ELECTRONIC SUPPLEMENTARY INFORMATION

Structural modeling of the AhR:ARNT complex in the bHLH-PASA-PASB region elucidates the key determinants of dimerization

Dario Corrada, Michael S. Denison, Laura Bonati

Methods

Model optimization

To remove bad contacts and adjust non-optimal lengths and angles, the AhR:ARNT models were subjected to energy minimization followed by a short MD simulation with the AMBER 14 software,[1] by using the *AMBER force field 99SB*.[2] The models were placed in a periodic octhaedron box with a minimum distance of 10 Å from the protein surface. The system was solvated using the TIP3P water model [3] and net charges were neutralized by adding counterion. The energy of each system was initially minimized with 500 steps of steepest descent plus 1500 steps of conjugate gradient. The minimization process was repeated twice: firstly, holding the backbone atoms with a restraint of 500 kcal $mol^{-1}Å^{-1}$ and then without restraints. Systems were subjected to MD simulations for 2100ps (time-step 0.002 ps) using the Particle-Mesh Ewald method (PME) for long-range Coulomb interactions.[4] Initially, systems were simulated for 100 ps in NVT conditions, ramping the temperature from 0 to 100 K and with backbone positions restraint of 4 kcal mol^{-1} $Å^{-1}$. Afterwards, systems were simulated in NPT conditions and backbone positions restraint of 1 kcal mol^{-1} $Å^{-1}$ for 1000 ps, ramping the temperature from 100 to 300 K in the first 750 ps. Finally, simulations were extended for further 1000 ps with no restraints.

Loop modeling

Missing residues in the PAS-A loops of the X-ray template structures were built for the AhR:ARNT model using the Rosetta all-atom de novo loop modelling method with the next generation kinematic closure (NGK) procedure.[5] The kinematic closure approach (KIC) randomly perturbs all but three pairs of φ/ψ angles in a backbone segment, the remaining pairs of torsion angles are solved analytically to close the chain break making a valid peptide segment.[6] The models proposed undergo Monte-Carlo simulated annealing for rotamer-based side-chain optimization ("repacking"). In the next generation variant, the annealing methods gradually ramp the weights of the repulsive and Ramachandran terms of the Rosetta energy function and neighbor-dependent Ramachandran propensities and ω angle sampling are included.[5] Fragment libraries of 3mers and 9mer were generated with the Robetta server.[7] The initial loop-build starts from the fragment data and the perturbations of coupled φ/ψ/ω torsions are sampled from the given fragment libraries.

Binding Free Energy

The binding free energy ($\Delta G_{binding}$) for dimer formation was calculated by means of the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method implemented in the AMBER software package.[8,9] Calculations were performed starting from an ensemble of snapshots sampled in the 1000 ps of the MD simulation. The input structures were treated in implicit solvent. The polar solvation term was approximated with the Generalized Born (GB) model [10] by using the OBC re-scaling of the effective Born radii.[11] A physiological salt concentration (0.154 M) was chosen to take into account the electrostatic screening effect of salt.[12] The non-polar solvation term was calculated as the product of the surface tension parameter (set to 0.0072 kcal mol^{-1} $Å^{-2}$), and the solvent accessible surface area (SA) was evaluated using the Linear Combination of Pairwise Overlap (LCPO) algorithm.[13]

Tables

Table S1: The most buried residues in the modeled AhR:ARNT PPI interfaces (ΔSASA values, obtained with the POPSCOMP method,[14] higher than 60 Å^2)

