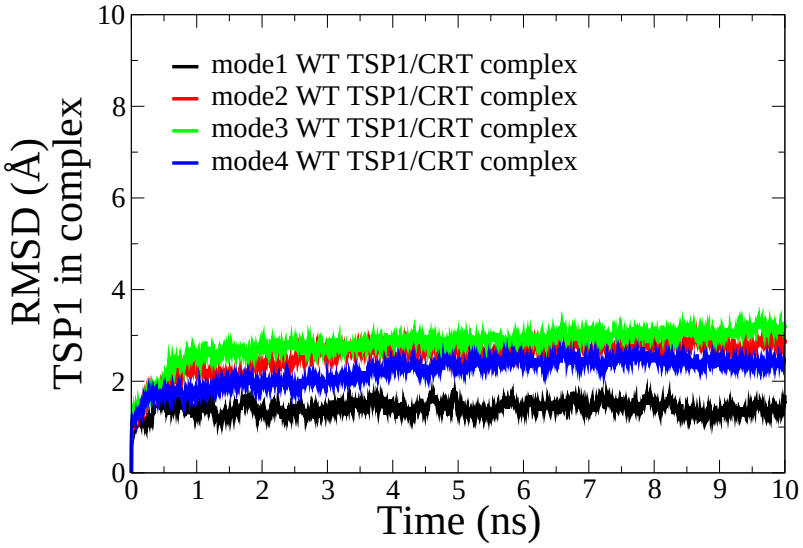
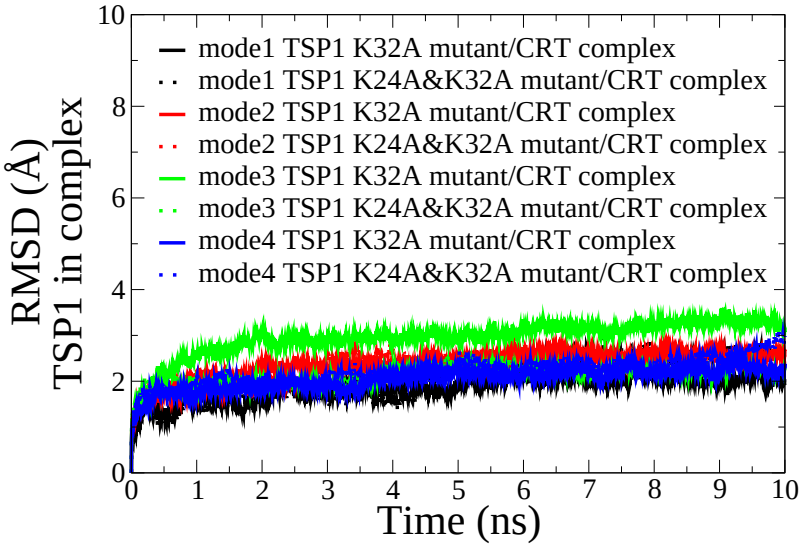


Figure 2S. Four binding modes predicted from protein docking. (A) mode 1; (B) mode 2; (C) mode 3; (D) mode 4. Mode 1 predicted from ZDOCK/RDOCK programs; Modes 2, 3, and 4 predicted from ZDOCK/ZRANK/RosettaDock programs. Red: TSP1; blue: CRT; yellow: binding sites. Images were generated with VMD software support.

