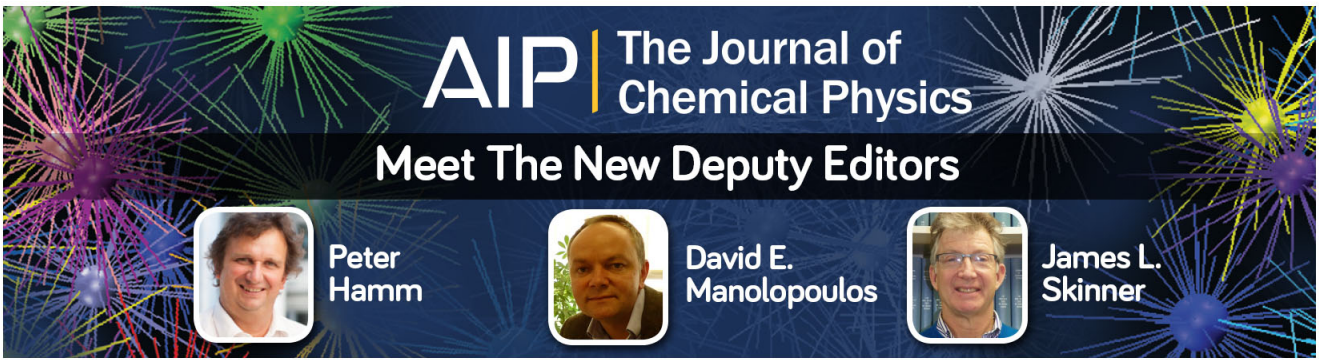
Identification of small-molecule binding pockets in the soluble monomeric form of the Aβ42 peptide

Maximillian Zhu, Alfonso De Simone, Dale Schenk, Gergely Toth, Christopher M. Dobson, and Michele Vendruscolo

Citation: The Journal of Chemical Physics **139**, 035101 (2013); doi: 10.1063/1.4811831
View online: http://dx.doi.org/10.1063/1.4811831
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/139/3?ver=pdfcov
Published by the AIP Publishing



Identification of small-molecule binding pockets in the soluble monomeric form of the Aβ42 peptide

Maximillian Zhu,[1] Alfonso De Simone,[1] Dale Schenk,[2] Gergely Toth,[1,a)]
Christopher M. Dobson,[1,a)] and Michele Vendruscolo[1,a)]
[1]*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom*
[2]*Elan Pharmaceuticals, South San Francisco, California 94080, USA*

(Received 17 February 2013; accepted 13 May 2013; published online 15 July 2013)

The aggregation of intrinsically disordered peptides and proteins is associated with a wide range of highly debilitating neurological and systemic disorders. In this work we explored the potential of a structure-based drug discovery procedure to target one such system, the soluble monomeric form of the Aβ42 peptide. We utilised for this purpose a set of structures of the Aβ42 peptide selected from clusters of conformations within an ensemble generated by molecular dynamics simulations. Using these structures we carried out fragment mapping calculations to identify binding "hot spots" on the monomeric form of the Aβ42 peptide. This procedure provided a set of hot spots with ligand efficiencies comparable to those observed for structured proteins, and clustered into binding pockets. Such binding pockets exhibited a propensity to bind small molecules known to interact with the Aβ42 peptide. Taken together these results provide an initial indication that fragment-based drug discovery may represent a potential therapeutic strategy for diseases associated with the aggregation of intrinsically disordered proteins. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4811831]

INTRODUCTION

The aggregation of intrinsically disordered peptides and proteins is associated with a wide range of human disorders, including Alzheimer's and Parkinson's diseases.[1,2] These diseases, for which at present there are no effective treatments, are increasingly common in our ageing society,[3] prompting a variety of therapeutic strategies to be proposed and pursued.[4,5] Among such strategies, increasing attention has been devoted to finding drug-like small molecules capable of interfering with the aggregation process of intrinsically disordered proteins, and of promoting their normal behaviour.[6–10] In this context, stabilizing the soluble monomeric form of these proteins is appealing because it can influence downstream aggregation events,[7,11] including the formation of small oligomeric species that are increasingly recognised as the origin of neuronal damage.[2,12,13] It has indeed been suggested that intrinsically disordered proteins may be targeted by identifying specific sequence regions that exhibit specific "molecular recognition features" (MoRFs).[9,14,15]

In this work we investigate an alternative approach to this problem, which is based on the structure-based search of potential binding pockets in intrinsically disordered proteins. Although structure-based drug discovery is a strategy that has been effective in identifying small-molecule ligands that bind to the native states of globular proteins,[16,17] there are two major challenges in the extension of this approach to intrinsically disordered proteins. The first is the existence of very substantial technical difficulties in acquiring accurate information about the structure and dynamics of disordered proteins by experimental methods[18–20] and the second is that the binding pockets in these molecules are likely to be present only transiently. Despite these problems, recent evidence indicates that disordered binding interfaces can be effectively targeted by small molecules.[8–10]

In order to explore the potential of this structure-based approach for disordered monomeric polypeptide chains, we have considered the 42-residue Aβ peptide (Aβ42), whose aggregation process is associated with the pathogenesis of Alzheimer's disease.[1,2] This peptide is highly disordered in solution, populating a heterogeneous ensemble of conformations,[21,22] a situation in sharp contrast to that of the fibrillar state of the peptide, which is ordered and has been characterised in general terms, notably by X-ray fibre diffraction,[23] electron microscopy and solid-state NMR[24] studies.