	4F3L model				4ZP4 model			
protein	region	PPI subregion	residue	ΔSASA (Å^2)	region	PPI subregion	residue	ΔSASA (Å^2)
AhR	bHLH	1	L42	69.91	bHLH	1	L42	80.25
		1	L52	119.31		1	E45	60.48
		1	L66	83.21		1	L49	64.01
		1	R77	76.74		1	L52	67.51
	bHLH/PAS-A linker	2	R93	74.74		1	L66	67.58
		2	Q99	78.53		1	L76	72.13
		2	W107	88.71	PAS-A	2	Q108	88.42
	PAS-A	2	L116	76.57		2	L110	74.28
		2	R236	95.86		2	E112	94.01
		2	H241	74.79		2	L116	65.43
						2	A119	75.59
						2	Q234	111.33
						2	R236	79.98
						2	K246	82.98
						2	G247	72.20
	PAS-A/PAS-B linker	3	Q267	108.08		2	K248	79.75
	PAS-B	4	L325	73.12	PAS-A/PAS-B linker	3	R275	86.26
		3	S352	66.00	PAS-B	4	Q317	83.32
ARNT	bHLH	1	Y108	60.78	bHLH	1	M105	82.57
		1	M115	102.92		1	Y108	147.49
		1	M139	98.12		1	E111	82.79
		2	R143	84.01		1	M115	108.89
	bHLH/PAS-A linker	2	S149	94.74		1	R133	63.26
		2	Y154	78.02		1	M139	123.88
		2	F158	70.88				
	PAS-A	2	K165	91.69	PAS-A	2	L159	98.76
		2	H166	73.32		2	Q162	61.99
		2	L167	94.30		2	E163	87.26
		2	I168	90.11		2	L167	89.39
		2	E170	61.21		2	I168	61.47
		3	R260	76.41		2	D191	60.31
		2	K313	101.13		3	D216	69.53
		3	R342	123.76				
	PAS-A/PAS-B linker	3	M354	61.79	PAS-A/PAS-B linker	3	V345	83.25
		3	Q359	102.41				
	PAS-B	3	R379	72.98	PAS-B	4	P449	71.16
		4	P449	81.86		4	Y450	125.52
		4	Y450	123.98				

