The role of conformational selection in the molecular recognition of the wild type and mutants XPA$_{67-80}$ peptides by ERCC1

Elisa Fadda*

Department of Chemistry, Maynooth University, Maynooth, Kildare, Ireland

ABSTRACT

Molecular recognition is a fundamental step in the coordination of biomolecular pathways. Understanding how recognition and binding occur between highly flexible protein domains is a complex task. The conformational selection theory provides an elegant rationalization of the recognition mechanism, especially valid in cases when unstructured protein regions are involved. The recognition of a poorly structured peptide, namely XPA$_{67-80}$, by its target receptor ERCC1, falls in this challenging study category. The microsecond molecular dynamics (MD) simulations, discussed in this work, show that the conformational propensity of the wild type XPA$_{67-80}$ peptide in solution supports conformational selection as the key mechanism driving its molecular recognition by ERCC1. Moreover, all the mutations of the XPA$_{67-80}$ peptide studied here cause a significant increase of its conformational disorder, relative to the wild type. Comparison to experimental data suggests that the loss of the recognized structural motifs at the microscopic time scale can contribute to the critical decrease in binding observed for one of the mutants, further substantiating the key role of conformational selection in recognition. Ultimately, because of the high sequence identity and analogy in binding, it is conceivable that the conclusions of this study on the XPA$_{67-80}$ peptide also apply to the ERCC1-binding domain of the XPA protein.

INTRODUCTION

The life of a cell depends on well-orchestrated molecular interactions that facilitate signaling through complex networks. Molecular recognition is the key first step for the realization of these interactions.[1] Finding which molecular determinants drive biomolecular recognition in the crowded, dynamic and heterogeneous cell environment is difficult. The idea that molecular recognition may hinge on shape complementarity between ligand and receptor was one of the first put forth[2] and, despite its simplicity, one of the most successful in explaining molecular interaction and binding in modern drug design.[3] The realization that conformational dynamics may play also a key role in recognition, forced a review of this simple theory, known as *lock and key* theory,[4] in favor of the *induced fit* theory,[5] which takes into account a conformational re-adjustment of the receptor upon ligand binding. The *induced fit* theory represents the rationale behind flexible docking methods[6,7] and it is predictive on its own when both ligand and receptor do not undergo large conformational changes.[8–10] Nevertheless, it is not fit to describe recognition between flexible or poorly structured proteins, where conformational dynamics is central to the protein structure and function.[11–13] Indeed, flexible proteins exist as an ensemble of different conformational substrates at equilibrium[14,15]; the relative populations of these substrates can be linked to the degree of conformational disorder of the protein.[16] The *conformational selection* theory[12,17,18] suggests that only one of the weakly populated substrates of a flexible protein is recognized by

the receptor. This event triggers a shift in the conformational equilibrium ensemble.[12,19] Recent work[13,18–21] reveals that the complex dynamics involved in ligand recognition and binding can be explained through a mechanism where both, conformational selection and induced fit play a role. Kinetic studies suggest that conformational selection is the dominant recognition mechanism,[22,23] especially when inherent structural flexibility allows for fast conformational transitions relative to the timescale of binding events.[22] Following recognition, an induced fit stage drives the structural and energetic mutual adjustment, which leads to the final complex.[13,18,19] Molecular dynamics (MD) simulation studies have also highlighted the possibility of a more complex scenario, where interactions at the binding interface, specifically triggered and stabilized by the binding partner, play a role in recognition and binding.[20] In this work I use MD simulations to investigate the molecular basis for the recognition and binding of a series of 14 residue peptides, namely the wild type XPA_{67-80} peptide and 4 mutants, to their protein target ERCC1. The sequence of the XPA_{67-80} peptide corresponds to the ERCC1 binding motif of the XPA protein.[24,25]

Xeroderma Pigmentosum complementation group A (XPA) and Excision Repair Cross Complementation group 1 (ERCC1) are essential proteins in nucleotide excision repair (NER), the principal pathway responsible for the repair of DNA bulky adducts, such as the ones caused by alkylation agents and radiation.[26] ERCC1 forms with XPF (also known as ERCC4) an endonuclease responsible for the excision on the 5' side of the lesion. One of the roles of XPA in the NER pathway is to recruit the ERCC1-XPF endonuclease and to position it correctly on the DNA strand for excision.[27–29] The ERCC1-binding domain of H. sapiens XPA is located between residues 59 and 114.[25] Within this region, the essential residues for binding are only 14, namely residue 67 to 80.[24,25] Experimental evidence[24] shows that a peptide with the same 14 residues sequence, named XPA_{67-80}, binds ERCC1 and inhibits the NER reaction in vitro. Additionally, an earlier MD conformational study[30] of the wild type XPA_{67-80} peptide both in solution and in complex with ERCC1, shows, in agreement with NMR data,[24] that the unbound XPA_{67-80} peptide has a high degree of disorder and that, within the microscopic timescales sampled, this disorder is determined by the dynamic interchange between hairpin-like structures.[30] As shown in Figure 1, panel a, the most populated hairpin conformation in solution is structurally similar to the conformation of the bound XPA_{67-80} peptide.[30]

The XPA section between residues 67 and 80 is highly conserved across many different species carrying the NER genes.[24,25] A number of highly conserved XPA_{67-80} residues interact directly with ERCC1 residues in the complex,[24,25,29–31] namely Asp 70, Thr 71, Phe 75, and Ile 76. Also, structural studies[24,29,30] suggest that

Gly 72 to 74 and Leu 77 are highly conserved because of their steric and hydrophobic complementarity to the narrow ERCC1 binding site.[24,30] The last three residues at each tail of the XPA_{67-80} hairpin, namely Glu 78 to 80 on one end, and Lys 67, Ile 68 and Ile 69 on the other end, are located outside of the ERCC1 binding pocket. Li et al.[25] have shown that deletion of the segment between Glu 78 to 84, named E motif, reduces binding of the XPA protein to ERCC1 in vitro by approximately 70%. As neither Glu 78 to 84, nor Lys 67 and Ile 68, located on the opposite end of the peptide, interact with ERCC1,[24,30] their role in ERCC1 of XPA binding is not immediately clear. In this work I address how specific XPA_{67-80} residues affect the conformational propensity of the peptide in solution and if this effect correlates with available mutagenic and binding data.[24,25,29] This is an important point as, on the basis of conformational selection, a significant shift in the XPA_{67-80} peptide conformational propensity in solution should affect its recognition and binding to ERCC1. The extensive MD simulations analyzed in this work provide a comparison of the conformational propensity of four different mutants, with sequences shown in Figure 2, against the wild type XPA_{67-80} peptide. More specifically, I analyzed the dynamics of 2 triple mutants, named M1 and M2, designed to understand the effect of terminal charged residues on the dynamics of the peptide, and of 2 single point mutants, named F75W and I68K, designed to enhance binding affinity relative to the wild type. The highest populated conformers determined through clustering analysis were tested for their ability to spatially fit in the ERCC1 binding site through structural alignment to the XPA_{67-80} peptide from the NMR structure of the complex.[24] Moreover, the stability of selected potential recognition complexes were analyzed through extensive MD simulations and compared the structure and interactions to the NMR complex.[24]