Here, we have adopted a strategy for identifying small-molecule binding sites in the Aβ42 peptide in which molecular dynamics simulations are combined with fragment-based drug design.[25–28] Given the challenges in obtaining atomistic descriptions of the conformational ensembles populated by intrinsically disordered regions,[18–20] molecular dynamics simulations represent a convenient tool in elucidating the structures and dynamics of these systems.[29–40] In this view, the approach that we describe extends to intrinsically disordered proteins a type of strategy that, from the relaxed complex method[41] to subsequent techniques,[42–48] has been aimed at introducing flexibility in docking. In fragment-based drug design a library of small-molecule fragments is screened to find those that have a propensity to bind to specific "hot spot" regions in a given conformation of a protein.[25–28] By using this approach we present evidence that in the case of the Aβ42 peptide it is possible to identify clusters of binding hot spots

a)Electronic addresses: mv245@cam.ac.uk; cmd44@cam.ac.uk; and gt293@cam.ac.uk

that could serve as binding sites for drug-like small molecules assembled from fragments.

METHODS

Replica-exchange molecular dynamics simulations

There are several different procedures that can be used to generate structural ensembles representing the conformational space of intrinsically disordered peptides and proteins, including molecular modelling,[49,50] molecular dynamics simulations,[29–38] and molecular simulations with NMR restraints.[18–20,51,52] In the present case an ensemble was obtained by performing replica-exchange molecular dynamics (REMD) simulations[53] using the GROMACS molecular dynamics simulations package,[54] following a protocol similar to one used recently.[55] The REMD method enhances the sampling of the conformational space of polypeptide chains by overcoming energy barriers that could otherwise trap the simulations in local minima.[53] The duration of the simulation was 100 ns, using an integration step of 2 fs, at 48 temperatures ranging from 276.1 to 376.9 K with the AMBER99SB force field[56] and the TIP3P water model;[57] this force field has been shown to reproduce with rather good accuracy NMR parameters in molecular dynamics simulations of other peptides and proteins.[58,59] The monomeric form of the Aβ42 peptide with charged termini was placed in a box of 7 × 7 × 7 nm^3 in periodic boundary conditions, with about 10 000 water molecules and three Na^+ counterions to neutralize the net charge of the peptide. As the protonation state of titratable groups is known to affect molecular interactions,[60–62] in all the simulations Glu and Asp residues were assumed to be protonated, while His residues were assumed to be not protonated, as detected experimentally at neutral pH.[63] The system was equilibrated for 500 ps before the production run. Replica exchange attempts were made every 250 ps resulting in a success rate of about 13%. We verified the structures obtained at 278 K with NMR results reported in the literature at similar temperatures;[21,36,64] chemical shifts were back-calculated from the structures using the CamShift method.[65]

Cluster analysis

Twenty thousand structures were taken from a 60–100 ns portion of the trajectory at 309.4 K at 2 ps intervals, and clustered by means of the GROMACS g_cluster tool using the single linkage algorithm, by which a structure is added to a cluster when its RMSD on all Cα atoms to any member of the cluster is less than cut-off of 2 Å. We thus identified 45 clusters with populations ranging from 0.05% to 2%, which were included in the docking analysis. Side-chain contacts were calculated as the pairwise average distance between all the atoms (other than Cα) of a side-chain with that of another.

Fragment-based mapping of binding hot spots

A wide range of approaches are available to perform fragment-based computational mapping of potentially druggable binding hot spots,[66–70] including the GRID method,[71] the multiple copy simultaneous search (MCSS) method,[72] the ROSETTALIGAND method,[73] and the mixed-solvent molecular dynamics (MixMD) method.[74] In this work, we identified binding hotspots of small molecular fragments by combining the FTMap[66] and FRED[75] methods. The FTMap method, which is based on a Fast Fourier Transform (FFT) correlation approach, was used as an initial screen of the 45 structures representative of the corresponding clusters that we identified, and the top ten structures were selected for a further fragment-based docking analysis. In order to perform the docking, among possible alternatives,[42,43] including GLIDE,[44] DOCK,[45] MolDock,[46] and GOLD,[76] we used here FRED,[75] which is a protein structure-based docking program that performs an exhaustive search that systematically samples multiple possible poses to a given resolution, and is thus more computationally expensive than FTMap.

We used a library of ten small organic molecular fragments (benzene, cyclohexane, cyclopropyl, dymethyl ketone, furan, imidazole, methanol, methylamide, oxazole, and pyrazole), which commonly appears in fragment libraries[26,77] and have the hydrophobic character expected to favour the binding to the hydrophobic regions of the Aβ42 peptide. For each fragment, exhaustive docking was performed on the surface of the Aβ42 peptide with a rotation step of 1.25 Å and translation step of 1 Å, and the 10 000 top ranked poses were retained and optimized based on a shape-based Gaussian scoring function.[78] The top 300 poses were then selected based on a consensus score of four scoring functions: Shapegauss, PLP (Piecewise Linear Potential), OEChemscore, and Screenscore, which are implemented in FRED.[75] In order to further optimise the structures, among a wide range of possible alternative methods,[79–84] we used SZYBKI (OpenEye Scientific, www.eyesopen.com) and the Merck Molecular Force Field MMFF94s,[85,86] where the partial charges of ligands were first calculated by Molcharge using the AM1BCC charges (OpenEye Scientific, www.eyesopen.com).