Figures

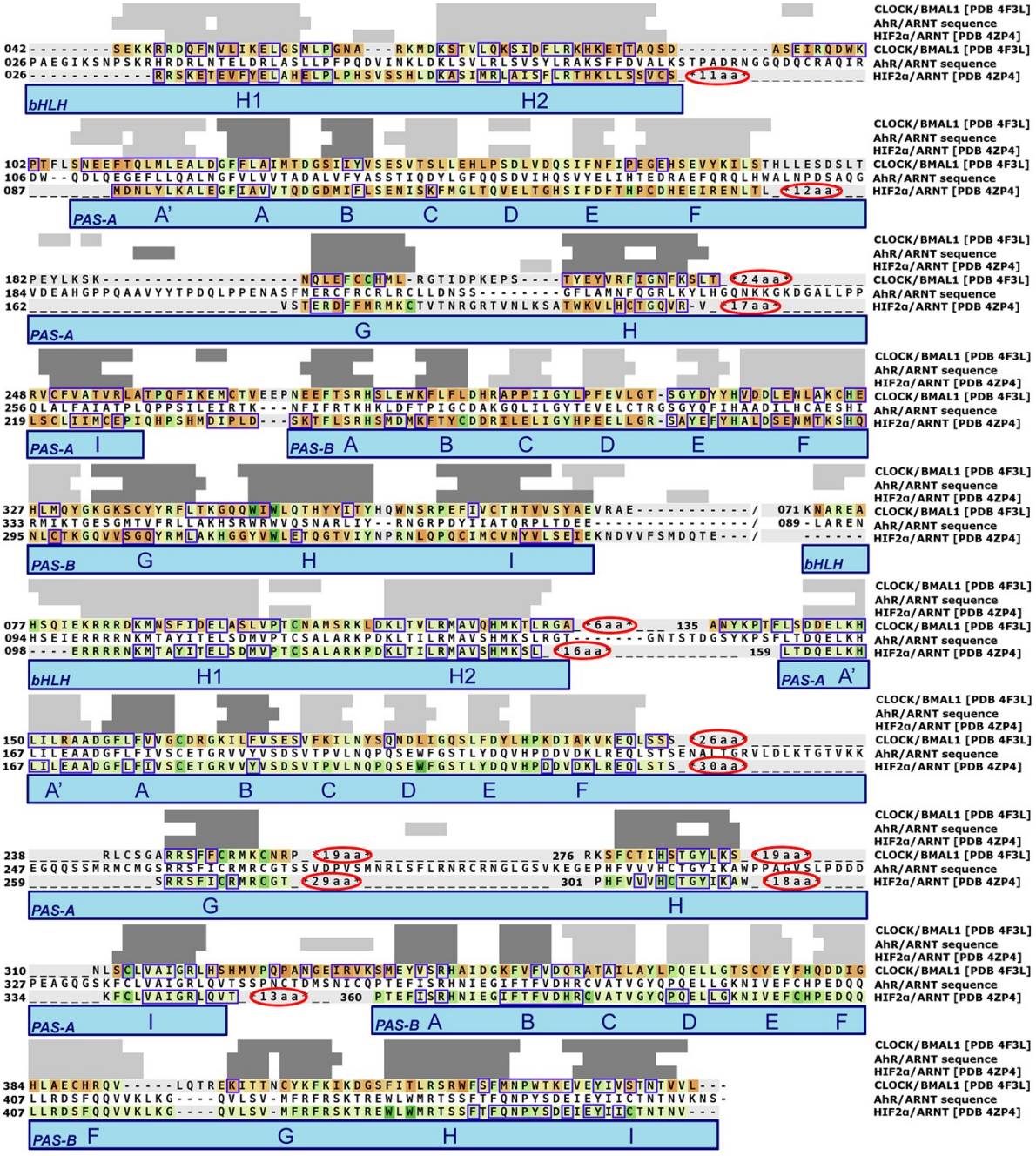

Fig. S1- Sequence alignment between target and templates. The heading bars depict the secondary structure assignment according to PSIPRED prediction[15] for the target sequence (AhR/ARNT) or DSSPcont attribution[16] for the templates (light grey: helices; dark grey: strands). Residues in the template sequences are colored according to the score assignment of the BLOSUM45 substitution matrix, based on the pairwise alignments with the target sequence (red, most dissimilar aligned residues; green, most similar aligned residues). Residues belonging to the template PPI interfaces (defined on the basis of the ΔSASA calculated with POPSCOMP[14]) are highlighted by blue boxes. Red circles depict regions of the AhR/ARNT sequences where no structural information is available (missing regions in the template structures). The lower cyan bar

depicts the boundaries of each domain, and contains the labels of the SS elements generally adopted for these domains.

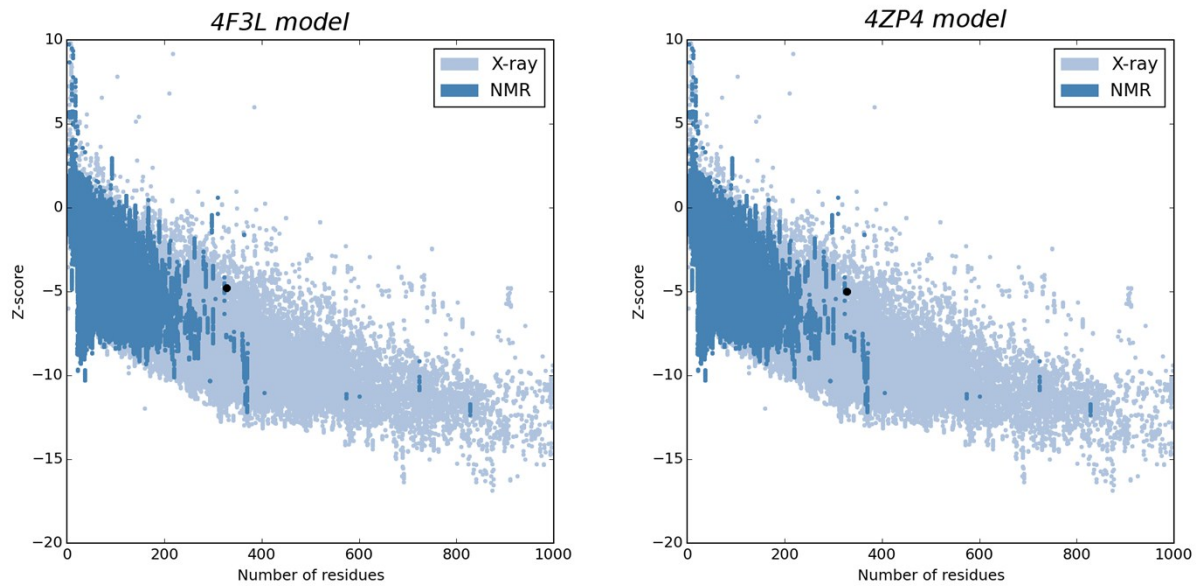

Fig. S2- Structural accuracy of the AhR:ARNT models, validated by ProSA knowledge-based potential.[17,18] The distribution of the Z-score is represented along the sequence lengths taken from a dataset of experimentally resolved 3D protein structures. Both the models presented in this work fall into such distribution.

