

Supplemental Information

A Dynamic Folded Hairpin Conformation

Is Associated with α -Globin Activation

in Erythroid Cells

Andrea M. Chiariello, Simona Bianco, A. Marieke Oudelaar, Andrea Esposito, Carlo Annunziatella, Luca Fiorillo, Mattia Conte, Alfonso Corrado, Antonella Prisco, Martin S.C. Larke, Jelena M. Telenius, Renato Sciarretta, Francesco Musella, Veronica J. Buckle, Douglas R. Higgs, Jim R. Hughes, and Mario Nicodemi

Supplemental Information

for

***A dynamic folded hairpin conformation is associated with
 α -globin activation in erythroid cells***

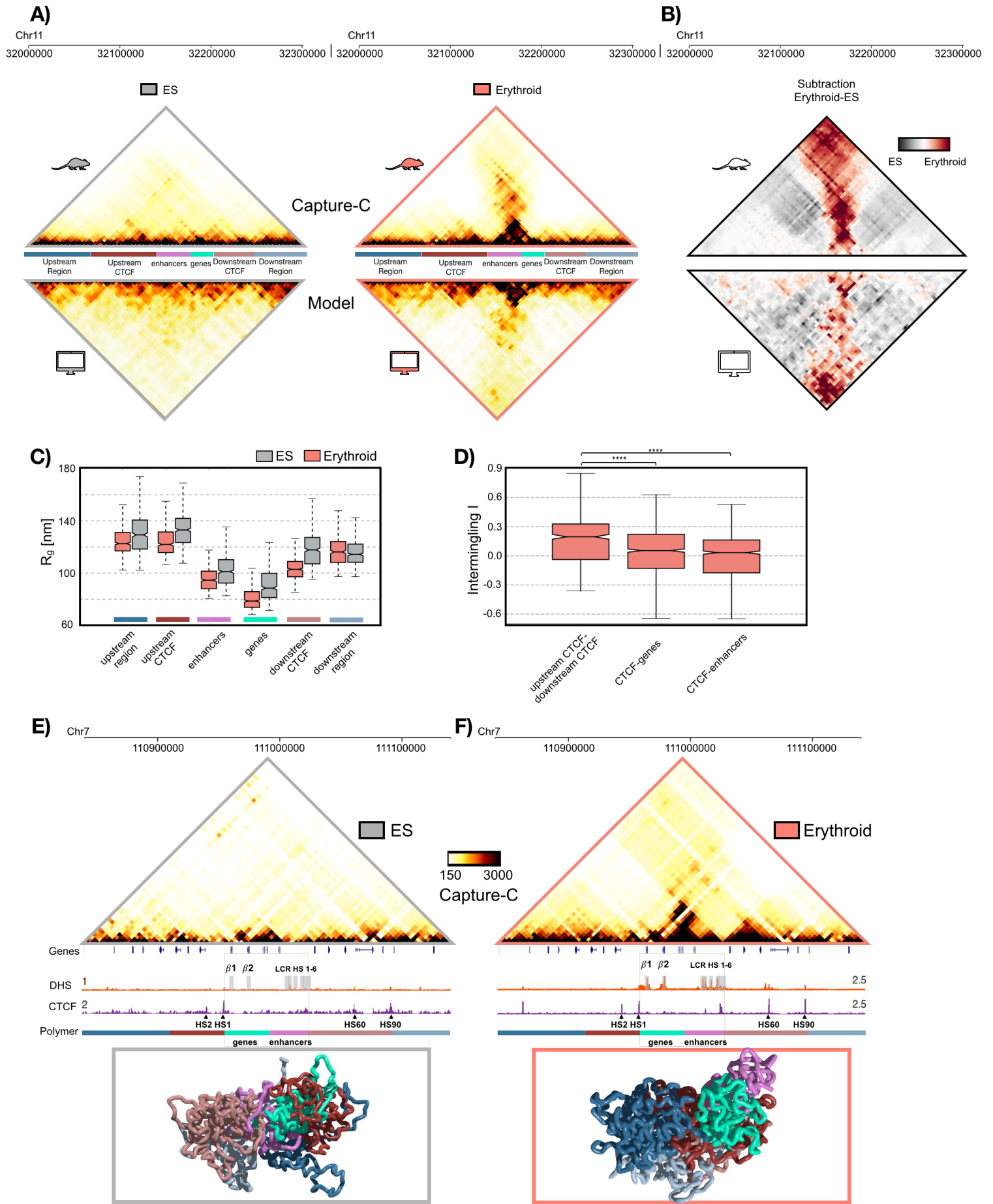


Figure S1 (related to Figure 1). Comparison between the α -globin Capture-C and SBS model inferred contact maps and model derived single-allele conformations of the β -globin locus

A) Capture-C data (top matrices, see Fig.1) and SBS polymer model derived (bottom matrices) contact maps, in ES (left) and erythroid (right) cells. Direct comparison shows that the model accurately recapitulates the experimental data (Pearson correlation $r=0.96$ and distance-corrected correlation $r'=0.87$ and $r'=0.91$ for ES and erythroid respectively). **B)** Subtraction map erythroid-ES of Capture-C (top) and simulated (bottom) contact matrices, normalized by the average contact at a fixed genomic distance in ES. The peculiar elongated shape of the red pixels along the anti-diagonal direction is also well captured by the simulated matrices. **C)** Distribution of the gyration radius in the different polymer sub-regions shown in Panel A and Figure 1A, according