A binding hot spot is defined in this study as a small surface area capable of binding multiple ligand fragments. In order to estimate the quality of a given hot spot we considered the potential ligand efficiency of the fragments that bind to it. The ligand efficiency is defined as[87] F_p/N_p, where N_p is the number of heavy atoms in a ligand probe p and F_p (in kcal/mol) is the binding free energy of the probe. In our calculations we considered the potential energy E_p in the MMFF94s force field,[85] rather than the binding free energy F_p of the probe. This approach represents an approximation, as the binding free energy could not generally be expected to be very accurately approximated by the potential energy of binding.[88] This type strategy was primarily adopted because of its computational efficiency with the aim of generating a small number of candidate fragments and small molecules to facilitate subsequent experimental studies of binding. The validity of this approximation has been discussed in the case of folded proteins, where enthalpic contributions were found to be larger than entropic ones in the binding of small fragments of the type considered here.[89] We should emphasise, however, that the role of entropic contributions in the case of disordered proteins may be greater and will require further studies to be fully clarified.

The normalisation of the binding free energy by the number of heavy atoms has been suggested to be a

useful means of evaluating the quality of hot spots because larger fragments tend to have better binding energy just because of their larger sizes.[87] We then defined the average potential ligand efficiency of a given hot spot as $L_e = \sum_p L_{ep}/N$, where $L_{ep} = E_p/N_p$ is the potential ligand efficiency and N is the total number of ligand probes that bind to the hot spot. More negative potential ligand efficiency values are thus indicative of better binding hot spots. For comparison, we considered the cases of structured proteins, the ATP binding site of p38 MAP kinase and the active site of β-secretase, finding similar potential ligand efficiency values (see Tables S1 and S2 in the supplementary material[90]). Structural representations of structures of the Aβ42 peptide and its ligands were created using PyMOL.[91]

Identification of binding pockets

As is common in fragment-based drug design procedures,[25–28] we identified potential small-molecule binding pockets as clusters of neighbouring binding hot spots.

Docking and molecular dynamics studies of possible binding modes between Aβ42 and curcumin, and Aβ42 and Congo red

Conformers of curcumin and Congo red were first generated by Omega v2.4.3, a multi-conformer structure database generation program by OpenEye Scientific,[92] using a 0.5 Å RMSD cut-off between conformers. Docking at the binding sites of Aβ42 of the two compounds were performed by FRED, using the scoring functions and method mentioned above. For each compound the most highly ranked binding mode was then used as the starting structure in molecular dynamics simulations performed using GROMACS with settings similar to those of the REMD method described above, but in this case at constant temperature (298 K) for 80 ns. We used force field parameters and topologies based on General Amber Force Field (GAFF) and AMBER99SB as prepared by ACPYPE/Antechamber for GROMACS (http://www.ccpn.ac.uk/software/ACPYPE-folder). The system was first energy minimized *in vacuo* and then in TIP3P water, equilibrated by slowly heating from 272 K to 298 K over a 500 ps period, before a production run of 80 ns was carried out. In total, we ran 20 trajectories of 80 ns, one for each of the 10 binding pockets and the 2 small molecules.

RESULTS

Generation and validation of an ensemble of conformations of the Aβ42 peptide

The first step of the procedure that we discuss in this work is the generation of an ensemble of conformations that represents the soluble monomeric form of the Aβ42 peptide (Figure 1(a)), which was carried out by replica-exchange molecular dynamics (REMD) simulations[53] in explicit solvent (see Methods section). In order to establish whether the Aβ42 structural ensemble generated by the procedure described above provides a good representation of the conformations that this peptide populates in solution, we

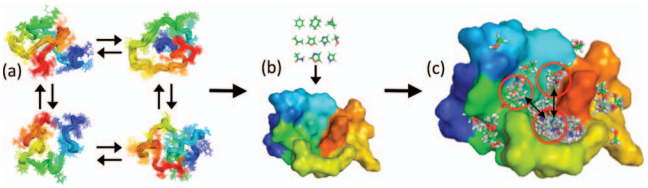


FIG. 1. Scheme illustrating the strategy discussed in this work in which molecular dynamics simulations are combined with computer-based fragment-based hot spot mapping to identify potential binding sites on the soluble monomeric form of the Aβ42 peptide. (a) Representative structures are selected by a clustering procedure within an ensemble of conformations representing the natively unfolded state of the Aβ42 peptide. (b) Hot spot regions are mapped on these structures using a set of small molecule fragments. (c) Neighbouring hot spot regions are identified as potential small-molecule binding sites.

investigated the agreement between various experimental and back-calculated structural parameters, where the latter were obtained from the ensemble of structures. The convergence of the simulations was monitored by following the evolution of the correlation between experimental and calculated chemical shifts of the Cα atoms[21] (Figure 2(a)), as well as the evolution of radius of gyration (Figure 2(b)) and solvent accessible surface area (Figure 2(c)). This analysis was performed for the conformations determined at 278 K, the temperature at which the experimental chemical shifts that we considered were measured[21] (Figures 3(a)–3(c)). After convergence, we found good correlations between experimental and back-calculated chemical shifts (the coefficients of correlation were 0.987 for Cα, 0.822 for Hα, and 0.796 for N). We also compared experimental and back-calculated ^3J-couplings,[36] as well as residual dipolar couplings (RDCs),[64] finding a good agreement also in these cases (Figures 3(d) and 3(e)). In particular, the level of such an agreement was found to be higher than that provided by the statistical coil model (SCM, Figures 3(d) and 3(e)), which has been found to describe accurately the dimension and structures populated by highly disordered states of proteins;[49] for the SCM ensemble and the present ensemble the RMSD values were, respectively, 1.22 Hz and 0.82 Hz for the ^3J-couplings, and 2.91 Hz and 2.07 Hz for the RDCs.