The MD results suggest that the recognition of the wild type XPA_{67-80} peptide by ERCC1 can be explained through conformational selection. Furthermore, the data show that the mutations studied enhance the conformational disorder of the XPA_{67-80} peptide in solution, significantly lowering the population of the potentially recognized conformers relative to the wild type. While the population of the recognized conformer in the wild type XPA_{67-80} peptide in solution is of 36%, the populations of the recognized conformers of M2, F75W and I68K drop to single digits, while none of the significantly populated M1 conformers satisfies the structural constraints set for recognition. These results also suggest that the enhanced conformational disorder relative to wild type caused by the deletion of the E motif in XPA could be one of the contributing factors for the experimentally determined 70% decrease in binding to ERCC1.[25] The analysis of the structural stability and dynamics of the potential recognition complexes leaves
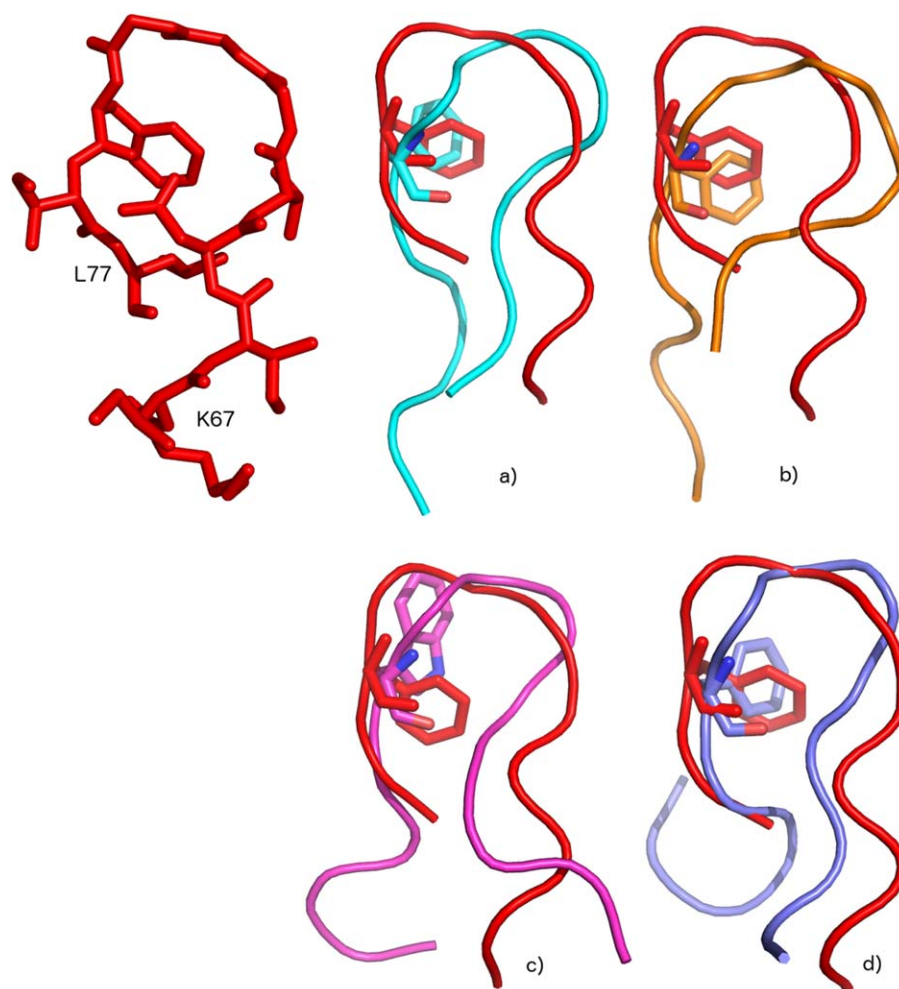
Figure 1

The NMR structure of the bound XPA$_{67-80}$ peptide (PDBid 2JNW) is shown in red on the top left hand side. The location of Lys 67 and Leu 77 are shown to help define the spatial orientation of the peptide. The positions of the terminal Glu 78 to 80 do not appear in the NMR structure. The structural alignment onto the NMR bound conformation of the wild type XPA$_{67-80}$ and of the mutants conformers, identified as positive candidates for recognition, are shown in the panels. The wild type XPA$_{67-80}$ peptide is shown in cyan in panel a), the M2 conformer in orange in panel b), the F75W conformer in magenta in panel c), and the I68K conformer in light blue in panel d). All conformers are rendered with cartoons with the Phe 75, Trp 75 in the F75W mutant, shown with sticks for clarity.

also room for an induced fit stage, following conformational selection and binding, in the realization of the final complex.[1]

METHODS

The conformational dynamics of 4, 14-residues long XPA$_{67-80}$ mutant peptides was studied by MD. The simulations and the clustering analysis were run with the GROMACS simulation package,[32] version 4.6.3. All peptides were built in a fully extended conformation with the academic version 9.7 (release 2014-1) of Schrödinger Maestro[33] with sequences shown in Figure 2. The peptides termini were capped with acetyl (ACE) and

N-methylamide (NME) residues. For each peptide, I selected 10 snapshots from an initial 50 ns trajectory. These snapshots were used as starting conformations for 10 independent (1 μs) MD simulations, for a total sampling time of 10 μs for each peptide. Further details on the MD set-up protocols can be found in the Supporting Information section. The AMBER99SB-ILDN force field[34] was chosen to represent the peptide and counterions in all MD simulations, while water molecules were represented by the TIP4P-Ew[35] potential. For the MD simulations of the wild type XPA$_{67-80}$ peptide[30] I used in this work as reference, the AMBER99SB force field[36] and the TIP3P water model[37] were used instead. To verify the consistency of the results obtained with the two different force fields, I ran the first 500 ns of each one of

