

Table S1. Top ten analogs identified by LOMETS for threading alignment

Rank	Protein domain Identity	Organism	PDB ID	Iden1 (%) ^a	Iden2 (%) ^b	Cov ^c	Norm. Z-score ^d
1	TBP	<i>Encephalitozoon cuniculi</i>	3eikA	14	17	0.67	4.01
2	TBP	<i>Sulfolobus acidocaldarius</i>	1mp9A	24	17	0.67	1.11
3	TBP	<i>Saccharomyces cerevisiae</i>	4b0aA	16	20	0.74	3.98
4	TBP	<i>Methanocaldococcus jannaschii</i>	2z8uA	16	13	0.63	3.87
5	TBP	<i>Pyrococcus woesei</i>	1aisA	16	15	0.68	3.87
6	TBP	<i>Sulfolobus acidocaldarius</i>	1mp9A	12	17	0.73	3.64
7	CPSF-100	<i>Saccharomyces cerevisiae</i>	2i7xA	15	16	0.90	0.53
8	TBP	<i>Homo Sapiens</i>	1jfiA	13	17	0.65	0.67
9	GS	<i>Chromohalobacter salexigens</i>	5dm3A	9	15	0.88	0.61
10	TBP	<i>Saccharomyces cerevisiae</i>	1ytbA	27	17	0.67	0.62

2

3 a) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

4 b) Iden2 is the percentage sequence identity of the whole template chains with query sequence

5 c) Cov- Represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of the
6 query protein.7 d) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a normalized Z-score>1 means a good
8 alignment and vice versa

- 9 CPSF-100- Cleavage and Polyadenylation Specificity Factor, GS- Glutamate Synthetase, PDB-
- 10 Protein Data Bank TBP- Tata Binding Protein.