Structural analysis of the Aβ42 ensemble

The analysis of the results of the simulations indicates that, under the conditions that we have investigated, Aβ42 populates a restricted but highly dynamical ensemble of conformations that is significantly more compact than that expected for a random coil,[49] in agreement with previous conclusions on this system.[21,93] In addition, the average value of the radius of gyration of the Aβ42 peptide in our simulations is of about 13 Å, compared to the SCM value of about 20 Å. The presence of transient structural motifs in Aβ42 is also reflected by the differences between the present results and those obtained by SCM (Figures 3(d) and 3(e)).

Analysis of the inter-residue distances in the ensemble of conformations that we have generated here (Figure 3(f)) suggests that the overall structure and dynamics of the

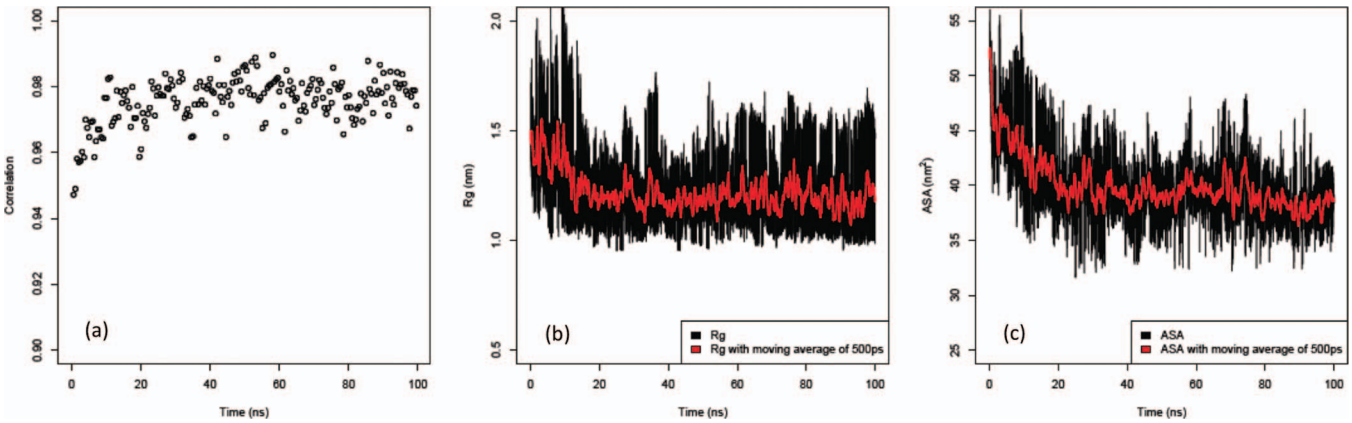


FIG. 2. Analysis of the convergence of the molecular dynamics simulations used in this work to generate an ensemble of structures representing the soluble monomeric form of the Aβ42 peptide: (a) Time series of the correlation between experimental and calculated Cα chemical shifts, which indicate that after about 40 ns (out of a total of 100 ns) a good correlation is reached between experimental and calculated chemical shifts. (b) Time series of the radius of gyration. (c) Time series of the solvent accessible surface area (SASA).

Aβ42 peptide are particularly strongly affected by the behaviour of specific regions of the amino acid sequence that have a high tendency to form turns, in particular Asp7-Tyr10, Asp23-Ser26, and Gly37-Val40, as well as by the interactions between the two main hydrophobic regions (residues Leu17-Ala21 and Ile31-Val36) of the polypeptide chain. In particular, the combination of this latter tertiary contact, which is observed in several clusters, with the transient formation of a turn in the Asp23-Ser26 region, appears to be the main driving factors for the quasi-hairpin-like structures often reported for the Aβ42 peptide, where the turn formation could be assisted by electrostatic interactions between Glu22, Asp23, and Lys28. These findings are in good agreement with structural insights drawn previous from nuclear Overhauser enhancement (NOE) data[21] as well as from molecular dynamics simulations.[35,37]

We next used a cluster analysis to find families of similar structures in the Aβ42 ensemble (see Methods section).

