Received: 11 October 2017 | Revised: 7 March 2018 | Accepted: 7 March 2018

DOI: 10.1002/bip.23117

ORIGINAL ARTICLE

WILEY Biopolymers

Molecular determinants of the interaction of EGCG with ordered and disordered proteins

Giuliana Fusco[1,2] | Maximo Sanz-Hernandez[1] | Francesco S. Ruggeri[2] | Michele Vendruscolo[2] | Christopher M. Dobson[2] | Alfonso De Simone[1]

[1]Department of Life Sciences, Imperial College London, London SW7 2AZ, United Kingdom

[2]Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

Correspondence
Alfonso De Simone, Department of Life Sciences, Imperial College London, London SW7 2AZ, United Kingdom.
Email: adesimon@imperial.ac.uk

Funding information
This research was supported by Parkinson's UK, Grant Number: G-1508; The UK Medical Research Council, Grant Number: MR/N000676/1; The UK Engineering and Physical Research Council, Grant Number: EP1579441; The Swiss National Science Foundation, Grant number: P300P2_171219; The Centre for Misfolding Diseases of the University of Cambridge

Abstract

The aggregation process of peptides and proteins is of great relevance as it is associated with a wide range of highly debilitating disorders, including Alzheimer's and Parkinson's diseases. The natural product (-)-epigallocatechin-3-gallate (EGCG) can redirect this process away from amyloid fibrils and towards non-toxic oligomers. In this study we used nuclear magnetic resonance (NMR) spectroscopy to characterize the binding of EGCG to a set of natively structured and unstructured proteins. The results show that the binding process is dramatically dependent on the conformational properties of the protein involved, as EGCG interacts with different binding modes depending on the folding state of the protein. We used replica exchange molecular dynamics simulations to reproduce the trends observed in the NMR experiments, and analyzed the resulting samplings to identify the dominant direct interactions between EGCG and ordered and disordered proteins.

1 | INTRODUCTION

The aggregation of proteins and peptides into amyloid fibrils is associated with a variety of neurodegenerative and non-neuropathic conditions, including Alzheimer's and Parkinson's diseases, type II diabetes and a number of forms of systemic amyloidosis.[1] In addition, several proteins have been found to adopt the amyloid state for functional purposes,[2–4] and there is much interest in the use of amyloid fibrils as novel biomaterials.[5,6] It is the significance of the amyloid state in human disease in particular, however, that has led to the considerable efforts to characterize its structural properties.[7–14]. It has also become increasingly evident that small and diffusible protein oligomers which represent intermediates in the formation of amyloid fibrils are likely to be the most highly toxic species giving rise to the onset and progression of these diseases.[15–22]

A number of strategies of molecular intervention to reduce the toxic effects of these oligomeric aggregates have been proposed, including the use of antibodies and molecular chaperones,[23–25] or the employment of small molecules.[26,27] Amongst the most promising compounds are natural products, several of which have been identified for their remarkable abilities to inhibit protein aggregation both in vitro and in vivo.[28–30] Thus, for example, polyphenols extracted from black tea have been shown to inhibit the aggregation of proteins associated with neurodegenerative disorders such as the amyloid-β peptide (Aβ) and α-synuclein, linked to Alzheimer's and Parkinson's diseases, by inducing the formation of different forms of oligomers that are non-toxic and highly stable.[31] Perhaps, the most interesting of these molecules is (-)- epigallocatechin-3-gallate (EGCG; Figure 1A), the major catechin found in green tea leaves.[30] EGCG has been associated with beneficial effects in cancer,[32] HIV pathogenesis,[33] and protein aggregation diseases;[34,35] indeed in the context of the latter it has been found to interact with a variety of proteins linked to amyloid formation such as α-synuclein,[36] Aβ,[37] transthyretin,[38] and huntingtin.[39] Despite the considerable degree of interest in this molecule, the molecular determinants of the interaction of EGCG with such disease-related proteins remains a matter of debate. For example, experimental evidence has been obtained that some

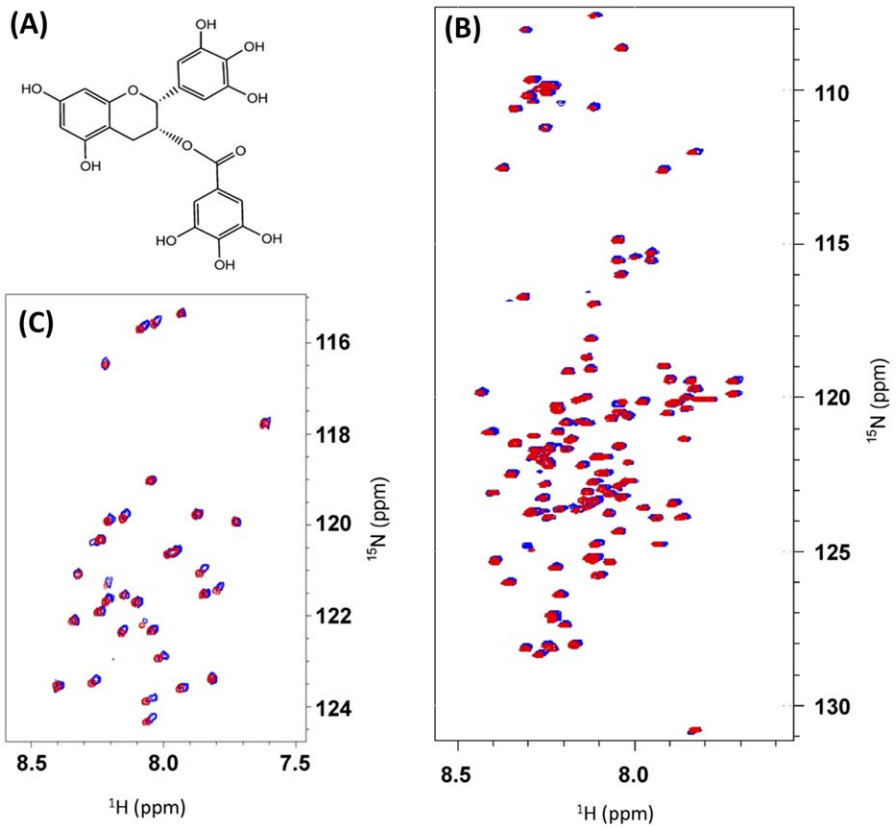

FIGURE 1 Interaction of EGCG and Aβ40 and α-synuclein. A, Chemical structure of EGCG. B,C, ^{1}H-^{15}N-HSQC spectra of α-synuclein B, and Aβ40 C, in isolation and in the presence of three molar equivalents of EGCG (red and blue spectra, respectively). The spectra were measured at 278 K using a 700 MHz NMR spectrometer. The α-synuclein sample was prepared at a protein concentration of 300 μM in 25 mM Tris-HCl (100 mM NaCl) buffer at pH 7.4. The Aβ40 sample was prepared at a protein concentration of 50 μM in 50 mM phosphate buffer at pH 7.4

oligomeric protein aggregates formed in the presence of EGCG are unstructured,[30] whereas other types of EGCG-induced oligomers contain significant levels of β-sheet structure.[40]