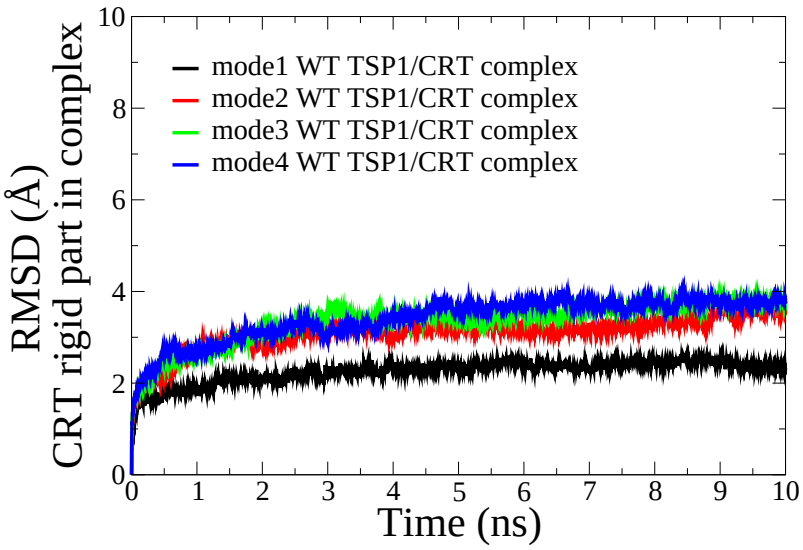
Root mean square deviation (RMSD) of TSP1 and CRT in TSP1-CRT complex for four predicated binding modes



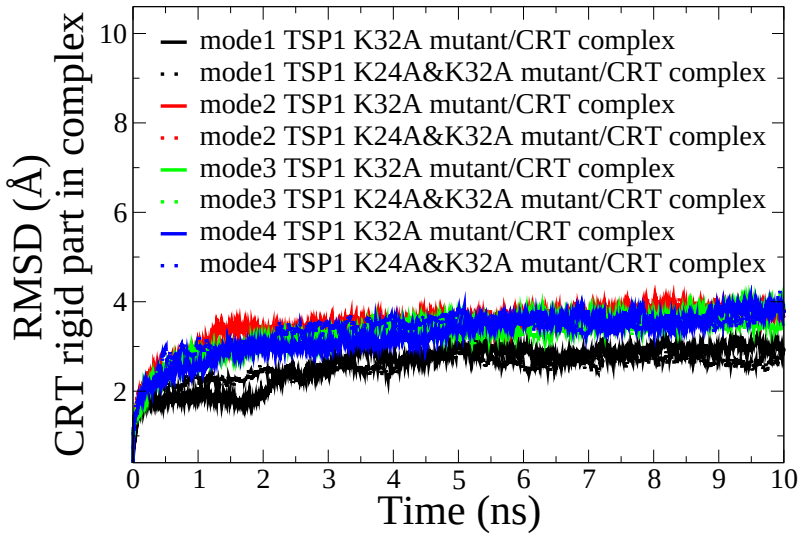
(A)



(B)



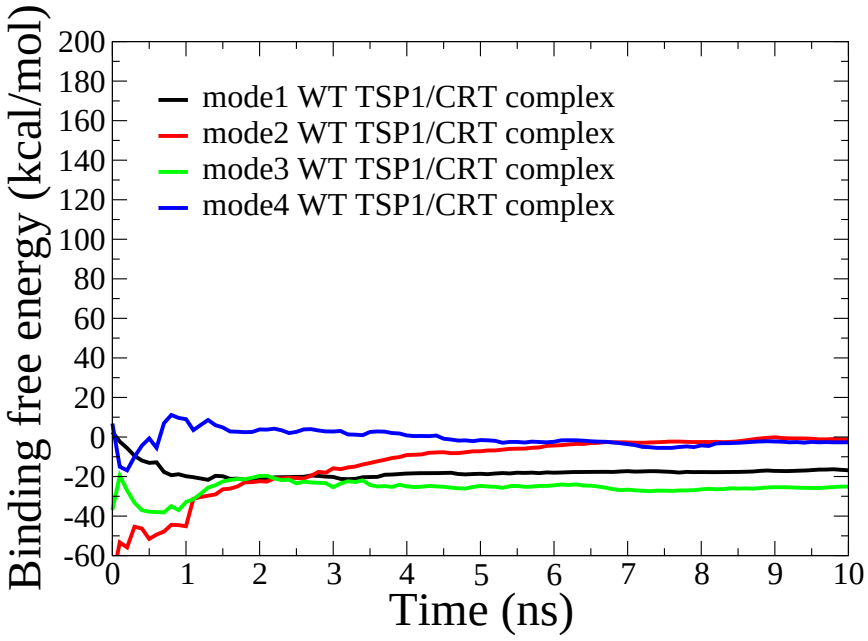
(C)