to the color scheme used there. Note how those regions are more localized in erythroid than in ES cells. The average gyration radius of the sub-regions considered is approximately 110nm and the maximum distances between the beads of each region is roughly as twice as the average gyration radius. **D)** Distribution of the intermingling in erythroid among the sub-regions containing CTCF, promoters and enhancers. Their coordinates are approximately overlapping with the color scheme of Panel A. Upstream-downstream CTCF regions tend to intermingle between them rather than with enhancer or gene regions, since their average intermingling is approximately zero. The stars indicate Mann-Whitney test $p\text{-val}<10^{-4}$. **E)** High resolution Capture-C data (4kb resolution) for a 300kb window around the β -genes for ES (left) and erythroid (right) cells. Gene annotation, DNaseI Hypersensitive Sites (DHS) and CTCF are shown below. The grey bars indicate the $\beta1$ and $\beta2$ gene promoters and the LCS HS 1-6 enhancers. The dashed box highlights the enhancer-promoter region. The colored bar represents the linear color scheme used for the polymer structures. The corresponding 3D conformations obtained from MD are reported below. Their corresponding contact maps are shown in Figure S4 (Pearson correlation with Capture-C is above $r=0.95$). Genomic coordinates: chr7:110,840,000-111,400,000 mm9. **F)** High resolution Capture-C data (4kb resolution) for a 300kb window around the β -globin genes for erythroid cells. Data from (Oudelaar et al., 2018).

