Supporting Information to

Probing native protein structures by chemical cross-linking, mass spectrometry and bioinformatics

Alexander Leitner et al.

Simulations

Generating virtual cross-links

Virtual cross-links in protein complexes were generated by interconnecting lysine residues with a "solvent path distance" of 9 to 24 Å. The solvent path distance corresponds to the shortest path between two amino acids, here lysine, in a grid where the path leads only through those grid cells that are occupied only by the solvent and not the protein complex. It can be calculated employing the following approach:

1.) Find all lysine residues in the protein structure.
2.) Place a large cubic grid around the first residue and label all grid cells according to their location within the solvent or protein.
3.) Check the solvent accessibility of the first lysine residue in the grid by counting the neighboring grid cells that carry a solvent label.
4.) Check whether any additional lysine residues are located within the grid and determine their solvent accessibility as in step 3.
5.) Set the distance of all grid cells containing the first residue to 0.
6.) Using a recursive-like algorithm that is based on the Travel Depth algorithm (S1) to calculate the depth of clefts in protein structures, calculate distances in an "onion-shell" fashion for all surrounding grid cells that carry a solvent label.
7.) Read out the solvent path distance between the first and the second lysine residue from the grid cells that are occupying the second lysine residues.

The cubic grid has an edge length of 2 x *k* and a grid cell length of *m*. *k* is the typical length of the employed cross-linker, i.e. the maximum distance that the second residue can have, while *m* approximates the dimension of the cross-linker's cross-section. For this study *k* was varied between 9 and 24 Å, *m* was fixed to 3 Å.

Lysine residues were required to have at least 40% of their neighboring grid cells to be occupied by solvent to be classified "accessible". Note that this calculation might not be appropriate for large oligomeric protein complexes which will require a more sophisticated determination of solvent accessibility.

Protein-Protein Docking

Native protein complexes from the Protein Data Bank (S2) as compiled to a protein-protein docking benchmark set by Chen and Weng (58) were downloaded from http://zlab.bu.edu/zdock/benchmark.shtml. The benchmark set consists of 54 protein complexes, of which 22 are enzyme-inhibitor complexes, 16 are antibody-antigen complexes, and the remaining 16 complexes are categorized as “other” or “difficult”. For each of the native complexes, 10,000 different conformations were generated with the docking protocol (60) from the Rosetta 3.1 Software Suite (http://www.rosettacommons.org/). For the docking calculations the protein backbone was held fixed while amino acid side-chains of both interaction partners were replaced by off-rotamer side chain conformations. After all conformations were generated, virtual cross-links were determined in the native crystal structure of the complex and afterwards filtered out from all 10,000 conformations.

In order to assess the upper limit of applying cross-linker data to the computational modeling of protein-protein complex, we have tested the best-case scenario for the methodology. The best case comprises:

1.) The use of the native complexes in their bound form.
2.) The assumption that all cross-links can be observed in the mass spectrometer, i.e. all combinations of lysine residues are considered that have a solvent path distance smaller than 9-24 Å.
3.) The generation of only near native conformations by allowing both binding partners to move only 3 Å apart and rotate only by 8°.

Additional references

S1. Coleman, R. G., and Sharp, K. A. (2006) Travel depth, a new shape descriptor for macromolecules: Application to ligand binding. *J. Mol. Biol.* 362, 441-458.

S2. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucl. Acids Res.* 28, 235-242.

Figure S1. Example of a false positive identification where b_{n-1} ions are identical. Diamonds in the spectrum indicate theoretical ions that were matched to peaks in the spectrum. In this case, b_{n-1} ions of both chains are identical and match on both peptide chains (also indicated with circles in the peptide sequence). The matched product ion is the only assigned fragment from the short peptide and therefore cannot be considered as evidence for the short peptide. In this experiment the proteins (rabbit kreatine kinase and chicken transferrin) were cross-linked individually and mixed only after the cross-linking reaction, confirming that this cross-link is a false positive identification.

Sequence: **GGVHVKLAHLSKHPK-KDPVLK-a6-b1** **Mr: 2443.454** **Charge: 4** **Error [ppm]: 4.7**
Proteins: **KCRM_RABIT- TRFE_CHICK**