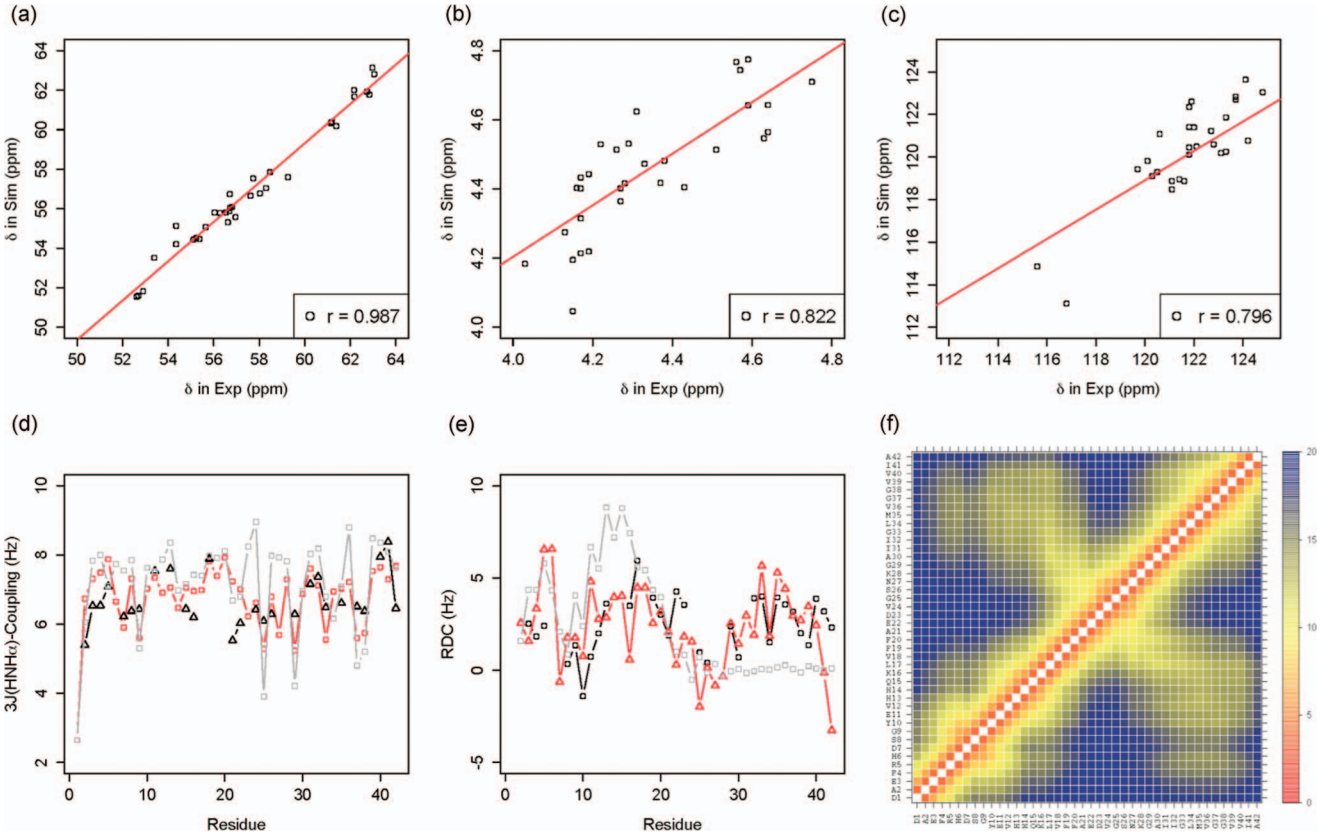


FIG. 3. Validation of the ensemble of conformations representing the soluble monomeric form of the Aβ42 peptide used in this work. (a)–(c) Correlation between experimental[21] and back-calculated chemical shifts: Cα (a), Hα (b), and N (c). (d) Comparison between experimental[36](black) and back-calculated (red) ^3J couplings (Hz). (e) Comparison between experimental[64] (black) and back-calculated (red) residual dipolar couplings (RDCs, Hz). For reference, ^3J couplings and RDCs are also shown as predicted by the statistical coil model[49] (grey). (f) Inter-residue distance map (Å).

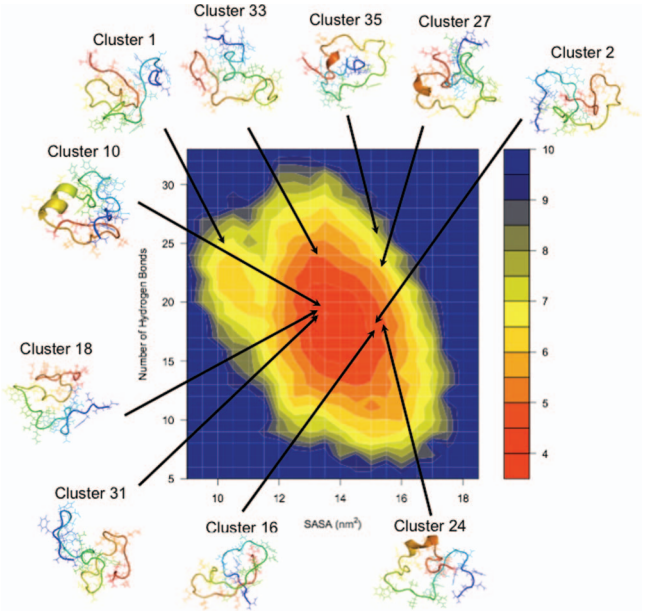


FIG. 4. Free energy landscape of the Aβ42 peptide as a function of the number of hydrogen bonds (backbone-backbone, backbone-sidechain, and sidechain-sidechain) and of the solvent-exposed surface area of hydrophobic residues. Hydrogen bonds were defined using the GROMACS g_hbond function, when hydrogen donors and acceptors are within 3.5 Å and the hydrogen-donor-acceptor angles are within 30°. The most populated clusters are found in different regions of the free energy landscape.