In the present study, we investigated further the effects of EGCG on the conformational properties of the monomeric states of Aβ40 and of α-synuclein by using solution-state NMR to analyze the populations of different forms of secondary structure in these intrinsically disordered systems. We then compared these results with those concerning the interaction between EGCG and natively folded proteins, showing that the modes of binding by EGCG are significantly different in structured and unstructured proteins. Finally, replica exchange molecular dynamics (REMD) simulations were employed to investigate in detail the molecular determinants of the binding affinities of EGCG with proteins. Taken together our results provide insights on the main interactions enabling EGCG to bind ordered and disordered polypeptide chains.

2 | MATERIALS AND METHODS

2.1 | Production, purification, and NMR experiments of individual protein systems

2.1.1 | α-synuclein

α-synuclein was purified in *Escherichia coli* using plasmid pT7-7 encoding for the protein as previously described.[41–43] The protein purity was analyzed by SDS-PAGE and the protein concentration was determined spectrophotometrically using $\varepsilon_{275} = 5600\ M^{-1}\ cm^{-1}$. NMR experiments were carried out using a 700 MHz spectrometer at 5°C and using a concentration of 300 μM in 25 mM Tris-HCl (100 mM NaCl) buffer at pH 7.4. Resonance assignments of the ^{1}H-^{15}N-HSQC spectrum of α-synuclein derived from our previous studies[41,42] and directly transferred easily to the spectra measured in the presence of EGCG. The assignment of the resonances of the backbone atoms was derived from the ^{1}H-^{15}N-HSQC in combination with 3D experiments (HNCO, CBCAcoNH, and HNHA) to obtain resonances of 6 atom types in total, namely Cα, Cβ, CO, NH, HN, Hα. BEST (Band-Selective Short Transient) implementation of HSQC and triple-resonance experiments[44] was used in order to acquire three-dimensional spectra in a short period of time, and to minimize the effects of the protein oligomerization in the NMR data acquisition. Fresh samples for the measurements of the protein in the presence of EGCG were employed for a maximum of 5 h during the NMR measurements.

2.1.2 | Aβ40

Aβ40 peptides obtained recombinantly were purchased from Alexo-Tech (Umea, Sweden) and handled on ice at all times. The powder was solubilized in 10 mM NaOH at concentrations of 4–7 mg mL^{-1} and stored at −80°C until required. To prepare NMR samples, Aβ40 stock solutions were diluted into sodium phosphate buffer (50 mM) to give a

final peptide concentration of 50 μM. The final buffer consisted of 20 mM sodium phosphate, 0.5 mM EDTA, 0.02% (w/v) NaN_3 and contained 10% D_2O. Samples were ultracentrifuged for 20 min at 8500*g* to sediment any aggregates of peptide present before being transferred to 3 mm MATCH NMR tubes (Bruker BioSpin AG, Fallenden, Switzerland). NMR experiments were carried out using a 700 MHz spectrometer at 5°C in 50 mM phosphate buffer at pH 7.4. Resonance assignments of the ^{1}H-^{15}N-HSQC spectrum of Aβ40 were derived from our previous studies[45] and directly transferred easily to the spectra measured in the presence of EGCG. The strategy of assignment of the 3D experiments followed the protocol described above for α-synuclein.

2.1.3 | Human lysozyme

Human lysozyme was expressed in *Pichia pastoris* and purified on an ion exchange column, as previously described.[46] ^{15}N ammonium sulfate was used to label the protein with ^{15}N. NMR experiments were carried out using a 700 MHz spectrometer at 310 K in a 50 mM phosphate buffer and pH 5.5, and using a concentration of lysozyme of 200 μM. Resonance assignments of the ^{1}H-^{15}N-HSQC spectrum of human lysozyme under these conditions were derived from our previous studies.[46]

2.1.4 | Human mAcP

Purification of mAcP was performed as described previously.[47] ^{15}N labelled protein was expressed as a GST fusion protein in the *E. coli* strain BL21-Gold(DE3) (Invitrogen), grown in minimal medium by using ^{15}N-enriched ammonium chloride, and purified using a glutathione column (Sigma–Aldrich). The GST/mAcP fusion protein was then cleaved with thrombin from human plasma (Sigma–Aldrich) in TRIS buffer. The eluted mAcP was then buffer exchanged into 30 mM ammonium carbonate buffer at pH 5.5 and then lyophilized. NMR experiments were carried out using a 700 MHz spectrometer at 310 K in a 50 mM acetate buffer and pH 5.5, and using a concentration of mAcP of 150 μM. Resonance assignments of the ^{1}H-^{15}N-HSQC spectrum of human mAcP under these conditions were derived from our previous studies.[47]

2.1.5 | EGCG samples

To avoid alteration of the experimental conditions of the protein samples upon addition of EGCG, for each protein system EGCG stocks of 5 mM were prepared by dissolving the catechin molecule in the equivalent buffer employed for the protein. Using this protocol, we generally observed no pH changes of the protein sample upon addition of 3 molar equivalents of EGCG and, in cases where minor pH changes occurred, pH was readjusted using NaOH or HCl.

2.1.6 | Atomic force microscopy (AFM) sample preparation and imaging

AFM measurements were made using previously published methods[48] by depositing on freshly etched bare mica substrates that are atomically flat (Supporting Information Figure S5). Preparation of the mica AFM samples was carried out at room temperature by the deposition of a 10 μL aliquot of the fully concentrated solution for 1 min. Then sample was then rinsed with ultrapure water and dried with a gentle flow of nitrogen.