	67	68	69	70	71	72	73	74	75	76	77	78	79	80
H. sapiens	K	I	I	D	T	G	G	G	F	I	L	E	E	E
M1	A	A	A	D	T	G	G	G	F	I	L	E	E	E
M2	K	I	I	D	T	G	G	G	F	I	L	A	A	A
I69K	K	K	I	D	T	G	G	G	F	I	L	E	E	E
F75W	K	I	I	D	T	G	G	G	W	I	L	E	E	E

Figure 2

The *H. sapiens* XPA ERCC1-binding motif sequence (NP_00371.1), shown in the first row, corresponds to the wild type XPA$_{67-80}$ peptide sequence. The XPA$_{67-80}$ mutant sequences are shown below. The mutations are highlighted in red for clarity. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the 10 independent trajectories of the M1 XPA$_{67-80}$ peptide also with the AMBER99SB force field in TIP3P water. The results in Supporting Information Table SI show that the two protocols produce highly consistent results. During the MD simulations the temperature was held constant at 300 K by a Langevin thermostat[38] with coupling time constant of 0.1 ps. The Berendsen barostat[39] was used to hold the pressure constant at 1 bar, with a time constant of 0.5 ps. The equations of motion were integrated using a leap-frog stochastic dynamics integrator with a 2 fs timestep. The linear constraint solver (LINCS) was used to constrain all bonds with hydrogen atoms.[40] Long range electrostatics were treated with the Particle Mesh Ewald (PME) method.[41,42] The maximum spacing for the Fast Fourier Transform (FFT) grid was chosen as 1 Å. In all simulations cutoff values for Coulomb were set to 12 Å, while van der Waals interactions were switched off between 10 and 11 Å.

Structure clustering was performed with the GROMACS tool *g_cluster*. Clusters were identified by means of the GROMOS algorithm[43] with a RMSD cutoff value of 1.5 Å as discussed in previous work.[30] Models of the recognition complexes were built by structural alignment of 10 of the highest populated conformers, one from each of the 10 (1 μs) simulations, using the 1st NMR structure of the XPA$_{67-80}$/ERCC1 complex (PDBid 2JNW) as a reference. The structural alignment was obtained based on a 10-point least square fit calculated with version 1.4.1 of the PyMOL Molecular Graphic System.[44] To find alternative conformations for potential recognition complexes, I also tried a series of rigid docking experiments with version 6.6 of Dock,[45] ClusPro[46,47] and PyDock.[48] All docking programs were initially tested on a positive control experiment, with the NMR structure of the XPA$_{67-80}$/ERCC1 complex (PDBid 2JNW) as starting coordinates. All docking methods were able to locate successfully the ERCC1 binding site, but failed to predict the correct pose. In fact, the highest scored poses almost universally presented the XPA$_{67-80}$ hairpin at 180° rotation, around the axis perpendicular to the hairpin plane, relative to the NMR bound conformation.

All recognition complexes built by structural alignment were visually inspected with the Pymol graphic user interface[44] for steric clashes and for major differences in the orientation of the peptide sidechains relative to the NMR structure.[24] The peptide conformations of all the mutants were ranked based on the backbone RMSD values relative to the bound[24] and solution XPA$_{67-80}$ conformations.[30] The peptides conformations that successfully passed this structural analysis were classified as potentially recognized structures. The recognition complexes of the wild type and F75W mutant XPA$_{67-80}$ with ERCC1 were further analyzed through 1 μs MD simulation each. Details on the set-up protocol used for the MD simulations of the recognition complexes can be found in the Supporting Information section.

RESULTS

In this section I present the results of the conformational analysis of the wild type and mutants XPA$_{67-80}$ peptides free in solution, together with the MD simulation of models of recognition complexes between the wild type and F75W XPA$_{67-80}$, and ERCC1. The wild type XPA$_{67-80}$ conformational dynamics discussed in earlier work[30] is analyzed here within the framework of the mutants dynamics, to aid the interpretation of the data and to make the work self-contained.

Wild type XPA67-80 peptide

In agreement with NMR data,[24] MD simulations have shown that the unbound XPA$_{67-80}$ peptide is poorly structured, but also that it has a high propensity for the formation of hairpins, stable at the microsecond timescale.[30] The most stable hairpins are held together by 3 hydrogen bonds between backbone atoms and are structurally similar to the bound XPA$_{67-80}$ peptide conformation from NMR in terms of the location of the residues within the hairpin structure, see Figure 1, panel a. More specifically, with reference to the *H. sapiens* numbering shown in Figure 2, Gly 72 to 74 form the hairpin turn, while Glu 78 to 80 are located on one tail and Lys 67, Ile 68 and Ile 69 on the other tail. The results of the

Table I

Average radius of gyration (Rg), total and average number of clusters, average population of the most populated clusters, namely cluster 1, 2, and 3

Peptide	Rg (Å)	Tot. No. clusters	Ave. No. clusters	Cluster 1 Pop. (%)	Cluster 2 Pop. (%)	Cluster 3 Pop. (%)
Wild type	*6.7*	*1145*	*115*	*46.9*	*9.9*	*5.0*
M1	8.6	4357	436	11.4	4.9	3.0
M2	7.5	2656	266	23.0	8.4	3.9
F128W	7.8	2035	204	31.7	5.2	3.8
I121K	7.6	2332	233	23.8	7.4	5.1

All values are obtained from the analysis of 10 (1 μs) independent MD trajectories. The RMSD threshold used for the clustering analysis is 1.5 Å.

clustering analysis of the 10 (1 μs long) independent MD simulations, summarized in Table I, show a high recurrence of a specific hairpin structure, shown in Figure 1, panel a. The combined analysis of the clustering data from earlier work[30] shows that the XPA$_{67-80}$ peptide free in solution is found in this hairpin conformation 36% of the time throughout 10 μs, see Table II. As shown in Supporting Information Table SII, this hairpin is the most populated conformer during 7 out of 10 μs. To assess how well this hairpin motif fits in the ERCC1 binding site, I ran a structural alignment of the representative (middle) structures from the highest populated clusters of the XPA$_{67-80}$ peptide in solution onto the XPA$_{67-80}$ bound structure from NMR.[24] The structural alignment, shown in Figure 3, scores an average backbone RMSD value of 2.2 Å, relative to the bound peptide conformation, see Supporting Information Table SII, and results in no steric clashes with binding site residues. As shown in Table II, I classified this hairpin as a positive recognition candidate. The 2nd most populated conformer of the wild type XPA$_{67-80}$ peptide unbound in solution is present for 7% of the time throughout the 10 μs, see Table II. This hairpin corresponds to the highest populated cluster identified during 2 μs, more specifically during simulation (s)2 and s8, see Supporting Information Table SII. Based on structural alignment, this hairpin does not fit the ERCC1 binding site, mainly due to a different orientation of the Phe 75 sidechain, see Supporting Information Figure S1, panel a. For the same reasons, as shown in Supporting Information Figure S1, panel b, also the 3rd most populated conformer, with a

relative population of 2% over 10 μs, does not fit the ERCC1 binding site. This latter conformer corresponds to the highest populated cluster from s9.[30]