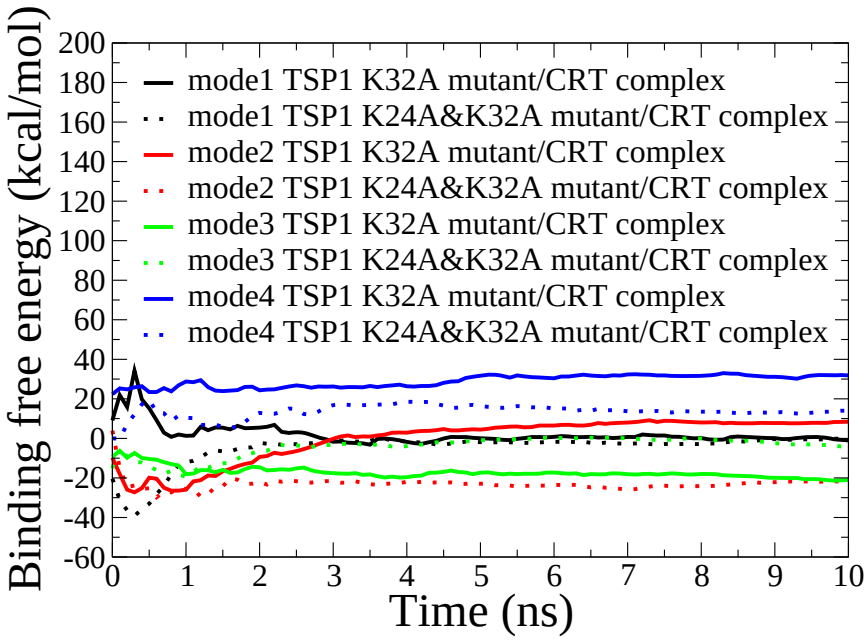


(D)

Figure 3S (A) Root mean squared deviation (RMSD) of TSP1 N-domain in WT TSP1-CRT complexes. (B) RMSD for TSP1 N-domain in TSP1 mutant-CRT complexes. (C) RMSD for CRT rigid region (N-domain and the partial C-domain) in WT TSP1-CRT complexes. (D) RMSD for CRT rigid region in TSP1 mutant-CRT complexes.

Binding free energy of the TSP1-CRT complex for four predicated potential binding modes





(B)

Figure 4S. (A) Binding free energy of the four WT TSP1-CRT complex candidates. (B) Binding free energy of TSP1 mutant-CRT complexes.