High-resolution images (1024 × 1024 pixels) were collected using an NX10 Atomic Force Microscope (Park Systems, Suwon South Korea) under ambient conditions and in amplitude modulation noncontact (NC-AM) mode. We performed all the measurements using PPP-NCHR cantilevers (Park Systems, Suwon South Korea) with resonance frequency of 330 kHz and typical apical radius of 8 nm. Raw images were flattened with the built-in software (XEI, Park System, Suwon South Korea) and roughness measurements (Supporting Information Figure S5) were performed by SPIP software (Image Metrology, Hørsholm, Denmark).[48]

2.1.7 | Force field employed in the molecular dynamics simulations

Two systems were simulated in this study using replica exchange molecular dynamics (REMD) simulations. These are the full length mAcP in its native state and the unfolded state of mAcP, both in the presence of the EGCG molecule. The simulation of the unfolded state included only the fragment spanning residues 5–23 of mAcP. The choice of simulating only a segment of the protein is due to the low compaction of disordered states relative to folded proteins, which would have required a significantly larger system to simulate the full length mAcP with consequent dramatic lowering of exchange rates between replicas of the REMD. Both systems were simulated using the GROMOS54a7 force field[49] for the protein and the PRODRG[50] topology for EGCG, following the procedure adopted for our previous REMD analyses.[51] As REMD was employed to enhance the sampling of different protein surfaces by EGCG, high temperatures might have compromised the structural integrity of the mAcP protein. As a result, in the simulations of the native state of mAcP, a Go-model based on the method employed in previous investigations of an homologous acylphosphatase[52] was used to maintain the native topology. The Go-model was only applied to secondary structure elements of the protein, by enabling the loop regions to have higher conformational flexibility at high temperatures. RMSDs of the whole protein at 310 K are shown in Supporting Information Figure S3d, indicating that this setup was successful in preserving the structure of the protein while ensuring a highly converging exploration of the possible protein-EGCG interactions (Supporting Information Figure S3a). In the case of the disordered state of the fragment 5–23 of mAcP, we restrain the overall radius of gyration to have a value of 14.0 Å. This restrained enabled to avoid the collapse of the disordered fragment, which is typically observed with standard force fields that are parameterized to simulate folded states of proteins.

2.1.8 | REMD setup

Simulations were carried out with the GROMACS program[53] by using the particle mesh Ewald method (grid spacing 0.12 nm) to model the electrostatic contributions. A distance cut-off of 1.4 nm was used to account for van der Waals interactions. Simulations were carried out with an integration time step of 2 fs. mAcP molecules were immersed

in a box filled with extended single point charge (SPCE) water molecules by applying periodic boundary conditions. After an initial energy minimization, systems were equilibrated for 1 ns by restraining the protein coordinates in order to equilibrate the water molecules and the thermodynamics properties of each replica. All REMD samplings were performed for 100 ns of each replica, for a total of 3.2 μs in each sampling. Convergence was assessed using a number of parameters (Supporting Information Figure S3), indicating that the simulation time employed was sufficient to obtain convergence. All REMD samplings were composed of 32 replicas ranging from 310.0 to 371.5 K. The temperatures followed this pattern: 310.00, 311.85, 313.70, 315.57, 317.44, 319.33, 321.22, 323.12, 325.03, 326.95, 328.88, 330.81, 332.76, 334.71, 336.67, 338.64, 340.63, 342.62, 344.62, 346.63, 348.65, 350.67, 352.71, 354.76, 356.82, 358.89, 360.96, 363.05, 365.15, 367.26, 369.37, 371.5 K. This pattern was optimized for ensuring homogeneous exchange frequencies between replicas (Supporting Information Figure S2). For consistency of the sampling, the two REMD were based on the same temperature pattern, although the larger system size in the unfolded simulation resulted in reduced, but still high, exchange frequencies (Supporting Information Figure S2). General statistics of the REMD simulations are reported in Supporting Information Table S1.

2.1.9 | Analysis of the EGCG-mAcP contacts in REMD simulations

The probability of contacts between EGCG and mAcP was assessed in the 310 K replica of the REMD by counting, for every frame, the number of contacts between the atoms from the EGCG molecule and the backbone amide N—H groups of mAcP, that is, the groups probed in the ^{1}H-^{15}N-HSQC NMR spectra, using a 5 Å cutoff for the distance between the EGCG atom and the amide nitrogen atom. To compute the probability of contact along the mAcP sequence, for each residue the average number of contacts that the amide N—H established with EGCG atoms in the whole REMD simulation was calculated. This number was normalized with respect to the average number of total contacts that all the backbone N—H amide groups of the protein established with EGCG atoms in the sampling.

3 | RESULTS AND DISCUSSION

3.1 | Conformational properties of Aβ40 and α-synuclein in the presence of EGCG

We first analyzed the binding of EGCG to natively disordered proteins whose aggregation is associated with neurodegenerative diseases. Our results show that most of the peaks in the ^{1}H-^{15}N-HSQC spectra of Aβ40 and α-synuclein are perturbed upon incubation with EGCG, a finding that is consistent with previous measurements[30,40] (Figure 1B, C). In the presence of an excess of EGCG relative to protein molecules, the NMR resonances of the protein are broadened beyond detection due to the conformational exchange between protein monomers, whose NMR resonance are visible, and slow-tumbling protein-EGCG aggregates, whose NMR resonance are invisible. By incubating the