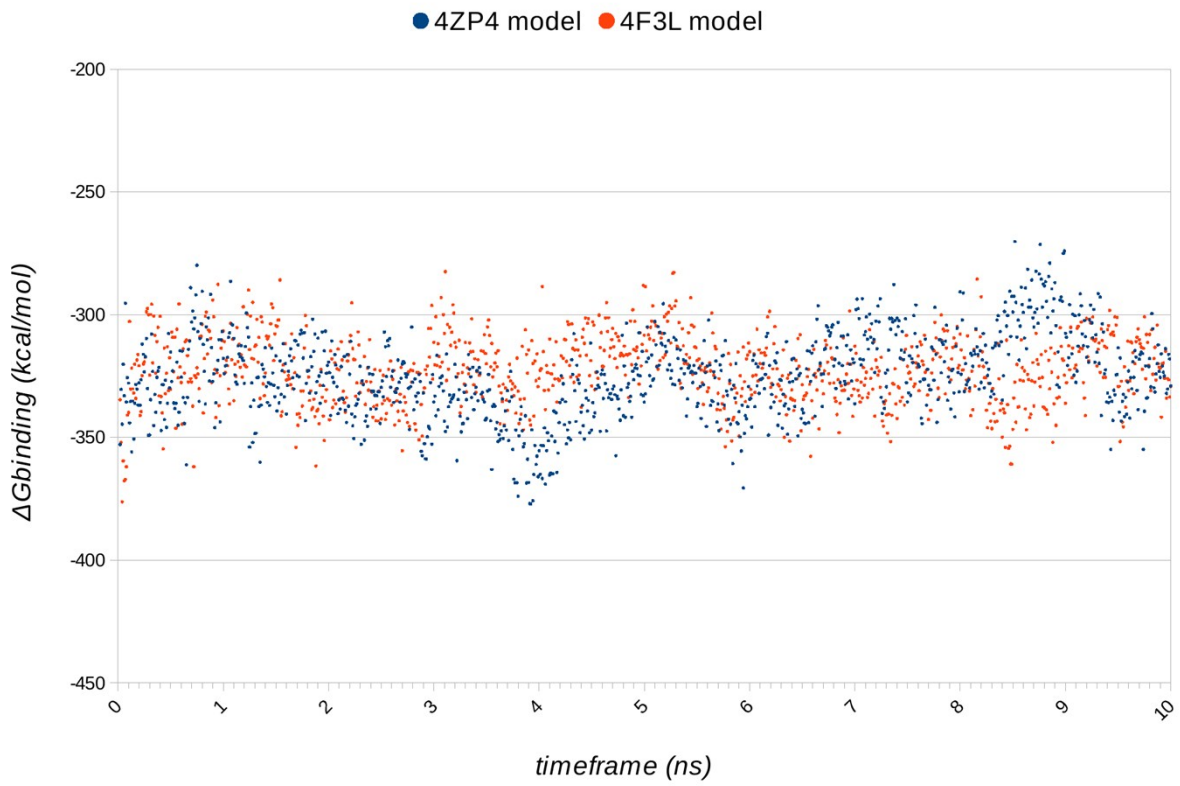

Fig. S3- Distribution of the $\Delta G_{binding}$ along 10 ns MD simulations of the AhR:ARNT dimer models. The $\Delta G_{binding}$ values are calculated with the MM-GBSA method.[8,9] For both the models, values are normally distributed (Shapiro-Wilk $W > 0.99$, ρ-value ≤ 0.01).