RMSD of single CRT

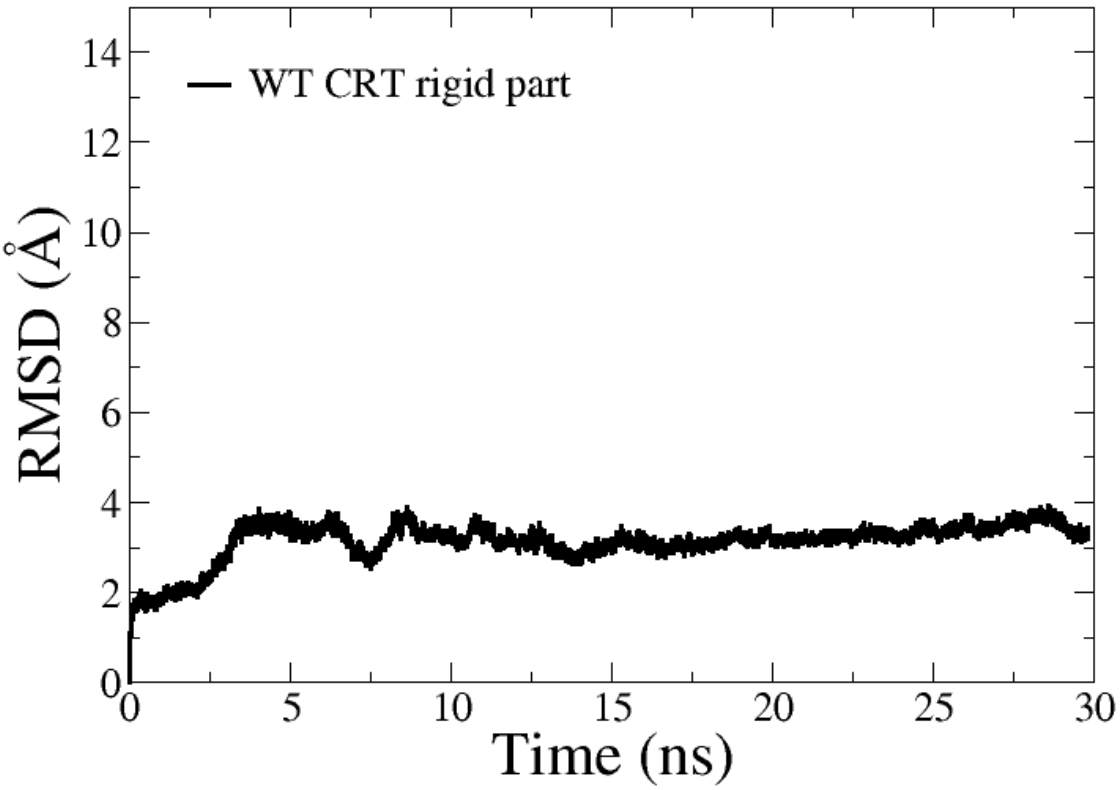


Figure 5S RMSD for CRT N-domain and the partial C-domain over the 30ns MD simulation

Conformational Change of CRT by binding to TSP1 with ANM-MD Simulations

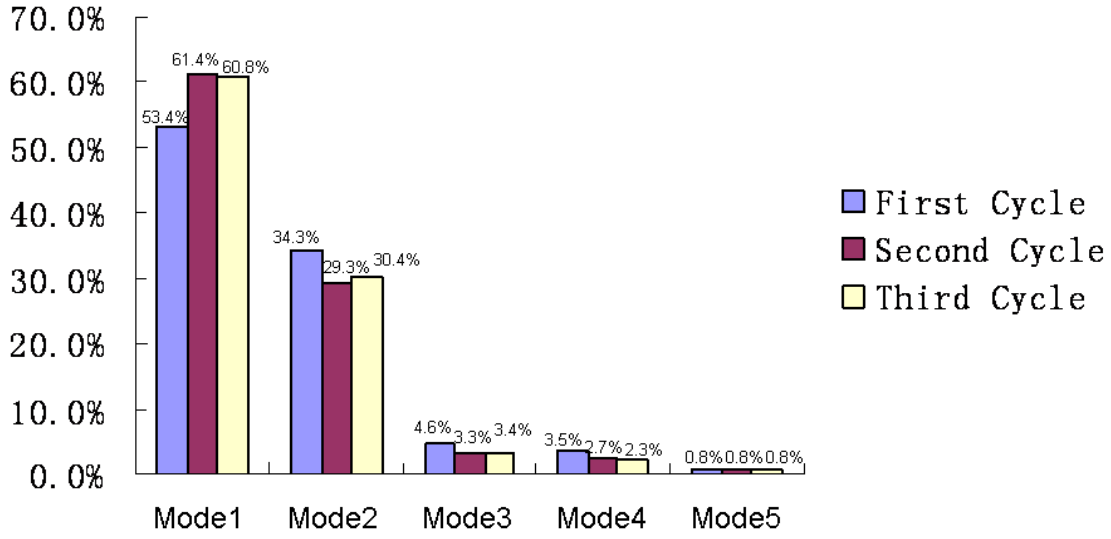
Mode selection in one cycle. For both single CRT and CRT in TSP1-CRT complex, the percentage of eigenvalues of the first mode accounted for about 60% of the whole motion, and the percentage of eigenvalues of the second mode accounted for about 30% of the whole motion. In sum, the first two modes accounted for about 90% of the whole motion, indicating that the first two modes could represent the motion (Fig. 5S). In addition, the first two modes' eigenvalues were sufficiently separate from those of the other modes. Therefore, the subset of modes including 1st and 2nd modes was distinctive. The degree of collectivity for the kth mode is calculated using:

$$K(k) = \frac{1}{N}\exp\left(-\sum_{i}^{N}\left[\Delta \overline{R}_i(k)^2 \right]\log\left[\Delta \overline{R}_i(k)^2 \right]\right) \quad (4S)$$

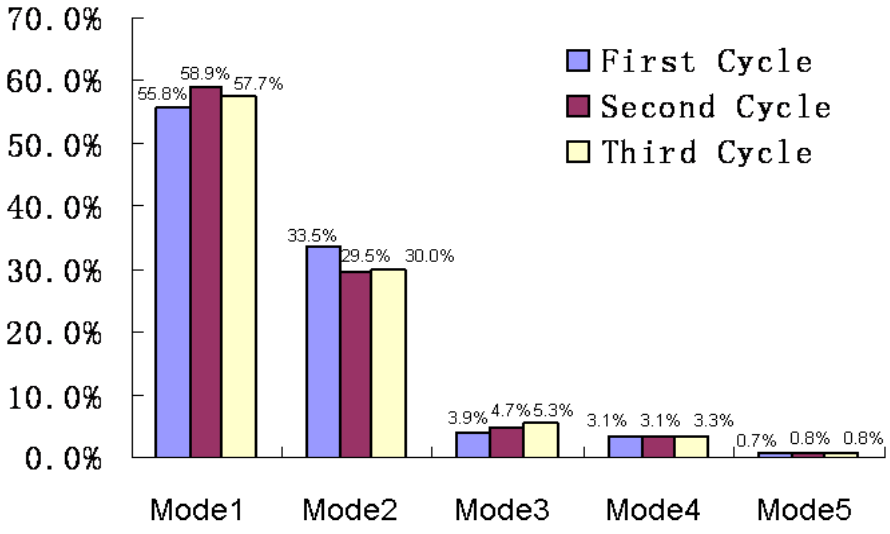
Where $\Delta\overline{R}_i(k)$ is the displacement of the ith residue driven by mode k. $\Delta\overline{R}_i(k)$ is normalized such that $\sum_i^N \left[\Delta\overline{R}_i(k)\right]^2 = 1$. It is used in order to estimate the degree of collectivity of each conformational change, reflecting the number of atoms which are significantly affected during the conformational change. The degree of collectivity, K, is confined to the interval between 1/N and 1. If K=1, the conformational change is maximally collective and all $\Delta\overline{R}_i(k)$ are identical. If K=1/N which is the minimal, only one atom is involved in the conformational change. Therefore, the more atoms involved in conformational change, the larger the value of K is. This can help to eliminate those motions that are mainly caused by loosely coupled N- or C-terminus.

In this study, because the flexible CRT P-domain always contributed more motion, the relatively low degree of collectivity was reasonable. All of the low-frequency modes (the first 5

modes in both the single CRT and CRT in complex) did not involve a large fluctuation of N- or C-terminus (Table 2S and Table 3S).