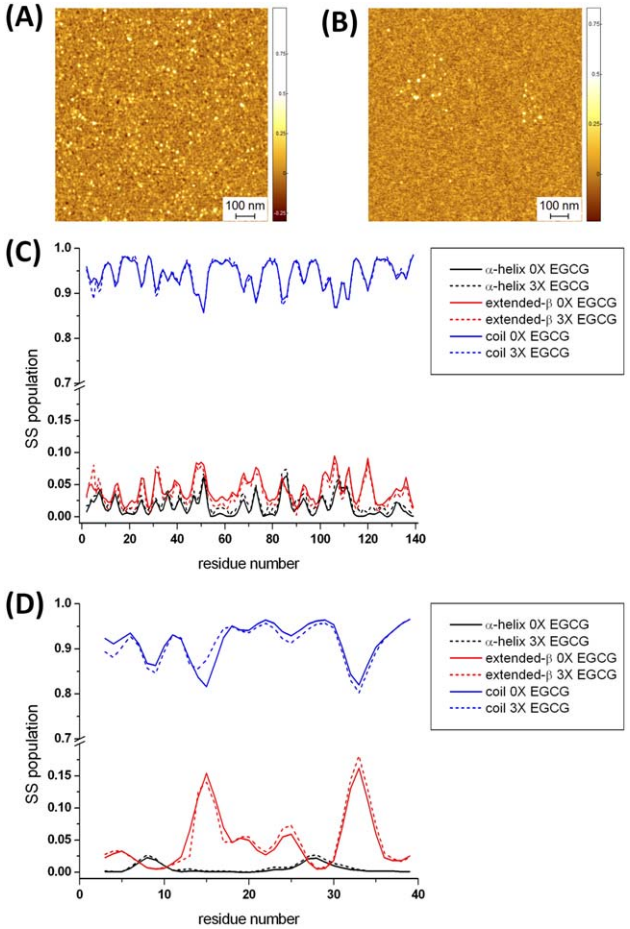

FIGURE 2 Effects of EGCG on the conformational properties of Aβ40 and α-synuclein. A,B, Properties of the oligomers formed by Aβ40 A, and α-synuclein B, after 48 h of incubation at 310 K with three molar equivalents of EGCG, as probed by AFM (scale bar 100 nm). Experimental conditions as in Figure 1. C,D, Population of residual secondary structure elements in Aβ40 C, and α-synuclein D, upon incubation at 278 K with three molar equivalents of EGCG (sample conditions as in Figure 1). The populations were estimated using δ2D[55] to analyze the measured NMR chemical shifts under these conditions. α-helix, extended-β population and coil (combining PPII and unstructured populations from δ2D) are shown using black, red and blue lines, respectively. Continuous and dashed lines are used for reporting the populations in the absence and presence of three molar equivalents of EGCG, respectively

proteins with relatively low ratios of EGCG relative to the protein molecules, however, the broadening of the protein resonances is not excessive and allows the acquisition of three-dimensional NMR experiments to measure the ^{13}C resonances of the protein. We therefore measured the backbone resonances, including ^{13}Cα, ^{13}Cβ, ^{13}CO, ^{1}Hα, ^{1}HN, and ^{13}N of Aβ40 and α-synuclein in the presence and absence of three molar excess of EGCG (Figure 1B, Supporting Information Figure S1A), and analyzed the chemical shifts of the resonances that were previously assigned[41,45] to assess the effects of the interaction with EGCG on the conformational properties of the monomeric states of the two proteins. Backbone chemical shifts of ^{13}Cα or ^{13}Cβ in particular are extremely sensitive probes of the secondary structure content

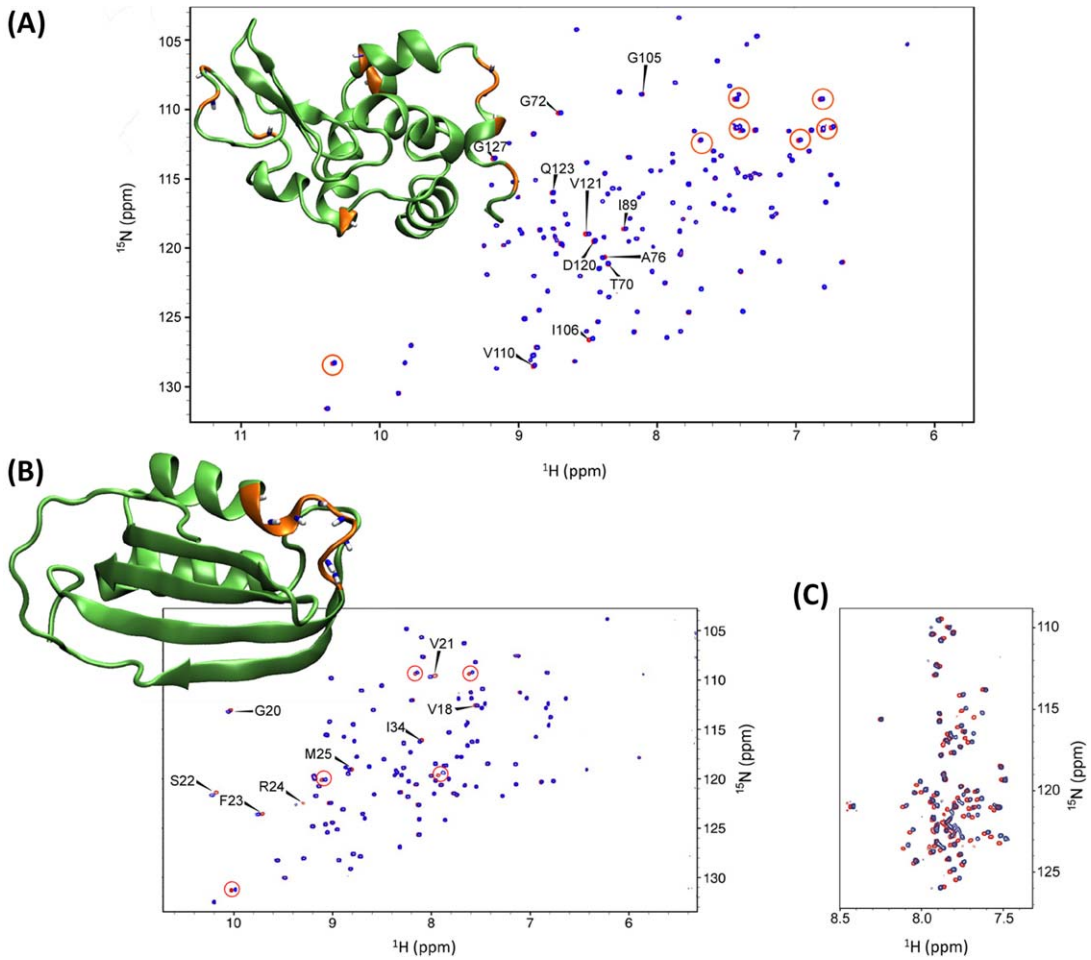

FIGURE 3 Binding of EGCG to lysozyme and mAcP. A, ^{1}H-^{15}N-HSQC spectra of human lysozyme in the absence (red) and presence (blue) of three molar equivalents of EGCG. The spectra were measured at 310 K in 50 mM phosphate buffer and pH 5.5, at a concentration of lysozyme of 200 μM using a spectrometer operating at a ^{1}H frequency of 700 MHz. B, ^{1}H-^{15}N-HSQC spectra of human muscle acylphosphatase (mAcP) in the absence (red) and presence (blue) of three molar equivalents of EGCG. The spectra were measured at 310 K in 50 mM acetate buffer and pH 5.5, and using a concentration of mAcP of 150 μM using a spectrometer operating at a ^{1}H frequency of 700 MHz. C, Same as panel B, but in the presence of 6 M guanidinium chloride

of disordered proteins,[54] and their values can be used in methods such as δ2D[55] to obtain accurate information on the nature of the residual structure and of the dynamical properties of the residues in the sequence.[56]