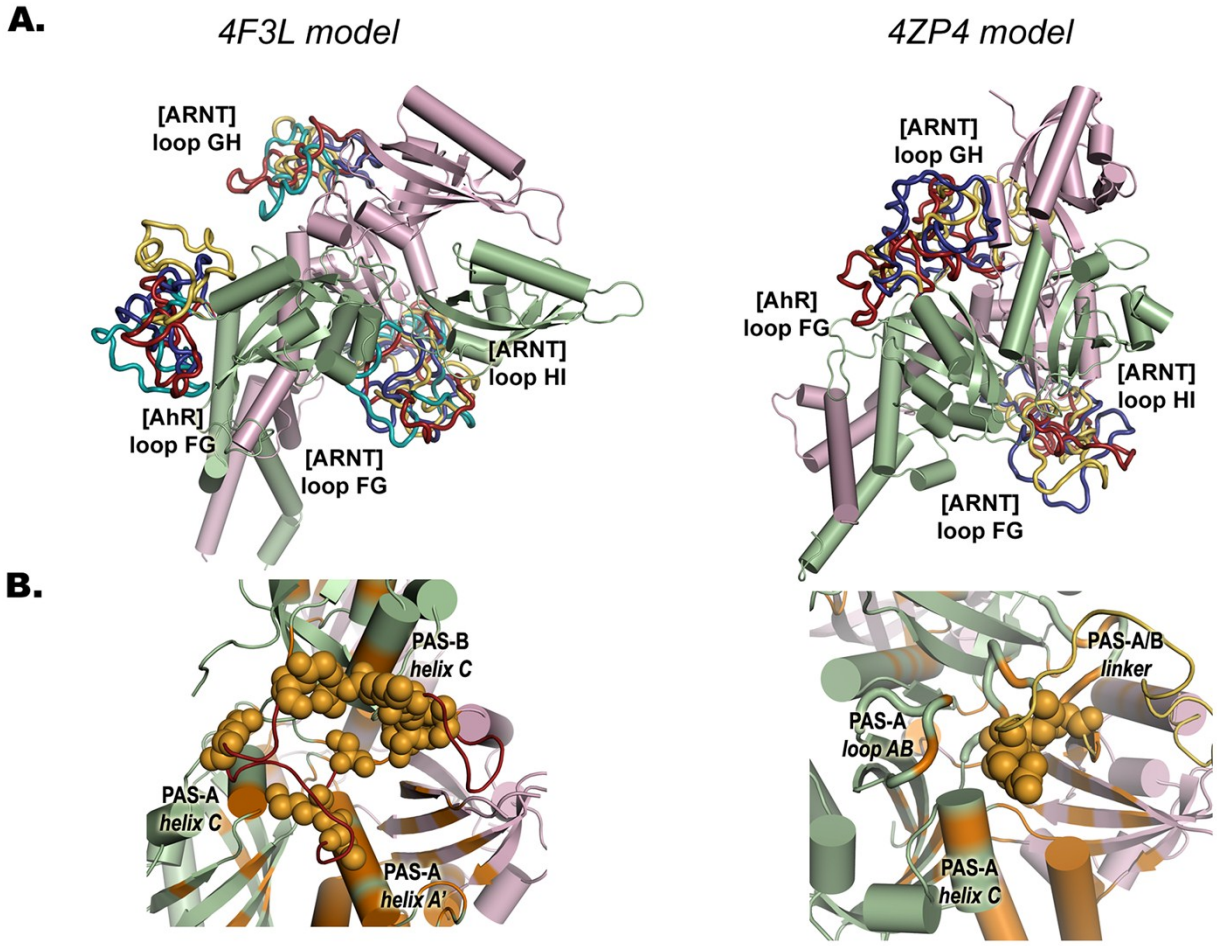

Fig. S4 Missing PAS-A loops modeled with Rosetta. (A) Mapping of reconstructed loops in the two AhR:ARNT model proposed; the representative conformations in the ensemble of 1000 generated models for each loop are colored in cyan, red, blue and yellow. (B) Putative extension to the dimerization interfaces provide by the ARNT FG loop. Residues belonging to the PPI interface are colored in orange, additional PPIs introduced by the modeled loop is represented in sphere.

References

1 D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatman, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossvary, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman (2014) AMBER 14. University of California, San Francisco.

2 V. Hornak V, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins,* 2006, **65**, 712–25.

3 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.,* 1983, **79**, 926.

4 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.,* 1993, **98**, 10089.

5 A. Stein and T. Kortemme, *PLoS One*, 2013, **8**, e63090.

6 D. J. Mandell, E. A. Coutsias and T. Kortemme, *Nat. Methods*, 2009, **6**, 551–2.

7 D. E. Kim, D. Chivian and D. Baker, *Nucleic Acids Res.*, 2004, **32**, W526-31.

8 N. Homeyer and H. Gohlke, *Mol. Inform.*, 2012, **31**, 114–122.

9 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke andA. E. Roitberg, *J. Chem. Theory Comput.*, 2012,**8**, 3314–3321.

10 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, **100**, 19824–19839.

11 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383–94.

12 J. Srinivasan and M. Trevathan, *Theor. Chem. Acc.*, 1999, 101, 426–434.

13 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.

14 J. Kleinjung and F. Fraternali, *Nucleic Acids Res.*, 2005, **33**, W342-6.

15 D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson and D. T. Jones, *Nucleic Acids Res*., 2013, **41**, W349-57.

16 C. A. F. Andersen, A. G. Palmer, S. Brunak and B. Rost B, *Structure*, 2002, **10**, 175–84.

17 M. Wiederstein and M. J. Sippl, *Nucleic Acids Res.*, 2007, **35**, W407-410.

18 M. J. Sippl, *Proteins*, 1993, **17**, 355–362.