M1 XPA$_{67-80}$ peptide

The M1 peptide was designed to understand the role of the Lys 67 to Ile 69 segment in the dynamics of the XPA$_{67-80}$ peptide. While the Ile 69 sidechain is found to bind in the hydrophobic pocket between Arg 144 and Leu 148 of ERCC1,[24,30] Lys 67 and Ile 68 do not form any stable contacts with ERCC1 residues. Nevertheless, the presence of a basic residue in the same position of Lys 67 in particular is highly recurrent among all different species carrying NER genes. As shown in Figure 2, in the M1 peptide Lys 67 and Ile 68 and 69 are mutated to Ala. The clustering analysis in Table I shows that 4357 clusters were identified during the MD simulations of the M1 peptide, almost four times the number of clusters identified for the wild type XPA$_{67-80}$ peptide.[30] Also, the significant increment in the radius of gyration (R_g) is indicative of the higher propensity for a more elongated, or extended, peptide conformation relative to the wild type. This conformational disorder is also reflected in the average cluster populations in Table I, where the highest populated clusters has an average population four times lower than the highest populated clusters in the wild type XPA$_{67-80}$ peptide. The highest populated conformer, identified during s1, see Supporting Information Table SIII, is present only 3.8% of the time during 10 μs. As shown in Supporting Information Figure S2, structural

Table II

Conformational propensity of the different peptides identified through the combined analysis of the 10 highest populated clusters from of the 10 (1 μs -long) independent MD trajectories

Peptide	Conf. 1 (%)	Rec	Conf. 2 (%)	Rec	Conf. 3 (%)	Rec
Wild type	*36.4*	*Yes*	*6.6*	*No*	*2.3*	*No*
M1	3.8	No	1.8	No	1.3	No
M2	7.2	No	5.4	Yes	5.0	No
F128W	9.4	No	8.7	No	6.6	No
I121K	11.7	No	4.3	Yes	2.6	No

The yes/no values in the columns titled "Rec" indicate the likelihood for a specific conformation to be recognized. This likelihood has been determined based on the conformer structural similarity to the bound and solution conformations of XPA$_{67-80}$, and on its fit into the ERCC1 binding site.
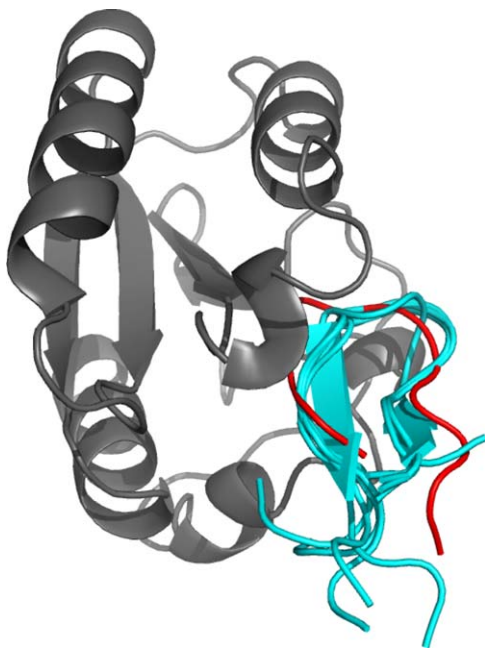
Figure 3

Structural alignment of the representative (middle) conformers of the highest populated clusters identified through the 10 μs MD simulation of the wild type XPA$_{67-80}$ in solution, shown in cyan, onto the bound XPA$_{67-80}$, shown in red (PDBid 2JNW). The ERCC1 central domain is rendered in grey cartoon. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

alignment on to the NMR peptide reveals major clashes with the ERCC1 binding site, suggesting that this hairpin is not recognized. As shown in Table II, the 2nd and 3rd most populated clusters have relative populations of 1.8% and of 1.3%, respectively. Structural alignment of these conformers on to the NMR bound peptide also reveals major clashes with the ERCC1 binding site, suggesting that none of the three highest populated clusters are recognized. All other conformers are formed for less than 1% of the time and their contribution is considered negligible.

M2 XPA$_{67-80}$ peptide

In the M2 mutant Glu 78 to 80 of XPA$_{67-80}$ have been replaced by Ala. As shown in Table I, this mutation also enhances the degree of disorder and the peptide extension relative to the wild type. Though, based on the average number of clusters, the mutation of the three negatively charged terminal residues seems to have a lower impact on the stability of conformers relative to the M1 mutation. This is probably due to the relative stability of the highest populated hairpin-like structure, see Supporting Information Table SIV, stabilized by the hydrophobic association between the sidechains of Lys 67 and Ile 68 with the Ala substituting the negatively charged Glu 77 to 80, which results twisting the peptide tails. Nevertheless,

as shown in Table II, the relative populations of the highest populated conformers are quite low, occurring for less than 10% of the time over 10 μs. Structural alignment of the middle structures on the NMR peptide shows that only the 2nd most populated cluster, identified during s3 and s7, see Supporting Information Table SIV, does not clash with residues in the ERCC1 binding site, see Figure 4. The backbone RMSD value relative to the peptide in the NMR complex is 3.1 Å, see Supporting Information Table SIV, while structural alignment to the wild type XPA$_{67-80}$ highest populated conformer results in a backbone RMSD value 2.1 Å, see Figure 1, panel b. Based on these data, the 2nd most populated conformer of the M2 mutant, with a relative population of 5.4% over 10 μs, is a positive candidate for recognition.