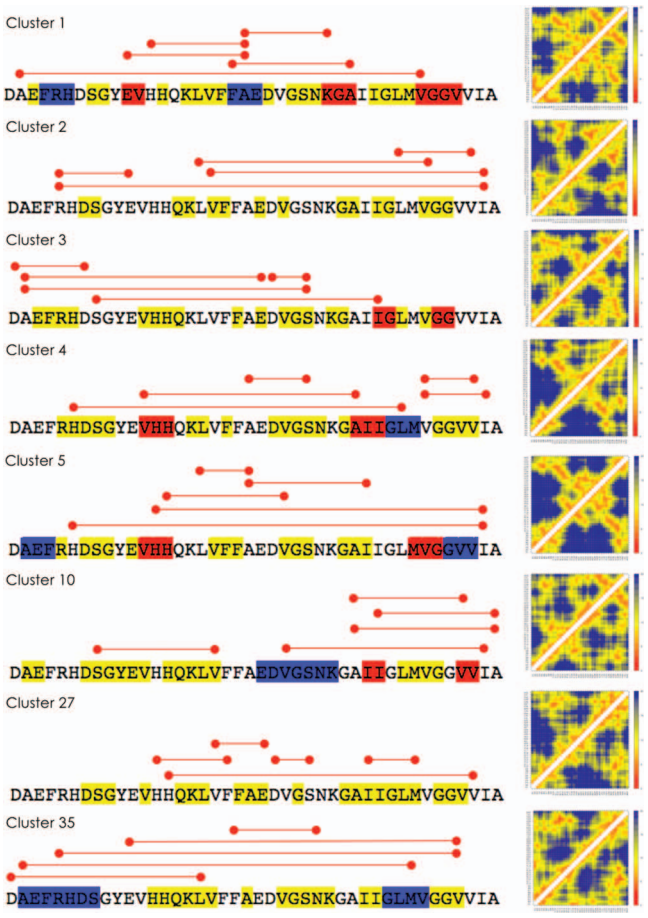


FIG. 5. Characterisation of eight representative clusters of structures within the ensemble of conformations of the soluble monomeric form of the Aβ42 peptide used in this work. We consider here the five most populated clusters (cluster 1–5) together with three more examples of clusters found to contain binding pockets (clusters 10, 27, and 35). Highly populated clusters may (as clusters 1 and 2) or may not (as clusters 3, 4, and 5) exhibit binding pockets. For each cluster we report the secondary structure elements determined by DSSP[106] (yellow: turns; blue: α-helices; red: β-sheets) and the side-chain distance maps. The five shortest long-range side-chain contacts (i.e., more than three residues apart along the amino acid sequence) are indicated by red lines.

This procedure resulted in 396 clusters, the 45 most populated of which were selected for further analysis. Taken together, these 45 clusters include about 67% of the members of the structural ensemble, illustrating its heterogeneity; the most populated clusters are shown in Figure 4 on a free energy landscape plotted as a function of the number of hydrogen bonds (backbone-backbone, backbone-sidechain, and sidechain-sidechain) and of the solvent-exposed surface area of hydrophobic residues. The secondary structure elements, side-chain distance maps, and long-range contacts most frequently observed in eight of the most populated clusters are shown as examples in Figure 5.

The features in these individual distance maps of clusters resemble in part the overall features exhibited in the average distance map (Figure 3(f)). For example, in cluster 1 the turns around Ser8-Gly9 and Gly25-Asn27 are very close to the corresponding regions of the first two turns identified in the general distance map, whereas cluster 2 exhibits the very prominent turn around residues Asp7-Try10 in the N-terminal region. Further, the characteristic contacts between the two hydrophobic regions (residues Leu17-Ala21 and Ile31-Val36) are also noticeable in all the distance maps shown in Figure 5. The identification of geometrically similar families of conformations of the Aβ42 peptide suggests that distinct sets of structurally related conformations exist in the ensemble.

Fragment-based hot spot mapping

Individual representative structures were selected from the top 45 clusters and used to identify binding hot spots using FTMap calculations[66] (see Methods section). The ten structures found to contain multiple binding hot spots in close proximity to each other were selected for further analysis (Table I and Figure 6) and used in an exhaustive rigid-body docking procedure using FRED[75] (see Methods section). Two representative structures are shown in Figure 7, where key residues forming the hot spots and the corresponding potential ligand efficiency values[87] (L_e, see Methods section) are listed.

We then examined those hot spots that were found to bind three or more different fragments. These hot spots exhibited potential ligand efficiency values (L_e) ranging from −0.8 to −1.5 kcal/mol, which are comparable to the ones that we observed for the model globular proteins that we studied (−1.3 to −1.5 kcal/mol, see Methods section and Tables S1 and S2 in the supplementary material[90]), and to those reported in the literature.[94] Consistent with previous observations,[89] we also found the number of hydrogen bonds formed by the fragments to be comparable to the number typically found in folded proteins, thus providing insight into the origin of the enthalpic contributions to binding (Figure 8). On average, we found a

TABLE I. List of the ten binding pockets (in roman numerals, column 2) and corresponding binding hot spots (in arabic numerals, column 3) identified within ten specific clusters of conformations (column 1) in the Aβ42 structural ensemble described in this work. Each of these ten clusters exhibits one binding pocket comprising between two and five binding hot spots; for example binding pocket VI is found in cluster 24 and comprises four hot spots. The remaining 35 clusters among the 45 that we analysed in detail did not exhibit binding pockets. The specific residues in the hot spots are also reported (column 4). The structures of the ten binding pockets are shown in Figure 6.