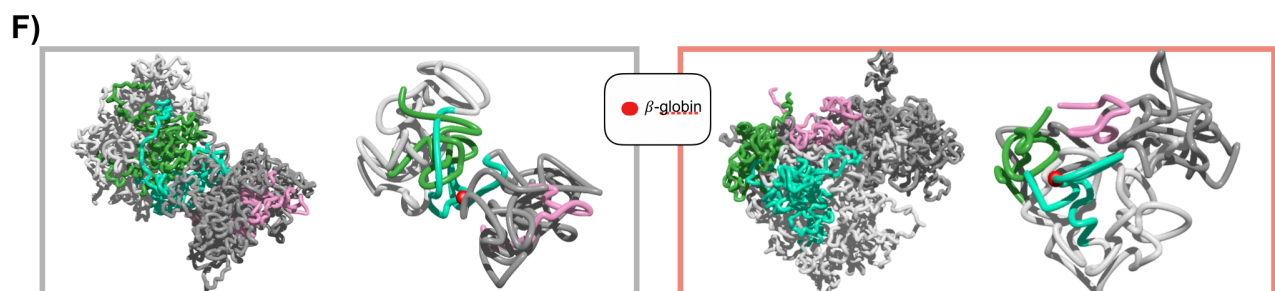
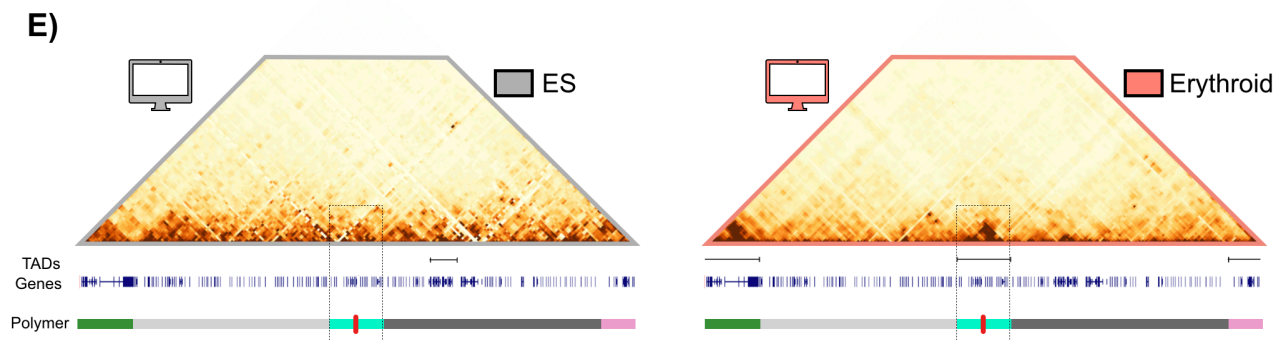
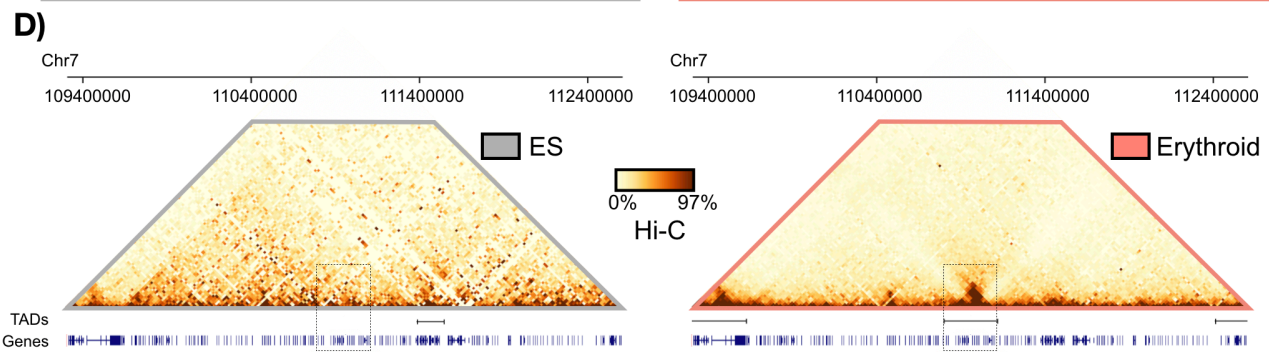