F75W XPA$_{67-80}$ peptide

The mutation of the highly conserved Phe75 to Trp was designed to increase the binding affinity of the peptide by enhancing the stacking interaction with Asn110 of ERCC1.[24,30] However, as shown in Table I, this mutation also leads to a higher conformational disorder of the peptide relative to the wild type. This is due to the occurrence of a set of hairpin motifs stabilized by the stacking of the Trp 75 sidechain onto the hydrophobic patch formed between Lys 67 and Ile 69, see Supporting Information Figure S3, panel a. These hairpins correspond to the highest populated cluster, with a 9.4% relative population over 10 μs, see Tables II and Supporting Information Table SIV. The 2nd highest populated conformers are stabilized by a similar interaction, shown Supporting Information Figure S3, panel b, where the Trp sidechain is stacked perpendicularly between the sidechains of Lys 67 and Ile 69. As shown in Table II, none of the 3 highest populated conformers formed by the F75W peptide can be accommodated in the binding site. This is essentially due to the orientation of the Trp sidechain relative to the Phe75 in the XPA$_{67-80}$ bound conformation. Nevertheless, the 4th most populated F75W hairpin motif, occurring 3.6% of the time over the 10 μs simulation, has the correct conformation to fit the ERCC1 binding site, with a backbone RMSD value against the NMR structure of the XPA$_{67-80}$ wild type peptide of 1.8 Å, see Supporting Information Table SV and Figure 1, panel c. The backbone RMSD value of this F75W conformer against the highest populated hairpin formed by the wild type peptide in solution is 3.4 Å. Structural alignment onto the bound peptide in the complex results in a good fit, with the Trp sidechain orientated correctly to stack with Asn 110 of ERCC1.[24,30]

I68K XPA$_{67-80}$ peptide

As the F75W, the I68K mutant was also designed to increase the peptide binding affinity for ERCC1, with a
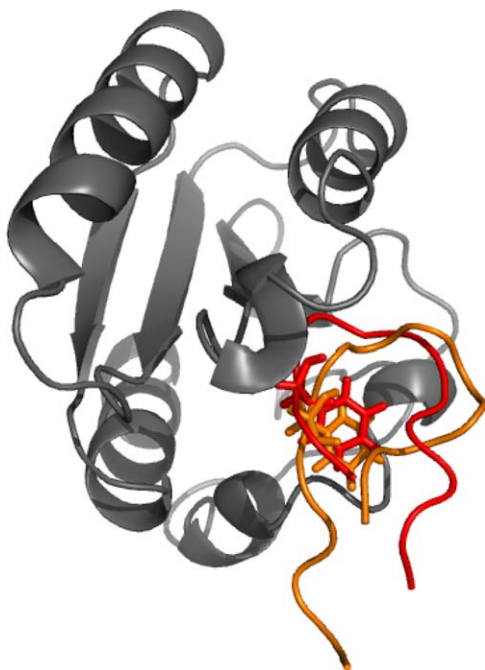
Figure 4

Structural alignment of the representative (middle) conformer of the potentially recognized cluster identified through the 10 μs MD simulation of the M2 XPA67-80 mutant in solution, shown in orange, onto the bound XPA67-80, shown in red (PDBid 2JNW). The ERCC1 central domain is rendered in grey cartoon. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

design strategy based on the idea that the presence of an extra positive charge on the N-terminal tail could enhance the stability of the recognized hairpin due to electrostatic interaction with the negatively charged C-terminal tail, see Figure 2. Nevertheless, similarly to the previous cases, as shown in Table I, this single point mutation also increases the conformational disorder and the extension of the peptide relative to the wild type. The increased number of clusters relative to the wild type peptide derives again from the stabilization of a wider set of hairpin motifs relative to the wild type XPA67-80. Indeed, the highest populated conformer is stabilized by the interaction between the Phe 75 sidechain and the hydrophobic part of the two sidechains of Lys 67 and Lys 68, stabilizing the Phe 75 sidechain in the wrong orientation for an interaction with the ERCC1.[24] Meanwhile, the 2nd highest populated conformer, shown in Figure 1, panel d, with relative population of 4.3% over 10 μs simulation, see Table II, has a suitable structure for recognition. The backbone RMSD values calculated relative to the bound XPA67-80 peptide, and relative to the wild type XPA67-80 peptide highest populated conformer, are 1.8 Å and 3.7 Å, respectively, see Supporting Information Table SVI. The 3rd highest populated conformer is a distorted hairpin motif, stabilized by an interaction

between the Phe 75 sidechain and the sidechain of Ile 76, and it does not structurally fit the ERCC1 binding site.

Wild type XPA67-80 peptide in a recognition complex model

The conformational stability of a potential recognition complex, obtained through structural alignment of the highest populated conformer of the wild type XPA67-80 peptide on the XPA67-80 peptide bound structure,[24] was studied through 1 μs MD simulation, during which the peptide was left completely unrestrained. As shown in Figure 5, the XPA67-80 hairpin structure is stable for most of the trajectory, held together by three intra-strand hydrogen bonds between backbone atoms. At around 650 ns the terminal hydrogen bond, between the backbone of Glu 78 and Ile 68, breaks causing a change in RMSD value. As also shown in Figure 5, during the 1 μs trajectory we did not observe any significant relaxation of the structure towards the XPA67-80 bound conformation, with an average backbone RMSD value between the two peptides of 2.8 Å. Most of the conserved interactions identified between the bound peptide and the ERCC1 binding site residues in the NMR structure are retained. More specifically, as shown in Table III, the hydrogen bond pattern between Gly 73 and 74 at the hairpin bend is similar to the one seen in the NMR complex structure, with the only difference that the carbonyl of Gly 74 interacts with Ser 142 hydroxyl group instead of the amide backbone. As in the NMR complex, Tyr 145 and 152 stack against the Gly-rich hairpin bend stabilizing its position, while Phe 75 is firmly stacked against the Asn 110 sidechain. One of the main differences between the two complexes is that the hydrophobic pocket formed between Arg 144 and Leu 148 of ERCC1 in the potential recognition complex is occupied by the Ile 69 sidechain, instead of Ile 76.