Cluster	Pocket ID	Hot spot ID	Key residues involved
Cluster 1	I	1	Leu17,Val18, Phe20
Cluster 1	I	2	Ala30, Gly33, Lys16, Gly33
Cluster 2	II	1	His6, His13, Leu17, Phe19
Cluster 2	II	2	His13, Leu17, Val36
Cluster 2	II	3	Tyr10, His13, Met35
Cluster 10	II	1	Leu17, Phe19, Phe20, Ile41, Val40
Cluster 10	II	2	His6, Phe20, Val39, Val40, Val18
Cluster 16	IV	1	Glu11, His13, Lys16, Leu17
Cluster 16	IV	2	Arg5, Asp7, Glu11, Lys16, Val24, Val36
Cluster 16	IV	3	Ala21, Asp23, Val24
Cluster 18	V	1	Leu17, His6, Ile3, Tyr10, Gly25, Asn27
Cluster 18	V	2	Lys16, Leu17, Phe19, Val24, Gly25
Cluster 18	V	3	Phe19, Glu22, Asp23, Val24
Cluster 18	V	4	His6, Try10, Ile31, Arg5
Cluster 18	V	5	Val24, Ile31, Gly33
Cluster 24	VI	1	His13, His14, Val18, Val24, Val39
Cluster 24	VI	2	His14, Met35, Val39, Val40
Cluster 24	VI	3	Val18, Phe19, Asp23
Cluster 24	VI	4	Val18, Asp23, Met35
Cluster 27	VI	1	Lys28, Val39, Ile31, Met35
Cluster 27	VI	2	Val13, His13, Val18, Asp23, Ser26, Lys28, Val39
Cluster 27	VI	3	His14, Leu17, Val39, Val40
Cluster 27	VI	4	Tyr10, Val12, Phe20
Cluster 31	VII	1	Lys16, Val18, Ile32, Leu34, Gln15
Cluster 31	VII	2	His14, Gln15, Lys16
Cluster 31	VII	3	His14, Ile41, Gln15
Cluster 33	IX	1	Gln15, Leu17, Gly33
Cluster 33	IX	2	Gln15, Leu17, Ser8, Gly9
Cluster 33	IX	3	Asp7, Ser8, Val18, Leu34
Cluster 35	X	1	Ala2, Leu17, Asn27, Met3, Val36
Cluster 35	X	2	Leu17, Phe19, Phe20
Cluster 35	X	3	Ile31, Val36, Val40

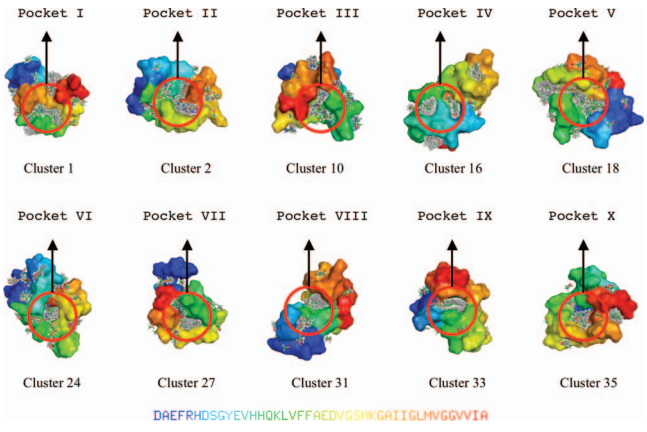


FIG. 6. Illustration of the ten binding pockets identified by fragment-probe mapping in the ten most populated clusters (see Table I) within the Aβ42 structural ensemble used in this work.

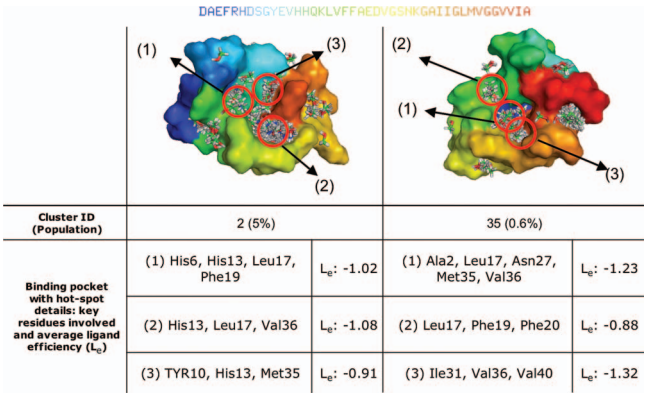


FIG. 7. Examples of adjacent binding pockets in the soluble monomeric form of the Aβ42 peptide identified through the approach described in this work. Results for clusters 2 and 35 (see Fig. 4) are shown together with a characterisation of the corresponding hot spots, the potential ligand efficiency (L_e) (in kcal/mol, see Methods section).

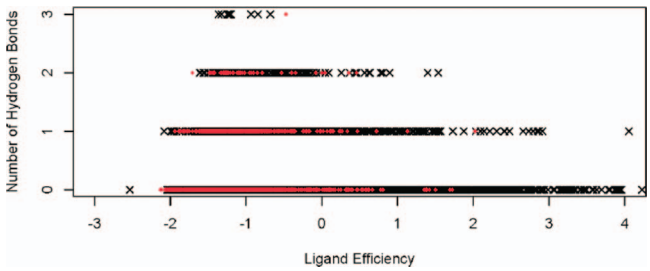


FIG. 8. Comparison between the potential ligand efficiency (L_e) (x-axis, in kcal/mol, see Methods section) and the number of hydrogen bonds (y axis) for all the poses of the fragments in the binding hot spots identified in this work within the Aβ42 structural ensemble (red circles) and those in model globular proteins (black crosses, see Methods section and Tables S1 and S2 in the supplementary material[90]); the red circles correspond to the hot-spot IDs listed in Table I.