The relatively small quantity of EGCG used in the NMR experiments was observed to still induce the aggregation of both proteins into oligomers, previously shown to be of spherical shape,[15,40] after 48 h of incubation at 310 K. In particular, atomic force microscopy (AFM,[48] Figure 2A,B) images revealed protein oligomers with similar properties to the aggregates observed upon incubation with larger quantities of EGCG.[15,30] The NMR experiments performed at 278 K enabled the detection of sharp signals in 3D NMR spectra of the monomeric states of the proteins in the presence of the EGCG (see Methods). Under these conditions, the chemical shift perturbation in the ^{1}H-^{15}N- HSQC spectra of both Aβ40 and α-synuclein resulting from the addition of three molar equivalents of EGCG are on average smaller than 0.05 ppm. The analysis by δ2D[55] indicates that the presence of EGCG caused no significant changes in the populations of the residual secondary structure of the monomeric states of the two intrinsically disordered proteins (Figure 2C,D). In particular, in the case of α-synuclein, where the NMR chemical shifts are indicative of an ensemble of conformations that is rather close to a random-coil state,[57,58] the presence of three equivalents of EGCG does not alter the very low content of secondary structure (Figure 2D), a finding consistent with the biophysical evidence that the oligomers of α-synuclein formed in the presence of EGCG are primarily disordered.[30] In the case of Aβ40 (Figure 1C, Supporting Information Figure S1B), the populations of secondary structure elements obtained by analyzing the chemical shifts using δ2D[55] are also indicative of a highly disordered state. Consistently with the case of α-synuclein, three molar equivalent excess of EGCG was not observed to modify the secondary structure content in Aβ40 (Figure 2C). The finding that the monomeric state of Aβ40 maintains a residual content of secondary structure in the presence of EGCG is consistent with the conformational properties of the Aβ40 oligomers generated in the presence of this molecule, as observed using ssNMR.[40]

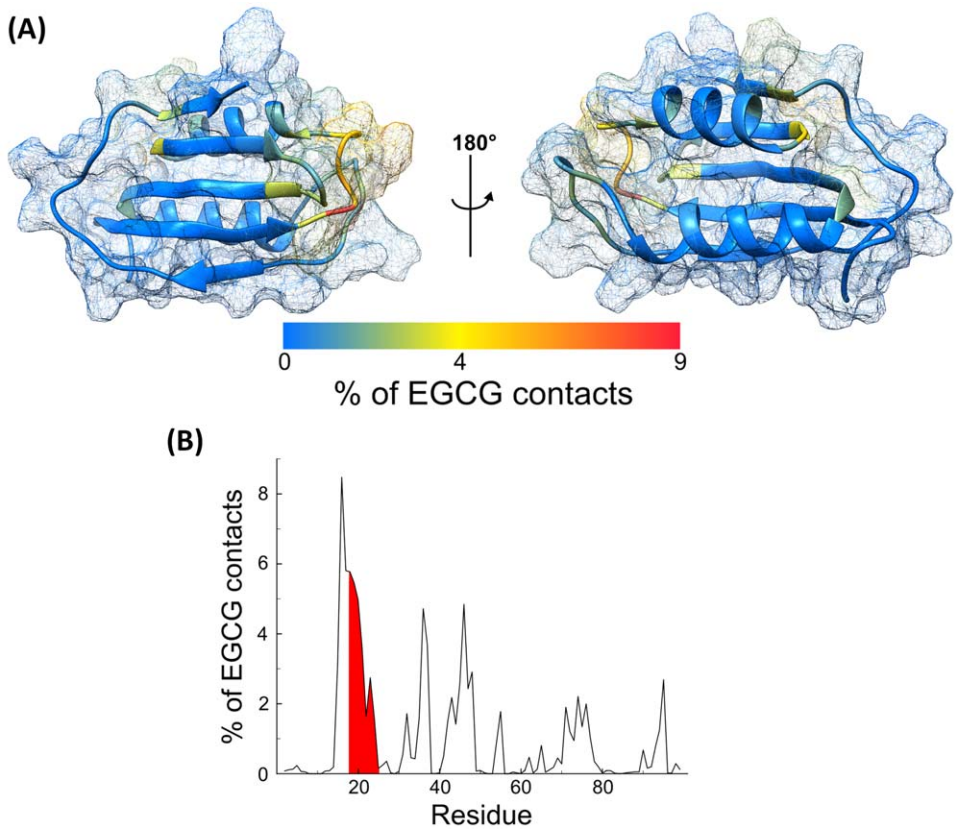

FIGURE 4 Probability of contacts with EGCG mapped on the structure of mAcP in REMD simulations. A, Structure of human mAcP shown with ribbons and surfaces, and colored using a scale ranging from blue (0% of EGCG contacts) to red (9% of probability contacts per residue). The contact map was calculated from replicas equilibrated at 310 K using the analysis described in the Methods section. B, Graph of the probability of contacts (%) between the residues of mAcP and EGCG, as analyzed on replicas equilibrated at 310 K

Taken together these data show that the EGCG molecule affects only weakly the conformational properties of the monomeric states of Aβ40 and α-synuclein, and therefore suggest that the ability to promote the formation of oligomers by this molecule is likely to originate from other factors, such as for example the ability of EGCG to promote protein–protein interactions.