F75W XPA67-80 peptide in a recognition complex model

A model for the recognition complex between the 4th highest populated conformer of the F75W XPA67-80 peptide mutant and ERCC1 was obtained by structural alignment of the peptide solution structure onto the bound peptide NMR structure, see Figure 6. The conformational stability of this complex was assessed during a 1 μs MD trajectory, where the ligand was left completely unrestrained. As shown in Figure 7, as in the case of the wild type XPA67-80 recognition complex, the structure of the F75W peptide does not completely relax to match the NMR bound structure over the 1 μs sampling timeframe. Nevertheless, the complex is stable and, as shown in Table IV, most of the hydrogen bonding and stacking interactions observed in the NMR complex are retained. Interestingly, from the structural point of view Trp 75 is quite effective in stacking Asn 110. As seen in the NMR
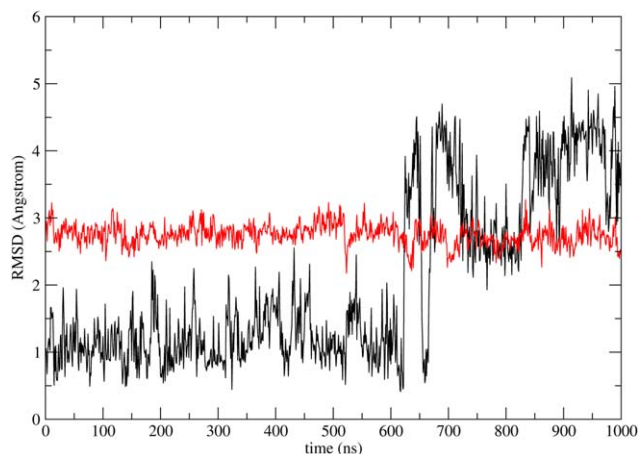
Figure 5

The backbone RMSD value calculated through the 10 μs trajectory of the wild type XPA$_{67\text{-}80}$ peptide in the model recognition complex is shown in black. The backbone RMSD value relative to the bound XPA$_{67\text{-}80}$ peptide conformation (PDBid 2JNW) is shown in red. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

structure the Ile 76 sidechain is well accommodated in the hydrophobic pocket formed between the sidechains of Arg 144 and Leu 148, the amide backbone group of Ser 142 forms a hydrogen bond with the carbonyl of Gly 74 and the hydroxyl group of the Tyr 145 sidechain forms a hydrogen bond with the carbonyl of Thr 71. In addition, as observed in the MD simulation of the wild type XPA$_{67\text{-}80}$ complex,[30] the Asp 70 forms a stable hydrogen bond with the Nε2 of His 149.

DISCUSSION

In this study I analyzed the conformational propensity in the μs timescale of 4 mutants of the XPA$_{67\text{-}80}$ peptide free in solution by means of MD simulations. The results are here compared to MD data obtained for the wild type XPA$_{67\text{-}80}$ peptide from previous work[30] and also to structural[24,29] and binding data[24,25] available. The MD data clearly suggests that the wild type XPA$_{67\text{-}80}$ peptide,

Table III

Average hydrogen bonding distances (Å) measured throughout the 1 μs MD trajectory of the potential recognition complex between the wild type (wt) XPA$_{67\text{-}80}$ peptide solution structure and ERCC1

Interacting residues and atoms		wt XPA$_{67\text{-}80}$/ERCC1	2JNW
Ser 142 (NH)	Gly 74 (O)	5.4	2.6
Ser 142 (OH)	Gly 74 (O)	2.9	5.0
Gln 107 (NH$_2$)	Gly 73 (O)	5.3	4.4
Asn 110 (CB)	Phe 75 (CG)	4.2	4.7
Tyr 145 (OH)	Thr 71 (O)	3.8	3.5
Tyr 145 (OH)	Thr 71 (OH)	5.3	6.4

The corresponding values from the NMR XPA$_{67\text{-}80}$/ERCC1 complex (PDBid 2JNW) are shown for comparison.
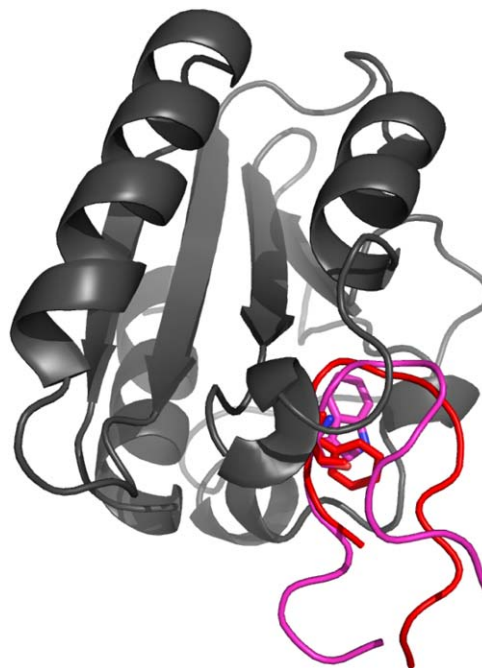

Figure 6

Structural alignment of the representative (middle) conformer of the potentially recognized cluster identified through the 10 μs MD simulation of the F75W XPA$_{67\text{-}80}$ mutant in solution, shown in purple, onto the bound XPA$_{67\text{-}80}$, shown in red (PDBid 2JNW). The ERCC1 central domain is rendered in grey cartoon. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

corresponding to the ERCC1 binding motif of XPA,[24,25] is poorly structured, adopting a range of hairpin-like motifs, stable within the μs time scale.[30] The dominant conformation, with a relative population of 36.4%, is a hairpin structurally similar to the XPA$_{67\text{-}80}$ peptide bound conformation,[24] held together by 3 hydrogen bonds between backbone atoms and characterized by the Phe 75 sidechain prevalently oriented underneath the hairpin loop.[30] As shown in Figure 1, panel a, the 3 hydrogen bonds constrain this XPA$_{67\text{-}80}$ solution structure to a narrow shape, where the tails are closer together than in the bound conformation, which notably has no intra-chain hydrogen bonds.[24,30] The conformational propensity of the wild type XPA$_{67\text{-}80}$ peptide in solution reveals that there is in fact a detectable structural order at the microscopic timescales, which could hold the key to understand its recognition and binding to ERCC1. Indeed, the results of this study suggest that conformational selection and its proposed role in driving recognition,[12,13,17,18,23,49–51] especially in flexible or poorly structured proteins,[18,50,51] can explain the initial recognition step in the XPA$_{67\text{-}80}$ peptide binding by ERCC1.