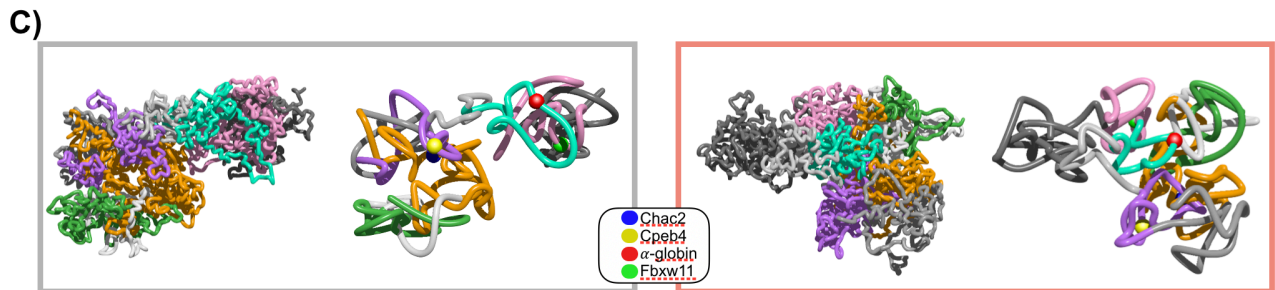
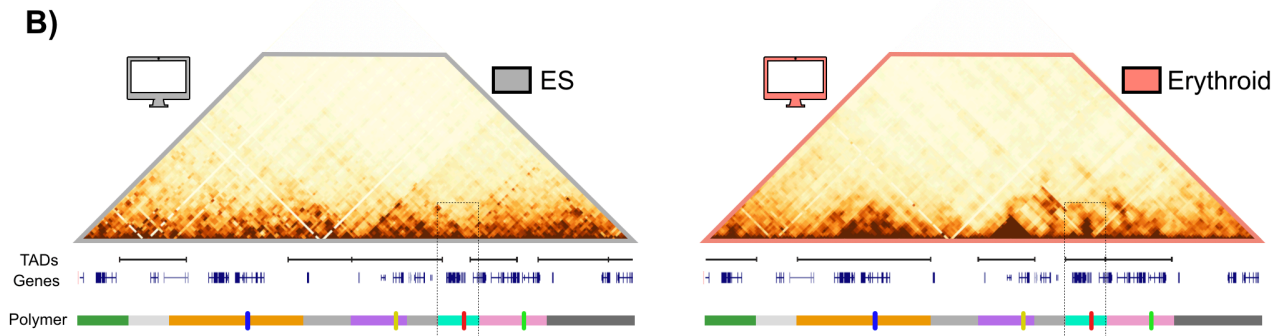
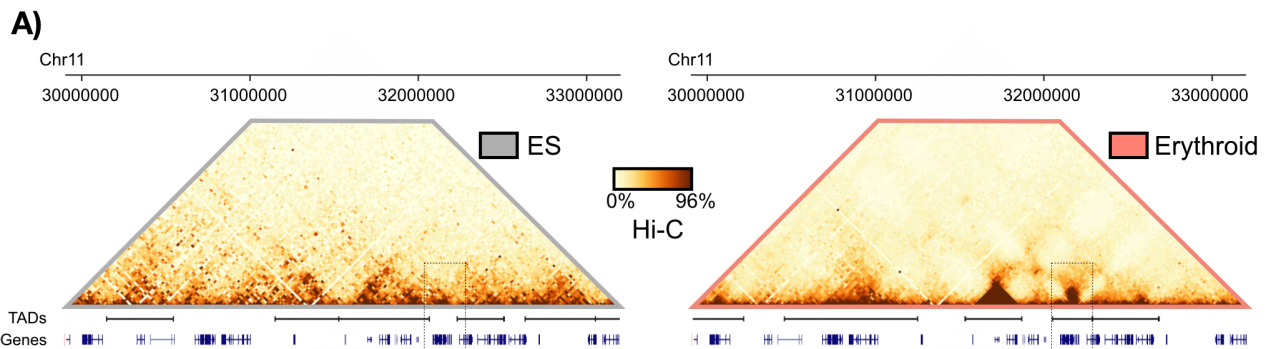