3.2 | Binding preferences of EGCG in natively folded proteins

We then examined the residue-specific binding of EGCG with two folded proteins, which have previously been largely characterized in great detail by NMR, namely human lysozyme,[46] a protein of similar size to α-synuclein, and human muscle acylphosphatase (mAcP)[47] of slightly smaller size. In the case of lysozyme, we found that the incubation with three molar equivalents of EGCG has a different effect on the ^{1}H-^{15}N-HSQC spectra than those measured in the cases of the disordered Aβ40 and α-synuclein (Figure 3, Supporting Information Figure S1). In particular, only a specific small subset of ^{1}H-^{15}N-HSQC peaks of backbone and side chain N-H groups are affected significantly by the presence of the EGCG, with the majority of the ^{1}H-^{15}N-HSQC peaks remaining unperturbed (Figure 3A, Supporting Information Figure S1C). Analysis of the resonances perturbed by EGCG indicates that the protein residues interacting most directly with the natural molecule are located primarily in exposed loops (T70, G72, A76, I89, G105, D120, V121, G127) or in the exposed N-terminal regions of the α-helices (I106, V110, Q123) of the protein. A similar behavior was observed in the case of mAcP (Figure 3B, Supporting Information Figure S1D), where the specific ^{1}H-^{15}N-HSQC peaks perturbed by the presence of EGCG are from residues located in the region spanning the catalytic loop (residues 18–22) and the exposed N-terminal residues of the helix 1 (residues 23–25).

Remarkably, the addition of 6 M guanidinium hydrochloride to unfold mAcP resulted in the majority of the ^{1}H-^{15}N-HSQC resonances of the backbone amide groups of the protein being perturbed by the presence of EGCG (Figure 3C). This finding reveals that when the protein is folded in its native state, only the exposed and highly dynamical region 18–25 interacts with EGCG, but when the same protein chain is denatured, its interaction with EGCG resembles that of the intrinsically disordered Aβ40 and α-synuclein proteins (Figure 3C).

Taken together these data suggest that the molecular affinity of EGCG is specific for highly mobile and exposed backbone and side chain regions, a condition that in a highly disordered state is uniformly matched throughout the protein sequence whereas in a folded state is only found in some specific regions of the protein.

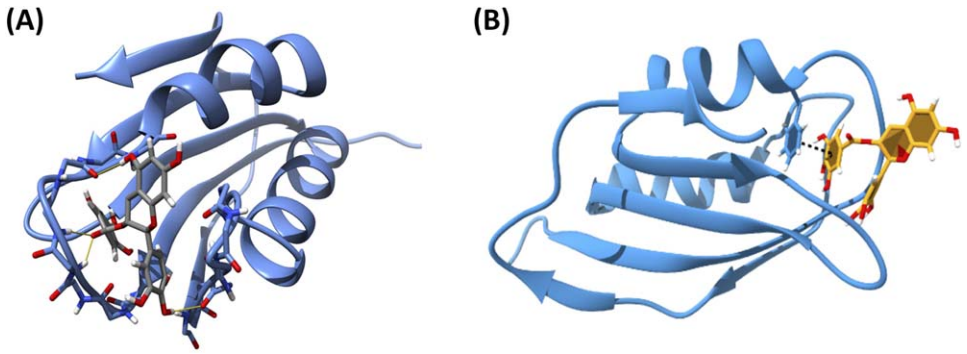

FIGURE 5 Principal direct interactions between EGCG and mAcP in REMD simulations. A, The primary sources of interactions between EGCG and the mAcP protein in the simulations are direct hydrogen bonds between donor/acceptor groups on the EGCG molecule and the main chain and side chain carbonyl and amide groups of the protein. B, The second most frequent interaction involves aromatic rings of EGCG and of the side chains of mAcP

3.3 | Molecular basis of the EGCG binding affinity investigated by molecular dynamics simulations

We next employed enhanced molecular dynamics (MD) simulations to investigate the nature of the interactions between EGCG and protein molecules. We aimed at reproducing the conditions of the NMR experiments with mAcP by using replica exchange MD (REMD) to sample the conformational space of EGCG around the protein in the presence of water molecules. The simulations were carried out using 32 replicas, each simulated for 100 ns, using a temperature range between 310 and 371.5 K (see Methods). The high acceptance rates of exchange steps between consecutive replicas (Supporting Information Figure S2) and the good convergence of the simulations (Supporting Information Figure S3) indicated that the simulation setup was appropriate for our targets. We first assessed if the simulations were able to reproduce the distributions of selective binding of EGCG to the surface of mAcP. This analysis was performed by mapping the mAcP:EGCG contacts as a function of the protein sequence (see Methods). Although the simulations do not comprehensively reproduce the selectivity of binding detected experimentally, the results from the simulations are remarkably consistent with those from the NMR experiments, and support the conclusion that EGCG has a specific preference for exposed loop regions of the protein (Figure 4). The region of the protein with highest degree of contact with EGCG in the simulations is the flexible loop spanning residues 16–25 (Figure 4B), which agrees well with the region 18–25 found experimentally (Supporting Information Figure S1D) to have the strongest interactions with the catechin molecule. We then extended the simulations to compare the binding of EGCG to the unfolded state of mAcP. To maintain a similar box size and number of water molecules, which is fundamental to ensure comparable exchange rates in the two REMD samplings, we simulated a segment of mAcP, residues 5–23, in the unfolded state. In the simulations of the folded state of mAcP, this fragment spans the regions establishing the highest (loop 16–20) and lowest (strand S1, residues 7–16) amount of interactions with EGCG (Figure 4). When this fragment is simulated in its unfolded state, however, the contacts with EGCG were observed to be distributed across the whole sequence, including the strand S1 that showed no contacts with EGCG in the simulations of the folded state of mAcP (Supporting Information Figure S4). Taken together, these findings indicate that the REMD simulations reproduce, at the very least, the trends identified using NMR experiments, with specific binding of EGCG to the dynamical exposed loops in the folded state of the mAcP protein (Figure 4) and nonspecific interactions throughout the main chain in its disordered state (Supporting Information Figure S4).

The high degree of convergence of the REMD samplings and the agreement with the trends indicated by the NMR experiments prompted us to analyze the simulations to identify the molecular determinants of the interaction between EGCG and mAcP. The analysis indicates that the majority of the protein–EGCG interactions result from hydrogen bonds involving amides or carbonyl groups of the protein backbone (Figure 5A). Moreover, the availability in the EGCG molecule of multiple groups that are able to engage in the formation of such hydrogen bonds enables multiple interactions to be established simultaneously. Another type of interaction indicated by the REMD sampling is the stacking of aromatic rings of EGCG and of protein sidechains (Figure 5B), whose accessibility is also enhanced in disordered protein states.