My working hypothesis is that the highest populated conformation of the wild type XPA$_{67\text{-}80}$ peptide in solution corresponds to a structure recognized by ERCC1. I
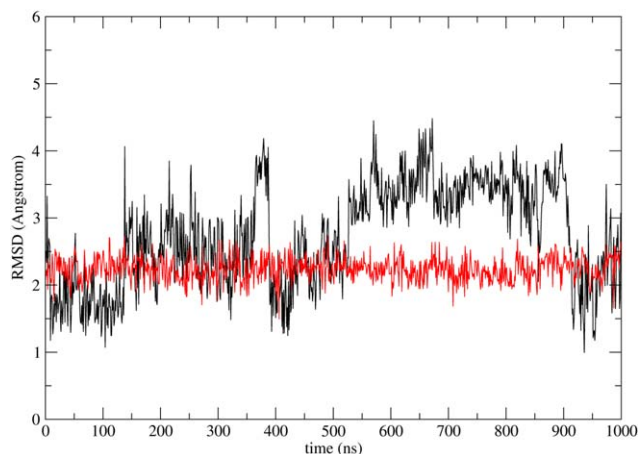
Figure 7
The backbone RMSD value calculated through the 10 μs trajectory of the F75W XPA$_{67-80}$ peptide in the model recognition complex is shown in black. The backbone RMSD value relative to the bound XPA$_{67-80}$ peptide conformation (PDBid 2JNW) is shown in red. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

tested this idea by assessing the fit of the wild type peptide solution structure in the ERCC1 binding site through structural alignment to the XPA$_{67-80}$ bound peptide from NMR[24] and by determining the conformational stability and dynamics of the model complexes through a 1 μs MD simulation. This potential recognition complex is stable throughout the trajectory, with most of the conformational dynamics affecting the hairpin tails, while the hairpin bend is firmly tucked in the interior of the narrow ERCC1 binding site. As shown in Table III, many of the same native contacts identified in the NMR structure[24] and monitored during the MD of the bound complex[30] are also retained. As shown in Figure 5, within the 1 μs timescale there are no significant conformational changes driving the XPA$_{67-80}$ solution structure closer to the peptide bound conformation from NMR. Nevertheless, slight structural modifications take place during the 1 μs trajectory that point in that direction, such as the break of the terminal hydrogen bond. Such

Table IV
Average hydrogen bonding distances (Å) measured throughout the 1 μs MD trajectory of the potential recognition complex, between the F75W XPA$_{67-80}$ peptide solution structure and ERCC1

Interacting residues and atoms		F75W XPA$_{67-80}$/ERCC1	2JNW
Ser 142 (NH)	Gly 74 (O)	3.1	2.6
Gln 107 (NH$_2$)	Gly 73 (O)	3.0	4.4
Asn 110 (CB)	Trp 75 (CD2)	4.1	4.7
His 149 (Nε2)	Asp 70 (Ox)	3.8	4.9
Tyr 145 (OH)	Thr 71 (O)	3.4	3.5
Tyr 145 (OH)	Thr 71 (OH)	6.4	6.4

The corresponding values from the NMR wild type XPA$_{67-80}$/ERCC1 complex (PDBid 2JNW) are shown for comparison.

conformational changes are consistent with an induced fit stage.

The relatively high occurrence (36%) of the highest populated wild type XPA$_{67-80}$ conformer in solution makes this hairpin the most likely structure to be recognized within the conformational ensemble. As shown in Table II, the mutation of the three terminal residues on both the C- and N-termini of the peptide to Ala and also the F75W and I68K single point mutations, all cause a significant increase in the peptide disorder, lowering the relative populations of all conformers to single digits. The potentially recognized structures in all mutants have a much lower occurrence than for the wild type. In fact, the relative population of the F75W XPA$_{67-80}$ recognized peptide is 10 fold less the population of the wild type recognized peptide. These data suggest that the increased disorder caused by mutations, such as the Glu 78-80 to Ala in the M2 mutant, leads to lower populations of the recognized conformers in solution. Within the framework of conformational selection, the results obtained for the M2 peptide can explain the 70% less binding observed the ΔE XPA mutant.[25] Based on the similar conformational trends obtained for all mutants and also considering the high level of sequence conservation of the XPA ERCC1-binding motif, I would expect a significant decrease in binding for all the XPA$_{67-80}$ mutants studied here, relative to the wild type.

CONCLUSIONS

Molecular recognition is still one of the big mysteries in chemical and molecular biology. Understanding how molecular recognition and binding occurs between highly flexible protein domains is even more complex. The conformational selection theory provides a very elegant explanation for molecular recognition,[12,17,18,50] especially in the context of seemingly unstructured protein regions. The results of this MD study strongly suggest that the initial step in the recognition of the wild type XPA$_{67-80}$ peptide by ERCC1 takes place through conformational selection. The analysis of the wild type XPA$_{67-80}$ peptide conformational dynamics in solution highlights a microscopic conformational propensity towards the formation of hairpin-like structures, relative to extended structures.[30] The relative populations of the different hairpin conformers suggest that a structural order does exist at the μs timescale. More specifically, a distinct hairpin conformation occurs 36% of the time over the 10 μs simulations, while all other structures have a significantly lower occupation. This high populated hairpin is structurally similar to the ERCC1-bound XPA$_{67-80}$ conformation,[24] it can be accommodated quite well into the ERCC1 receptor binding site in a recognition complex model, and in such complex it forms important and stable interactions with residues in the ERCC1 binding site

previously identified in structural studies.[24,30] These results all suggest that the highest populated hairpin formed by the wild type XPA_{67-80} peptide in solution is a good candidate for recognition by ERCC1. This study also shows that the highly conserved, C- and N-terminal residues of the XPA_{67-80} peptide, which do not interact with ERCC1, do play a significant role in controlling the peptide conformational propensity. In particular, all the mutations studied here cause a significant increase in the conformational disorder of the peptide relative to the wild type. The MD simulations show that this disorder can be caused, as in the case of F75W and I68K, by the stabilization of a higher number of hairpin conformers relative to wild type, or to the higher occurrence of extended structures, as in the case of M1 and M2. Particularly interesting is the increase in conformational disorder observed for the M2 peptide, where the 3 terminal Glu 78-80 are mutated to Ala. For this peptide the only conformer fit for recognition occurs only 3.6% of the time over the 10 µs simulation, 10 fold less populated than the wild type recognized hairpin. These data suggest that the 70% decrease in binding, measured for a XPA mutant with a deleted E-motif[25] (ΔE), can be caused by an increased conformational disorder that affects primarily its recognition by ERCC1. Based on the comparison between the enhanced conformational disorder of the M2 XPA_{67-80} mutant relative to wild type and the experimentally determined binding deficiency of the analogous ΔE XPA mutant, it is expected for all the other mutants studied here to show diminished to negligible binding relative to the wild type XPA_{67-80}. In conclusion, although the results of this study support conformational selection as the primary mechanism for ERCC1- XPA_{67-80} recognition, they are also consistent with the contribution of an induced fit step, guiding the conformational adjustment of the peptide conformation when in the binding site.[1] The timescales of this process may be too long to be observed through the MD simulations analyzed here. Due to the high sequence identity, it is likely that the findings of this conformational study also apply to the ERCC1-binding region of the XPA protein and to its molecular recognition mechanism.