Figure S2 (Related to Figure 1 and STAR Methods). Model derived single-allele conformations of the extended α - and β -globin loci.

A) Hi-C contact matrices (20 kb resolution) of the α -globin locus in ES (left) and erythroid (right) cells. TADs and gene annotation are shown below. Dashed boxes highlight the erythroid TAD containing the α -globin genes. Genomic coordinates: chr11:29,900,000-33,200,000, mm9. Data from (Oudelaar et al., 2018) (erythroid) and from (Giorgetti et al., 2016) (ES). **B)** Contact matrices (20 kb resolution) of the α -globin locus in ES (left) and erythroid (right) cells, obtained from the polymer model. The polymer color scheme shown below is chosen according to the erythroid TAD boundaries (Oudelaar et al., 2018). Pearson correlation coefficients with Hi-C data are $r=0.94$ for ES and $r=0.97$ for erythroid. The HiCRep SCC coefficient in ES is $SCC=0.63$ and in erythroid $SCC=0.67$. **C)** 3D conformations (full and coarse-grained) of the α -globin locus (20 kb resolution) derived from the SBS models of ES (left) and erythroid (right) cells. In ES, the polymer appears much more intermingled than in erythroid case. Some important key genes are also shown. **D)** Hi-C contact matrices (20 kb resolution) of the β -globin locus in ES (left) and erythroid (right) cells. TADs and gene annotation are shown below. Dashed boxes highlight the erythroid TAD containing the β -globin genes. Genomic coordinates: chr7:109,307,000-112,607,000, mm9. Data from (Oudelaar et al., 2018) (erythroid) and from (Giorgetti et al., 2016) (ES). **E)** Contact matrices (20 kb resolution) of the β -globin locus in ES (left) and erythroid (right) cells, obtained from the polymer model. The polymer color scheme shown below is chosen according to the erythroid TAD boundaries. Pearson correlation coefficients with Hi-C data are $r=0.91$ for ES and $r=0.96$ for erythroid. The HiCRep coefficient in ES is $SCC=0.58$ and in erythroid $SCC=0.45$. **F)** 3D conformations (full and coarse-grained) of the β -globin locus (20 kb resolution) derived from the SBS models of ES (left) and erythroid (right) cells. In ES, the β -globin gene is located in a region poorly defined (colored in cyan) and highly intermingled with the rest of the polymer. In erythroid, such region is highly defined.

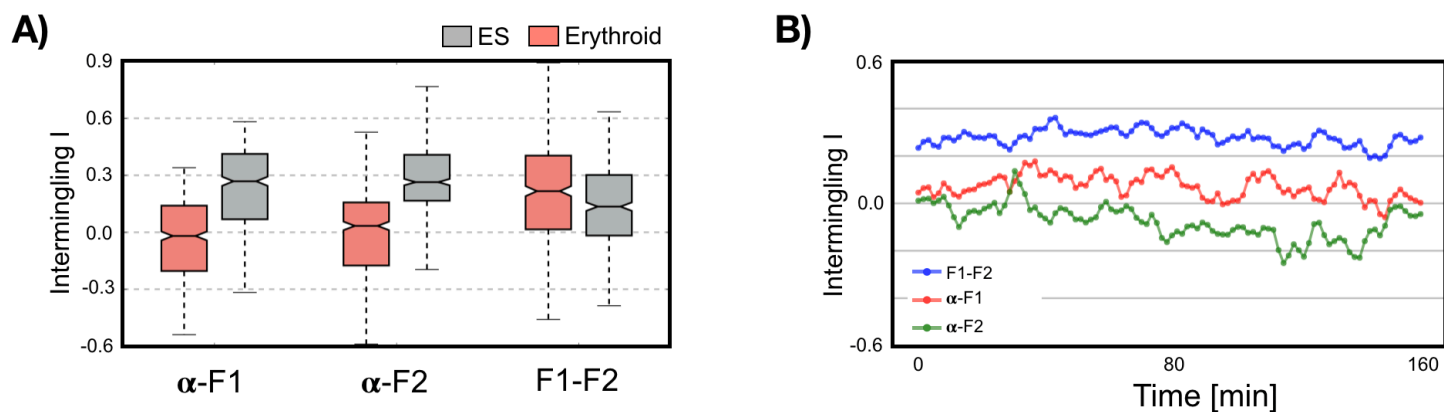


Figure S3 (Related to Figure 2). Analysis of the structural properties reveals heterogeneity and stability in the α -globin higher-order architecture.

A) The full distribution of intermingling among the regions shown in Figure 2A highlights the level of structural heterogeneity in the ensemble of polymer 3D structures. **B)** The dynamics of the Intermingling index I between the regions α -domain, F1 and F2 during a typical Molecular Dynamics run for an erythroid polymer. It exhibits small, thermal fluctuation around its equilibrium value for physical time scales of roughly tens of minutes or hour.