(A)



(B)

Figure 6S Percentage of eigenvalues of the first 5 modes of 3 circles (the second cycle started from mode2 of the first cycle; the third cycle started from mode2 of the second cycle) (A) single CRT; (B) CRT in complex

Table 2S Degree of collectivity for single CRT (the second cycle started from mode2 of the first cycle; the third cycle started from mode2 of the second cycle)

	K(1) mode1	K(2) mode2	K(3) mode3	K(4) mode4	K(5) mode5
Cycle1	0.42	0.55	0.40	0.53	0.25
Cycle2	0.42	0.51	0.37	0.61	0.24
Cycle3	0.40	0.55	0.42	0.56	0.25

Table 3S Degree of collectivity for CRT in complex (the second cycle started from mode2 of the first cycle; the third cycle started from mode2 of the second cycle)

	K(1) mode1	K(2) mode2	K(3) mode3	K(4) mode4	K(5) mode5
Cycle1	0.24	0.30	0.36	0.31	0.14
Cycle2	0.24	0.38	0.36	0.25	0.11
Cycle3	0.24	0.43	0.34	0.27	0.10

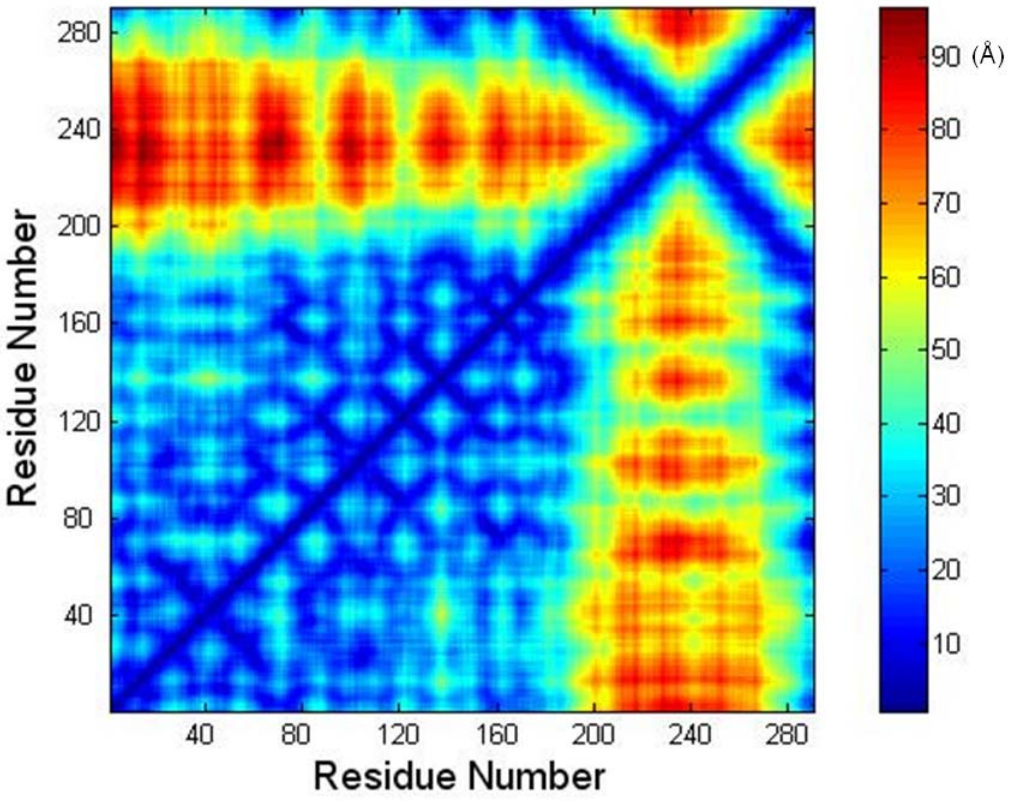


Figure 7S Distance matrix for CRT alone (bottom right) and CRT in the TSP1-CRT complex (top left) based on the final structures from ANM restraint MD simulations.