4 | CONCLUSIONS

This study has shown that the interaction between EGCG and disordered proteins, represented in this study by Aβ40 and α-synuclein, does not significantly alter their conformational preferences and suggest that other mechanisms are responsible of the induction of oligomeric protein aggregates by this molecule, for example EGCG-mediated induction of protein–protein interactions. We also found that EGCG binds proteins with very different specificities depending on their conformational state; in folded states there is selective binding in specific regions of the protein structure, notably exposed and highly dynamical loops, while disordered states are bound uniformly along the protein sequence. Enhanced molecular simulations identified two main types of interactions between EGCG and protein molecules, namely hydrogen bonding between the molecule and the protein backbone and stacking interactions including aromatic groups. Both types of interaction can be readily formed by disordered proteins as a result of

the accessibility of backbone atoms and aromatic rings in such states. We suggest, therefore, that the ability of EGCG to establish simultaneously multiple hydrogen bonds and aromatic interactions offers insights on the molecular basis by which this molecule is able to mediate protein-protein interactions and induce their aggregation into nontoxic forms of protein oligomers.

ACKNOWLEDGMENTS

This research was supported by Parkinson's UK (G-1508), the UK Medical Research Council (MR/N000676/1), the UK Engineering and Physical Research Council (EP1579441), the Swiss National Science Foundation (P300P2_171219) and the Centre for Misfolding Disease of the University of Cambridge.

ORCID

Giuliana Fusco http://orcid.org/0000-0002-3644-9809
Maximo Sanz-Hernandez http://orcid.org/0000-0001-8522-8730
Francesco Simone Ruggeri http://orcid.org/0000-0002-1232-1907
Michele Vendruscolo http://orcid.org/0000-0002-3616-1610
Christopher M Dobson http://orcid.org/0000-0002-5445-680X
Alfonso De Simone http://orcid.org/0000-0001-8789-9546

REFERENCES

[1] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2017**, *86*, 27.

[2] R. Narayanaswamy, M. Levy, M. Tsechansky, G. M. Stovall, J. D. O'Connell, J. Mirrielees, A. D. Ellington, E. M. Marcotte, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10147.

[3] M. Sunde, A. H. Kwan, M. D. Templeton, R. E. Beever, J. P. Mackay, *Micron* **2008**, *39*, 773.

[4] S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale, R. Riek, *Science* **2009**, *325*, 328.

[5] C. A. Hauser, S. Maurer-Stroh, I. C. Martins, *Chem. Soc. Rev.* **2014**, *43*, 5326.

[6] D. Papapostolou, A. M. Smith, E. D. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell, D. N. Woolfson, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10853.

[7] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, R. Tycko, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16742.

[8] M. D. Tuttle, G. Comellas, A. J. Nieuwkoop, D. J. Covell, D. A. Berthold, K. D. Kloepper, J. M. Courtney, J. K. Kim, A. M. Barclay, A. Kendall, W. Wan, G. Stubbs, C. D. Schwieters, V. M. Lee, J. M. George, C. M. Rienstra, *Nat. Struct. Mol. Biol.* **2016**, *23*, 409.

[9] A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang, L. Jiang, M. A. Arbing, B. L. Nannenga, J. Hattne, J. Whitelegge, A. S. Brewster, M. Messerschmidt, S. Boutet, N. K. Sauter, T. Gonen, D. S. Eisenberg, *Nature* **2015**, *525*, 486.

[10] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel, D. Eisenberg, *Nature* **2007**, *447*, 453.

[11] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, *Science* **2008**, *319*, 1523.

[12] A. W. Fitzpatrick, G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaroniec, L. Wang, V. Ladizhansky, S. A. Muller, C. E. MacPhee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5468.

[13] A. W. P. Fitzpatrick, B. Falcon, S. He, A. G. Murzin, G. Murshudov, H. J. Garringer, R. A. Crowther, B. Ghetti, M. Goedert, S. H. W. Scheres, *Nature* **2017**, *547*, 185.

[14] T. P. Knowles, A. De Simone, A. W. Fitzpatrick, A. Baldwin, S. Meehan, L. Rajah, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, *Phys. Rev. Lett.* **2012**, *109*, 158101.

[15] G. Fusco, S. W. Chen, P. T. F. Williamson, R. Cascella, M. Perni, J. A. Jarvis, C. Cecchi, M. Vendruscolo, F. Chiti, N. Cremades, L. Ying, C. M. Dobson, A. De Simone, *Science* **2017**, *358*, 1440.

[16] N. Cremades, S. I. Cohen, E. Deas, A. Y. Abramov, A. Y. Chen, A. Orte, M. Sandal, R. W. Clarke, P. Dunne, F. A. Aprile, C. W. Bertoncini, N. W. Wood, T. P. Knowles, C. M. Dobson, D. Klenerman, *Cell* **2012**, *149*, 1048.

[17] S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi, F. Chiti, *Nat. Chem. Biol.* **2010**, *6*, 140.

[18] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, M. Stefani, *Nature* **2002**, *416*, 507.

[19] D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, D. Selkoe, *J. Nat.* **2002**, *416*, 535.

[20] B. Winner, R. Jappelli, S. K. Maji, P. A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar, S. Campioni, C. Tzitzilonis, A. Soragni, S. Jessberger, H. Mira, A. Consiglio, E. Pham, E. Masliah, F. H. Gage, R. Riek, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4194.

[21] O. W. Wan, K. K. Chung, *PLoS One* **2012**, *7*, e38545.

[22] R. Kayed, Y. Sokolov, B. Edmonds, T. M. McIntire, S. C. Milton, J. E. Hall, C. G. Glabe, *J. Biol. Chem.* **2004**, *279*, 46363.

[23] S. I. A. Cohen, P. Arosio, J. Presto, F. R. Kurudenkandy, H. Biverstal, L. Dolfe, C. Dunning, X. Yang, B. Frohm, M. Vendruscolo, J. Johansson, C. M. Dobson, A. Fisahn, T. P. J. Knowles, S. Linse, *Nat. Struct. Mol. Biol.* **2015**, *22*, 207.

[24] M. Dumoulin, A. M. Last, A. Desmyter, K. Decanniere, D. Canet, G. Larsson, A. Spencer, D. B. Archer, J. Sasse, S. Muyldermans, L. Wyns, C. Redfield, A. Matagne, C. V. Robinson, C. M. Dobson, *Nature* **2003**, *424*, 783.

[25] P. Sormanni, F. A. Aprile, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9902.

[26] D. M. Walsh, M. Townsend, M. B. Podlisny, G. M. Shankar, J. V. Fadeeva, O. El Agnaf, D. M. Hartley, D. J. Selkoe, *J. Neurosci.* **2005**, *25*, 2455.