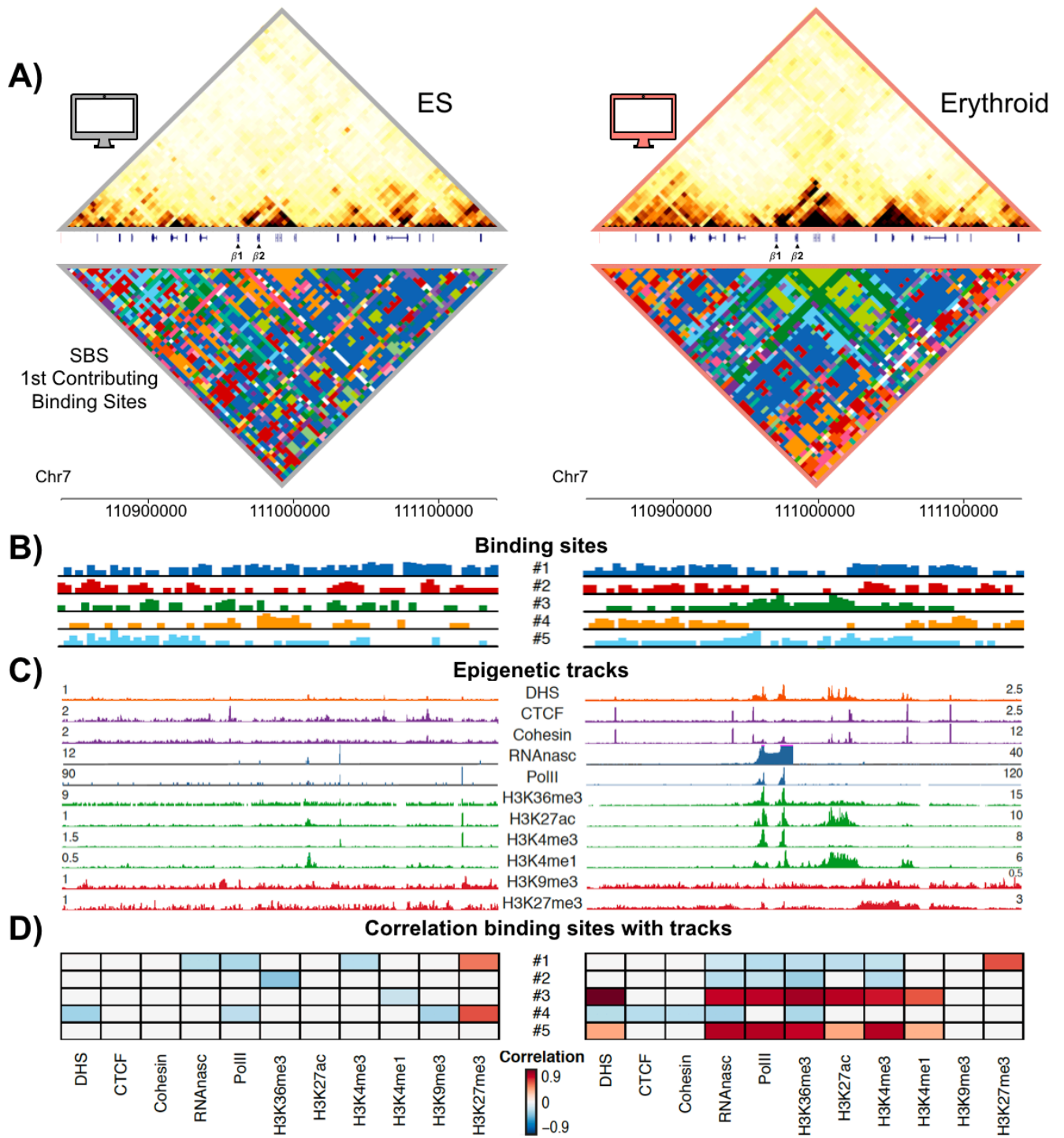


Figure S4 (Related to Figure 4). Binding domains contributing to the 3D structure of the β -globin locus correlate with distinct epigenetic features.

