Appendix 2

Table A2.1. A summary of the base-pairs in the target sequences that are in contact with the protein. The first two columns give the name of the protein family and the representative PDB structure. We provide the length of the target sequences for each family, the number of base-pairs in contact with the protein and the number of base-pairs that are not in contact with the protein. We also give the number of base-pairs that are variable in the target site definition: these are also separated into those in contact with the protein and those that are not. Base-pairs are almost always variable if not contacted by the protein. They are also sometimes variable if they are interacting.

Table A2.2. List of PDB codes in multi-specific families that contain more than one complex structure. The PDB code, chain or domain ID, subfamily (see appendix 1) and bound DNA sequence is given. Structures in the same subfamily bind the same DNA sequence and structures in different subfamilies bind different DNA sequences.

Table A2.3. The PET91 substitution matrix. The single-letter amino acid codes are provided in the column and row headings. Scores for each pairwise mutation is normalized between 0 and 100.

Table A2.1

Family		Total number of base-pairs			Total number of variable base-pairs		
		Target sequence	Contacted	Non-contacted	Target sequence	Contacted	Non-contacted
Pu1 ETS domain	1pueE	5	2	3	2	-	2
Prd paired domain	1pdnC	15	5	10	13	2	10
Trp repressor	1trrA	4	3	1	2	1	1
Loop-sheet-helix	1tsrB	5	3	2	3	2	1
Leucine Zipper	2dgcA	4	3	1	3	2	1
Papillomavirus-1 E2	2bopA	6	3	3	2	-	2
TBP	1ytbA	8	8	-	4	4	-
T-domain	1xbrA	8	4	4	3	-	3
RHR	1nfkA	5	3	2	1	1	-

Table A2.2

PDB code	Chain/ domain	Subfamily	DNA sequence bound
Homeodomain			
1fjl,1hdd *	A, B	12	-TAAT-
1au7,1oct *	A2	15	-AAAT- (-TAAT-)
1au7	A1	individual	-ATAC-
1oct	A1	individual	-ATGC-
1yrn	A	individual	-TGTA-
1yrn	B	16	-CATC-
ββα-zinc finger			
1aay,1zaa	A1,A3,C1,C3	18	-GCG-
"	A2,C2	7	-TGG-
2drp	A1,B1	11	-GAT-
"	A2,B2	17	-AGG-
1ubd	C1	37	-GAC-
"	C2	35	-GGA-
"	C3	41	-AAT-
1mey	C1,F1	26	-GAA-
"	C2,F2	31	-GCA-
"	C3,F3,G	2	-GAG-
Hormone receptor			
2nll	A,B	1	-AGGTCA-
1hcq	A,B,E,F	1	-AGGTCA-
1lat	A,B	1	-AGGTCA-
1glu	A,B	2	-AGAACA-
HLH			
1hlo,1an2,1an4	A,B	1	-.CAC-
1mdy	A,B	2	-ACAG-

Table A2.3

Amino acid	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	X	-
C	100	50	40	35	40	35	30	25	25	25	35	30	25	25	40	20	40	30	50	10	50	0
S	50	100	55	55	55	55	55	50	50	45	45	50	50	40	45	35	45	35	35	40	50	0
T	40	55	100	50	55	50	50	50	50	45	45	45	50	45	50	40	50	35	35	25	50	0
P	35	55	50	100	55	45	45	45	45	50	50	50	45	40	40	35	45	25	25	20	50	0
A	40	55	55	55	100	55	50	50	50	50	45	40	45	45	45	40	50	30	35	20	50	0
G	35	55	50	45	55	100	50	55	50	45	40	35	40	35	35	30	45	25	25	15	50	0
N	30	55	50	45	50	50	100	60	55	55	60	50	55	40	40	35	40	30	40	30	50	0
D	25	50	50	45	50	55	60	100	65	60	55	45	50	35	40	30	40	20	30	15	50	0
E	25	50	50	45	50	50	55	65	100	60	55	45	50	40	40	35	40	25	30	15	50	0
Q	25	45	45	50	50	45	55	60	60	100	65	55	55	45	40	40	40	25	30	25	50	0
H	35	45	45	50	45	40	60	55	55	65	100	60	50	40	40	40	40	40	50	35	50	0
R	30	50	45	50	40	35	50	45	45	55	60	100	65	50	40	35	40	30	30	60	50	0
K	25	50	50	45	45	40	55	50	50	55	50	65	100	50	40	35	40	25	30	35	50	0
M	25	40	45	40	45	35	40	35	40	45	40	50	50	100	60	70	60	50	40	30	50	0
I	40	45	50	40	45	35	40	40	40	40	40	40	40	60	100	60	70	55	45	25	50	0
L	20	35	40	35	40	30	35	30	35	40	40	35	35	70	60	100	60	60	45	40	50	0
V	40	45	50	45	50	45	40	40	40	40	40	40	40	60	70	60	100	45	40	20	50	0
F	30	35	35	25	30	25	30	20	25	25	40	30	25	50	55	60	45	100	85	50	50	0
Y	50	35	35	25	35	25	40	30	30	30	50	30	30	40	45	45	40	85	100	50	50	0
W	10	40	25	20	20	15	30	15	15	25	35	60	35	30	25	40	20	50	50	100	50	0
X	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	0
-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0