[27] P. Kuner, B. Bohrmann, L. O. Tjernberg, J. Naslund, G. Huber, S. Celenk, F. Gruninger-Leitch, J. G. Richards, R. Jakob-Roetne, J. A. Kemp, C. Nordstedt, *J. Biol. Chem.* **2000**, *275*, 1673.

[28] Z. Ignatova, L. M. Gierasch, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13357.

[29] M. Perni, C. Galvagnion, A. Maltsev, G. Meisl, M. B. Muller, P. K. Challa, J. B. Kirkegaard, P. Flagmeier, S. I. Cohen, R. Cascella, S. W. Chen, R. Limboker, P. Sormanni, G. T. Heller, F. A. Aprile, N. Cremades, C. Cecchi, F. Chiti, E. A. Nollen, T. P. Knowles, M. Vendruscolo, A. Bax, M. Zasloff, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1009.

[30] D. E. Ehrnhoefer, J. Bieschke, A. Boeddrich, M. Herbst, L. Masino, R. Lurz, S. Engemann, A. Pastore, E. E. Wanker, *Nat. Struct. Mol. Biol.* **2008**, *15*, 558.

[31] G. Grelle, A. Otto, M. Lorenz, R. F. Frank, E. E. Wanker, J. Bieschke, *Biochemistry* **2011**, *50*, 10624.

[32] J. Jankun, S. H. Selman, R. Swiercz, E. Skrzypczak-Jankun, *Nature* **1997**, *387*, 561.

[33] G. Fassina, A. Buffa, R. Benelli, O. E. Varnier, D. M. Noonan, A. Albini, *Aids* **2002**, *16*, 939.

[34] Y. Levites, O. Weinreb, G. Maor, M. B. Youdim, S. Mandel, *J. Neurochem.* **2001**, *78*, 1073.

[35] N. Lorenzen, S. B. Nielsen, Y. Yoshimura, B. S. Vad, C. B. Andersen, C. Betzer, J. D. Kaspersen, G. Christiansen, J. S. Pedersen, P. H. Jensen, F. A. Mulder, D. E. Otzen, *J. Biol. Chem.* **2014**, *289*, 21299.

[36] J. Bieschke, J. Russ, R. P. Friedrich, D. E. Ehrnhoefer, H. Wobst, K. Neugebauer, E. E. Wanker, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7710.

[37] K. Rezai-Zadeh, G. W. Arendash, H. Hou, F. Fernandez, M. Jensen, M. Runfeldt, R. D. Shytle, J. Tan, *Brain Res.* **2008**, *1214*, 177.

[38] N. Ferreira, I. Cardoso, M. R. Domingues, R. Vitorino, M. Bastos, G. Bai, M. J. Saraiva, M. R. Almeida, *FEBS Lett.* **2009**, *583*, 3569.

[39] D. E. Ehrnhoefer, M. Duennwald, P. Markovic, J. L. Wacker, S. Engemann, M. Roark, J. Legleiter, J. L. Marsh, L. M. Thompson, S. Lindquist, P. J. Muchowski, E. E. Wanker, *Hum. Mol. Genet.* **2006**, *15*, 2743.

[40] J. M. Lopez del Amo, U. Fink, M. Dasari, G. Grelle, E. E. Wanker, J. Bieschke, B. Reif, *J. Mol. Biol.* **2012**, *421*, 517.

[41] G. Fusco, A. De Simone, T. Gopinath, V. Vostrikov, M. Vendruscolo, C. M. Dobson, G. Veglia, *Nat. Commun.* **2014**, *5*, 3827.

[42] G. Fusco, T. Pape, A. D. Stephens, P. Mahou, A. R. Costa, C. F. Kaminski, G. S. Kaminski Schierle, M. Vendruscolo, G. Veglia, C. M. Dobson, A. De Simone, *Nat. Commun.* **2016**, *7*, 12563.

[43] G. Fusco, A. De Simone, P. Arosio, M. Vendruscolo, G. Veglia, C. M. Dobson, *Sci. Rep.* **2016**, *6*, 27125.

[44] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014.

[45] F. N. Newby, A. De Simone, M. Yagi-Utsumi, X. Salvatella, C. M. Dobson, M. Vendruscolo, *Biochemistry* **2015**, *54*, 6876.

[46] A. De Simone, F. A. Aprile, A. Dhulesia, C. M. Dobson, M. Vendruscolo, *Elife* **2015**, *4*, e02777.

[47] G. Fusco, A. De Simone, S. T. Hsu, F. Bemporad, M. Vendruscolo, F. Chiti, C. M. Dobson, *Biomol. NMR Assign.* **2012**, *6*, 27.

[48] F. S. Ruggeri, S. Vieweg, U. Cendrowska, G. Longo, A. Chiki, H. A. Lashuel, G. Dietler, *Sci. Rep.* **2016**, *6*, 31155.

[49] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, W. F. van Gunsteren, *Eur. Biophys. J.* **2011**, *40*, 843.

[50] A. W. Schuttelkopf, D. M. van Aalten, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1355.

[51] L. Vitagliano, L. Esposito, C. Pedone, A. De Simone, *Biochem. Biophys. Res. Commun.* **2008**, *377*, 1036.

[52] A. De Simone, A. Dhulesia, G. Soldi, M. Vendruscolo, S. T. Hsu, F. Chiti, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21057.

[53] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, *29*, 845.

[54] C. Camilloni, D. Schaal, K. Schweimer, S. Schwarzinger, A. De Simone, *Biophys. J.* **2012**, *102*, 158.

[55] C. Camilloni, A. De Simone, W. F. Vranken, M. Vendruscolo, *Biochemistry* **2012**, *51*, 2224.

[56] J. M. Krieger, G. Fusco, M. Lewitzky, P. C. Simister, J. Marchant, C. Camilloni, S. M. Feller, A. De Simone, *Biophys. J.* **2014**, *106*, 1771.

[57] J. H. Lee, F. Li, A. Grishaev, A. Bax, *J. Am. Chem. Soc.* **2015**, *137*, 1432.

[58] A. S. Maltsev, J. Ying, A. Bax, *Biochemistry* **2012**, *51*, 5004.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Fusco G, Sanz-Hernandez M, Ruggeri FS, Vendruscolo M, Dobson CM, De Simone A. Molecular determinants of the interaction of EGCG with ordered and disordered proteins. *Biopolymers*. 2018;e23117. https://doi.org/10.1002/bip.23117