A) Contact map obtained from the 3D polymer model (top matrices). The agreement with experimental Capture-C is high (in ES, Pearson $r=0.95$ and distance corrected Pearson $r'=0.90$, in erythroid, $r=0.96$ and $r'=0.92$). The HiCRep coefficient in ES is $SCC=0.52$ and in erythroid $SCC=0.77$). On the bottom, the most contributing binding sites for each contact is shown. Gene annotations is shown in the middle. **B)** Top five most contributing binding sites to the locus architecture, sorted by their contribution to the overall contact map of the whole locus. In ES, the first most contributing accounts for about 35% of total contacts, while in

erythroid the contacts are associated with more complex set of binding site types. **C)** Epigenetic features of the β -globin locus. DHS = DNaseI Hypersensitive Sites; RNAnasc = nascent RNA expression; PolII = RNA Polymerase II occupancy. **D)** Comparisons of the predicted binding domains to chromatin data show that the dominating binding domain in ES cells is correlated with repressive chromatin marks, while the binding sites associated with the erythroid structure also correlate with chromatin marks of active transcription. Values represent significant Pearson correlation coefficients.

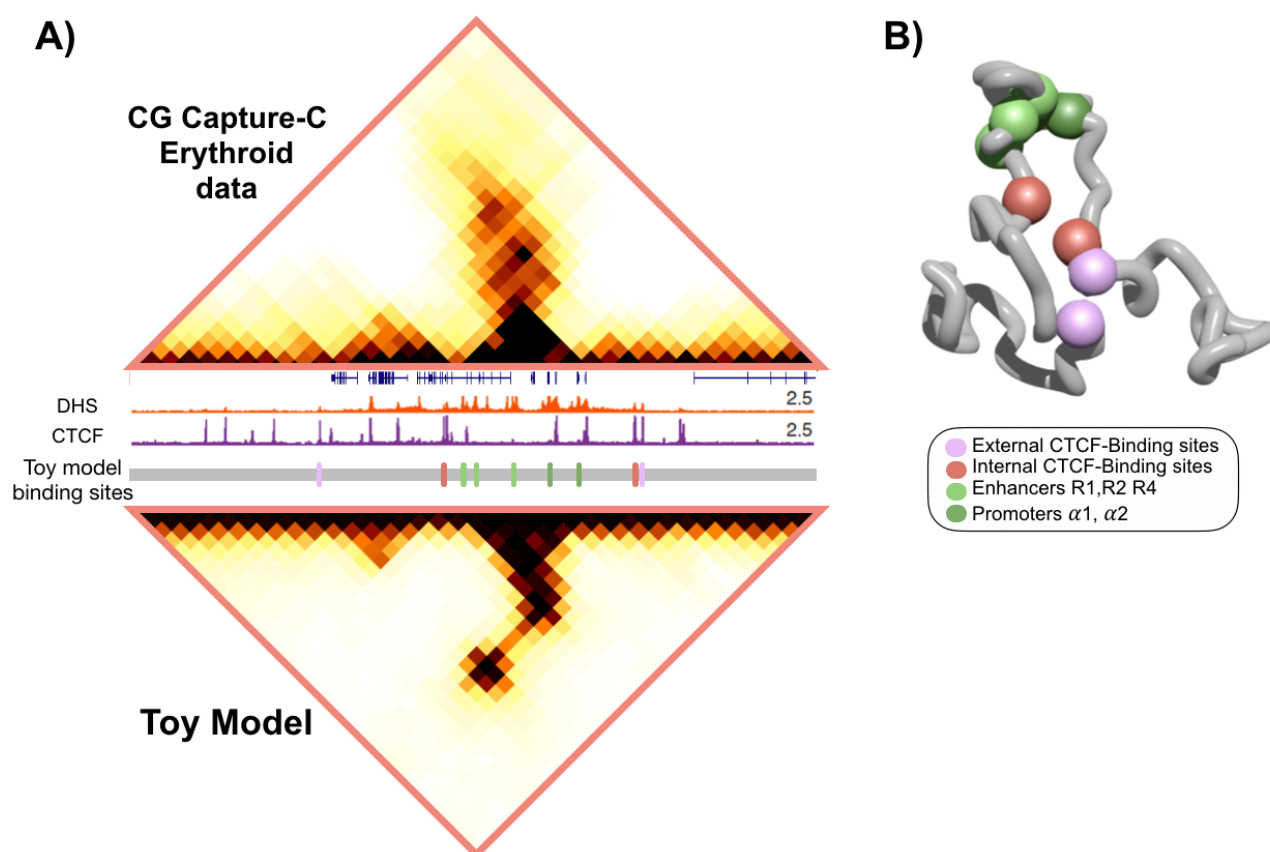


Figure S5 (Related to Figure 5). A simple toy, hairpin-shaped model can help visualising the folded hairpin structure of the α -globin in erythroid data.

A) Coarse-grained (CG) Capture-C data in erythroid cells (top matrix) compared with contact map resulting from the toy model (bottom matrix) with a symmetric arrangement of binding sites corresponding to promoters, enhancers and CTCF sites. **B)** A hairpin-like folded structure obtained from the MD simulation of the toy model.

Tables

Locus	α -globin			
Cell Type	Erythroid		ES	
	# Binding Sites	# Cognate Binders	# Binding Sites	# Cognate Binders
Type 1	104	20	135	27
Type 2	91	18	79	15
Type 3	91	18	78	15
Type 4	73	14	51	10
Type 5	86	17	52	10
Type 6	60	12	68	13