ACKNOWLEDGMENTS

REFERENCES

1. Baron R, McCammon JA. Molecular recognition and ligand association. Ann Rev Phys Chem 2013;64:151–175.
2. Fischer E. Einfluss der configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges 1894;27:2984–2993.
3. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Ann Rev Biophys Biomol Struct 2003;32:335–373.
4. Jorgensen WL. Rusting of the lock and key model for protein-ligand binding. Science 1991;254:954–955.
5. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci U S A 1958;44:98–104.
6. Mizutani MY, Takamatsu Y, Ichinose T, Nakamura K, Itai A. Effective handling of induced-fit motion in flexible docking. Proteins 2006;63:878–891.
7. B-Rao C, Subramanian J, Sharma SD. Managing protein flexibility in docking and its applications. Drug Discov Today 2009;14:394–400.
8. Evenäs J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R, Kay LE. Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. J Mol Biol 2001;309:961–974.
9. Kim E, Lee S, Jeon A, Choi JM, Lee HS, Hohng S, Kim HS. A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. Nat Chem Biol 2013;9:313–318.
10. Vellore NA, Baron R. Molecular dynamics simulations indicate an induced-fit mechanism for LSD1/CoREST-H3-histone molecular recognition. BMC Biophys 2013;6:15.
11. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit 2010;23:105–116.
12. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 2009;5:789–796.
13. Feixas F, Lindert S, Sinko W, McCammon JA. Exploring the role of receptor flexibility in structure-based drug discovery. Biophys Chem 2014;186:31–45.
14. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KF, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 2008;320:1471–1475.
15. Mittermaier AK, Kay LE. Observing biological dynamics at atomic resolution using NMR. Trend Biochem Sci 2009;34:601–611.
16. King Frederick K, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. Nature 2007;448:325–329.
17. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Prot Sci 1999;8:1181–1190.
18. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trend Biochem Sci 2010;35:539–546.
19. Weikl TR, Paul F. Conformational selection in protein binding and function. Prot Sci 2014;23:1508–1518.
20. Peters JH, de Groot BL. Ubiquitin dynamics in complexes reveal molecular recognition mechanisms beyond induced fit and conformational selection. PLoS Comput Biol 2012;8:e1002704.
21. Grünberg R, Leckner J, Nilges M. Complementarity of structure ensembles in protein-protein binding. Structure 2004;12:2125–2136.
22. Vogt AD, Pozzi N, Chen Z, Di Cera E. Essential role of conformational selection in ligand binding. Biophys Chem 2014;186:13–21.
23. Vogt AD, Di Cera E. Conformational selection is a dominant mechanism of ligand binding. Biochemistry 2013;52:5723–5729.
24. Tsodikov OV, Ivanov D, Orelli B, Staresincic L, Shoshani I, Oberman R, Schärer OD, Wagner G, Ellenberger T. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. EMBO J 2007;26:4768–4776.
25. Li L, Peterson CA, Lu XY, Legerski RJ. Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision-repair. Mol Cell Biol 1995;15:1993–1998.
26. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. DNA repair pathways as targets for cancer therapy. Nat Cancer Rev 2008;8:193–204.
27. Croteau DL, Peng Y, Van Houten B. DNA repair gets physical: mapping an XPA-binding site on ERCC1. DNA Repair (Amst) 2008;7:819–826.

28. Shell SM, Chazin WJ. XPF-ERCC1: on the bubble. Structure 2012; 20:667–675.

29. Tripsianes K, Folkers G, Zheng C, Das D, Grinstead JS, Kaptein R, Boelens R. Analysis of the XPA and ssDNA-binding surfaces on the central domain of ERCC1 reveals evidence for subfunctionalization. Nucleic Acid Res 2007;35:5789–5798.

30. Fadda E. Conformational determinants for the recruitment of ERCC1 by XPA in the Nucleotide Excision Repair (NER) pathway: structure and dynamics of the XPA binding motif. Biophys J 2013; 104:2503–2511.

31. Orelli B, McClendon TB, Tsodikov OV, Ellenberger T, Niedernhofer LJ, Schärer OD. The XPA-binding domain of ERCC1 Is Required for Nucleotide Excision Repair but Not Other DNA Repair Pathways. J Biol Chem 2010;285:3705–3712.

32. Hess B, Kutzner C, van der Spoel D, Lindhal E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulations. J Chem Theory Comput 2008;4:435–447.

33. Maestro, version 9.7. New York, NY; 2014.

34. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958.

35. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 2004; 120:9665–9678.

36. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 2002;124:11258–11259.

37. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.

38. Grest GS, Kremer K. Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev A 1986;33:3628–3631.

39. Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford, New York: Oxford University Press; 1987.

40. Blomberg MRA, Siegbahn PEM. The mechanism for proton pumping in cytochrome c oxidase from an electrostatic and quantum chemical perspective. Biochim Biophys Acta 2012;1817:495–505.

41. York DM, Darden TA, Pedersen LG. The effect of long-range electrostatic interactions in simulations of macromolecular crystals. A comparision of the Ewald and truncated list methods. J Chem Phys 1993;99:8345–8348.

42. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103: 8577–8593.

43. Daura X, Antes I, van Gunsteren WF, Thiel W, Mark AE. The effect of motional averaging on the calculation of NMR-derived structural properties. Prot-Struct Funct Genet 1999;36:542–555.

44. Cardinale D, Guaitoli G, Tondi D, Luciani R, Henrich S, Salo-Ahen OH, Ferrari S, Marverti G, Guerrieri D, Ligabue A, Frassineti C, Pozzi C, Mangani S, Fessas D, Guerrini R, Ponterini G, Wade RC, Costi MP. Protein-protein interface-binding peptides inhibit the cancer therapy target human thymidylate synthase. Proc Natl Acad Sci U S A 2011;108:542–549.

45. Mukherjee S, Balius TE, Rizzo RC. Docking validation resources: protein family and ligand flexibility experiments. J Chem Inform Model 2010;50:1986–2000.

46. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 2004;32:W96–W99.

47. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45–50.

48. Man-Kuang Cheng T, Blundell TL, Fernandez-Recio J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. Proteins 2007;68:503–515.

49. Tsai CJ, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci USA 1999;96:9970–9972.

50. Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R. Structured disorder and conformational selection. Proteins 2001;44:418–427.

51. Wang Q, Zhang P, Hoffman L, Tripathi S, Homouz D, Liu Y, Waxham MN, Cheung MS. Protein recognition and selection through conformational and mutually induced fit. Proc Natl Acad Sci USA 2013;110:20545–20550.