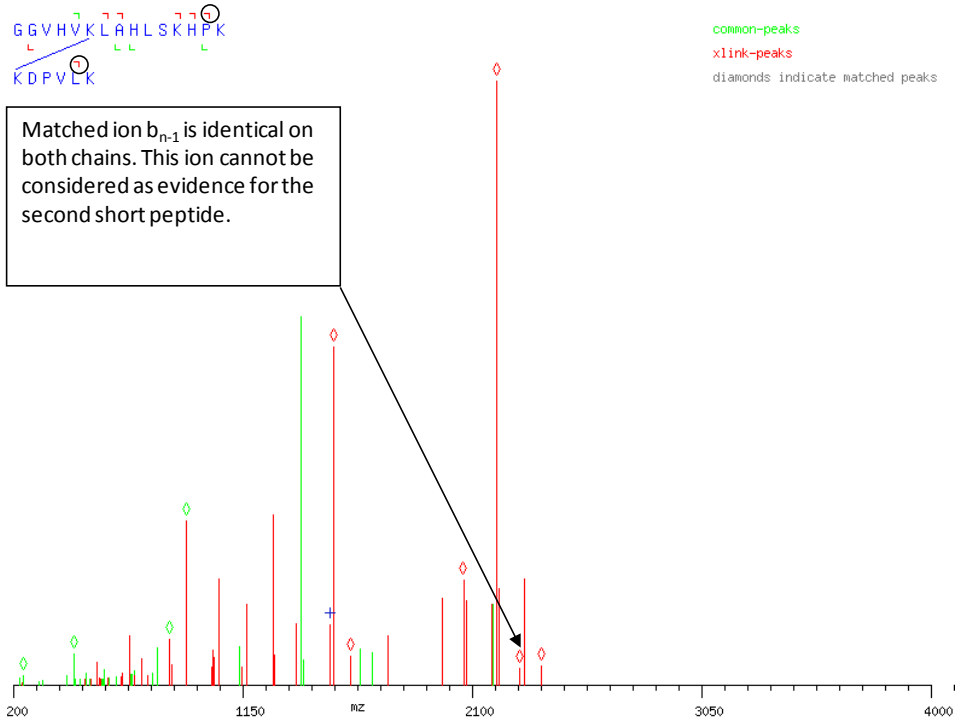

Table S1: Matched ions alpha-chain.

alpha_common_b_standard_plus1	*58.03*	*115.05*	214.12	351.18	450.25	-	-	-	-	-	-	-	-	-	-
alpha_xlink_b_-NH3_plus1	-	-	-	-	-	1397.82	1510.90	1581.94	1719.00	1832.08	1919.11	2047.21	2184.27	**2281.32**	**2409.41**
alpha_xlink_b_standard_plus1	-	-	-	-	-	1414.84	1527.93	1598.96	1736.02	1849.11	1936.14	2064.23	2201.29	**2298.35**	**2426.44**
AA	G	G	V	H	V	K	L	A	H	L	S	K	H	P	K
alpha_common_y_-NH3_plus1	-	-	-	-	-	-	1013.59	900.51	829.47	692.41	579.33	492.29	364.20	227.14	*130.09*
alpha_common_y_standard_plus1	-	-	-	-	-	-	1030.62	917.53	846.49	709.44	596.35	509.32	381.23	244.17	*147.11*
alpha_xlink_y_-NH3_plus1	2427.42	2370.40	2313.38	2214.31	2077.25	1978.19	-	-	-	-	-	-	-	-	-

Table S2: Matched ions beta-chain.

beta_xlink_b_-NH3_plus1	1857.09	1972.11	2069.17	2168.23	**2281.32**	**2409.41**
beta_xlink_b_standard_plus1	1874.11	1989.14	2086.19	2185.26	**2298.35**	**2426.44**
AA	K	D	P	V	L	K
beta_common_y_-NH3_plus1	-	554.32	439.29	342.24	243.17	*130.09*
beta_common_y_standard_plus1	-	571.35	456.32	359.27	260.20	*147.11*
beta_xlink_y_-NH3_plus1	2427.42	-	-	-	-	-
beta_xlink_y_standard_plus1	2444.45	-	-	-	-	-

Table S3: Matched ions table.

type	position	ion th	peak	delta	intensity
alpha_common_b_standard_plus1	5	450.246	450.242	0.005	5.073
alpha_common_y_standard_plus1	2	244.166	244.163	0.003	1.506
alpha_common_y_standard_plus1	7	846.495	846.513	-0.018	7.437
alpha_common_y_standard_plus1	8	917.532	917.550	-0.018	27.078
alpha_xlink_b_-NH3_plus1	7	1510.900	1510.896	0.004	9.836
alpha_xlink_b_standard_plus1	12	2064.233	2065.252	-1.019	17.255
alpha_xlink_b_standard_plus1	13	2201.292	2201.285	0.008	100.000
alpha_xlink_b_standard_plus1	14	2298.345	2298.315	0.030	2.739
alpha_xlink_b_standard_plus1	7	1527.926	1527.897	0.029	55.920
alpha_xlink_b_standard_plus1	8	1598.963	1598.929	0.034	4.663
alpha_xlink_y_standard_plus1	14	2387.429	2387.427	0.002	3.198
beta_xlink_b_standard_plus1	**5**	**2298.345**	**2298.315**	**0.030**	**2.739**

Tables S1-S3: In Tables S1 and S2 all theoretical ions from both peptide chains are listed. Matched ions are indicated by red (cross-link ions), green (common ions) and blue (NH_3 loss ions) background. Identical ions that emerge if the terminal amino acid of both peptides is the same are highlighted in dark blue. In Table S3 matched ions to peaks of the spectrum are summarized. The ion that is matched by both chains is again indicated in blue.