Type 7	63	12	60	12
Type 8	44	8	71	14
Type 9	54	10	38	7
Type 10	52	10	59	11
Type 11	62	12	55	11
Type 12	60	12	49	9
Type 13	54	10	43	8
Type 14	46	9	49	9
Type 15	51	10	42	8
Type 16	47	9	36	7
Inert sites	162	-	235	-
Total number of beads	1200		1200	
Total number of binders		201		186

Table S1 (Related to Figure 4). Details of the simulated polymer model from Capture-C data, for the α -globin in ES and erythroid. For each binding site, the number of sites along the chain and the number of cognate binders is reported. The distributions along the chain of the first 5 types are shown in Figure 4B.

Locus	β -globin			
Cell Type	Erythroid		ES	
	# Binding Sites	# Cognate Binders	# Binding Sites	# Cognate Binders
Type 1	138	27	149	29
Type 2	76	15	88	17

Type 3	106	21	58	11
Type 4	74	14	62	12
Type 5	92	18	73	14
Type 6	74	14	62	12
Type 7	61	12	54	10
Type 8	65	13	69	13
Type 9	45	9	51	10
Type 10	40	8	50	10
Type 11	57	11	67	13
Type 12	50	10	51	10
Type 13	50	10	41	8
Type 14	50	10	41	8
Type 15	34	6	33	6
Type 16	31	6	39	7
Inert sites	157	-	212	-
Total number of beads	1200		1200	
Total number of binders		204		190

Table S2 (Related to Figure S4). Details of the simulated of the polymer model from Capture-C data, for the β -globin in ES and erythroid. For each binding site, the number of sites along the chain and the number of cognate binders is reported. The distributions along the chain of the first 5 are shown in Figure S4B.

SBS Polymer simulation details			
Parameter	Value	Parameter	Value
Bead diameter	$\sigma=1$	Binder diameter	$\sigma=1$
LJ repulsive ($K_B T$ units)	$\epsilon=1$	LJ attractive ($K_B T$ units)	$\epsilon=12$
FENE length constant (σ)	$R_0=1.6$	Bead-Binder interaction range (σ units)	$R_{int}=1.5$
Stiffness (kb units)	4 (Capture-C model)	Bead-Binder Interaction intensity ($K_B T$ units)	$E_{int}=8.16$
	20 (HiC model)		
FENE Spring constant ($K_B T/\sigma^2$ units)	K=30		

Table S3 (Related to STAR Methods). Summary of the Molecular Dynamics parameters employed to simulate the SBS model.