REFERENCES

1. Tan, K., Duquette, M., Liu, J. H., Zhang, R., Joachimiak, A., Wang, J. H., and Lawler, J. (2006) The structures of the thrombospondin-1 N-terminal domain and its complex with a synthetic pentameric heparin, *Structure 14*, 33-42.
2. Guo, L., Groenendyk, J., Papp, S., Dabrowska, M., Knoblach, B., Kay, C., Parker, J. M., Opas, M., and Michalak, M. (2003) Identification of an N-domain histidine essential for chaperone function in calreticulin, *The Journal of biological chemistry 278*, 50645-50653.
3. Goicoechea, S., Pallero, M. A., Eggleton, P., Michalak, M., and Murphy-Ullrich, J. E. (2002) The anti-adhesive activity of thrombospondin is mediated by the N-terminal domain of cell surface calreticulin, *The Journal of biological chemistry 277*, 37219-37228.
4. Goicoechea, S., Orr, A. W., Pallero, M. A., Eggleton, P., and Murphy-Ullrich, J. E. (2000) Thrombospondin mediates focal adhesion disassembly through interactions with cell surface calreticulin, *The Journal of biological chemistry 275*, 36358-36368.
5. Murphy-Ullrich, J. E., Gurusiddappa, S., Frazier, W. A., and Hook, M. (1993) Heparin-binding peptides from thrombospondins 1 and 2 contain focal adhesion-labilizing activity, *The Journal of biological chemistry 268*, 26784-26789.
6. Chen, R., Li, L., and Weng, Z. (2003) ZDOCK: an initial-stage protein-docking algorithm, *Proteins 52*, 80-87.
7. Chen, R., and Weng, Z. (2003) A novel shape complementarity scoring function for protein-protein docking, *Proteins 51*, 397-408.
8. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations., *Journal of computational chemistry 4*, 187–217.

9. Li, L., Chen, R., and Weng, Z. (2003) RDOCK: refinement of rigid-body protein docking predictions, *Proteins 53*, 693-707.

10. Wiehe, K., Pierce, B., Mintseris, J., Tong, W. W., Anderson, R., Chen, R., and Weng, Z. (2005) ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5, *Proteins 60*, 207-213.

11. Pierce, B., and Weng, Z. (2007) ZRANK: reranking protein docking predictions with an optimized energy function, *Proteins 67*, 1078-1086.

12. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations, *Journal of molecular biology 331*, 281-299.

13. Feig, M., Karanicolas, J., and Brooks, C. L., 3rd. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology, *Journal of molecular graphics & modelling 22*, 377-395.

14. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of computational chemistry 26*, 1668-1688.

15. Ganoth, A., Friedman, R., Nachliel, E., and Gutman, M. (2006) A molecular dynamics study and free energy analysis of complexes between the Mlc1p protein and two IQ motif peptides, *Biophysical journal 91*, 2436-2450.

16. Wang, W., Lim, W. A., Jakalian, A., Wang, J., Wang, J., Luo, R., Bayly, C. I., and Kollman, P. A. (2001) An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis, *Journal of the American Chemical Society 123*, 3986-3994.

17. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., 3rd. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Accounts of chemical research 33*, 889-897.
18. Wang, J., Morin, P., Wang, W., and Kollman, P. A. (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA, *J Am Chem Soc 123*, 5221-5230.
19. Li, L., Uversky, V. N., Dunker, A. K., and Meroueh, S. O. (2007) A computational investigation of allostery in the catabolite activator protein, *J Am Chem Soc 129*, 15668-15676.
20. Efron, B., and Tibshirani, R. J. (1998) *An Introduction to the Bootstrap*, Chapman & Hall, New York.
21. Isin, B., Schulten, K., Tajkhorshid, E., and Bahar, I. (2008) Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes, *Biophysical journal 95*, 789-803.