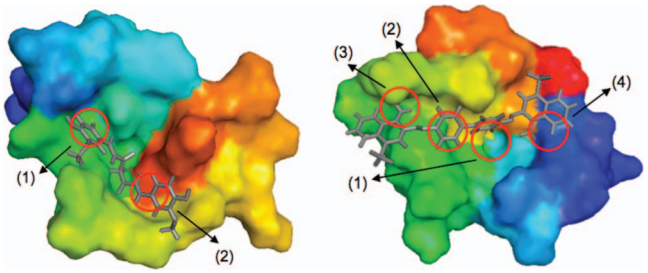


FIG. 10. Top binding modes of curcumin (left) and Congo red (right) with the Aβ42 peptide, which were identified through the analysis of the fragment-based mapping of the binding hot spots; hot spot labels refer to Table I and Figure 6 (pocket II in cluster 2 for curcumin and pocket V in cluster 18 for Congo red). Molecular dynamics simulations of the complexes show that the ligands remain bound over a 80 ns period.

value of slightly less than 1 hydrogen bond per fragment per hot spot for the Aβ42 peptide, which is similar to that observed for fragments of the same size extracted from high resolution structures in the PDB.[89]

Further examination of these results revealed that the Aβ42 peptide exhibits regions of the amino acid sequence with different propensity to bind small molecular fragments (Figure 9). In particular, the central hydrophobic cluster (CHC) region (residues Leu17-Ala21) has a high propensity to form binding hot spots and to bind small molecular fragments. We found that residues Phe4, Tyr10, Leu17, Phe19, Ile31, and Met35 are involved in many of the hot spots identified in the mapping. A list of the binding sites, corresponding hot spots and key residues involved is provided in Table I. It is interesting to note that the type of residues (in particular Phe, Tyr, Leu, Ile, and Met) involved in forming binding sites identified here for an intrinsically disordered peptide correspond quite closely with those that have been described for structured proteins.[95,96]

Small-molecule interactions with potential binding pockets

We identified potential binding pockets by clustering neighbouring binding hot spots. To examine the significance of these potential binding pockets, we performed docking studies of two compounds, curcumin and Congo red, that have been shown to inhibit the aggregation of the Aβ peptide.[97,98] Although the mechanism of action of these compounds on the behaviour of the Aβ peptide is still unclear, there is evidence suggesting that the peptide may bind to them either individually or in small oligomeric assemblies formed through detergent-like interactions.[97–99]

We performed docking calculations for the two compounds for possible binding modes in a systematic manner within the binding pockets that we identified in this work (Table I and Figure 6). The resulting top binding modes exhibit peptide-ligand interaction energies comparable with those seen in studies of globular proteins (see Methods section). Two examples of such top binding modes are illustrated in Figure 10. We found that the aromatic rings of these compounds play an important role in binding to the hot spots, in particular through ring stacking interactions with the side-chains of residues in the hot spots themselves (e.g., Phe4, His6, Tyr10, Phe19, and Phe20). These interactions have been suggested as being an important feature in other small molecules capable of binding the Aβ peptide, such as polyphenols,[100] catechins,[101] and ketones.[102]

We then carried out molecular dynamics simulations in order to probe the tendency of the small molecules to remain bound to the identified binding pockets (see Methods section). These simulations revealed that at least in some of these complexes the ligands remain bound to the Aβ42 peptide over a period of 80 ns at room temperature. The results of these simulations indicate that the interactions of the ligands with many

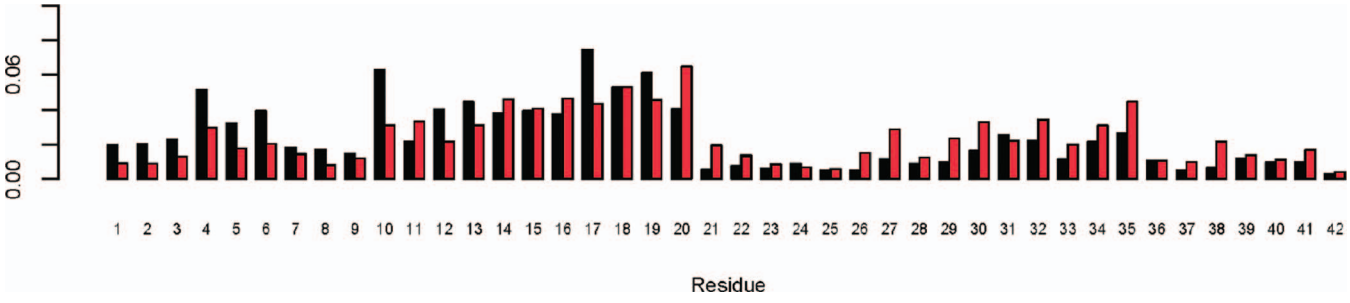


FIG. 9. Residue-specific probability of binding small molecular fragments in hot spots of the Aβ42 peptide, calculated by FTMap (see Methods section). Non-bonded (black bars) and hydrogen bond (red bars) interactions are shown separately. The central hydrophobic region (CHC, residues Leu17-Ala21) is particular involved in hot spot formation.

of the residues identified in the hot spots are key to strong binding.

DISCUSSION AND CONCLUSIONS

The drug discovery approach that we have described in this work is based on the idea of extending to intrinsically disordered proteins the well-established observation that a significant proportion of the free energy of binding in conventional protein-ligand complexes derives from relatively small regions of the protein surface, known as hot spots.[103] Ligands that bind simultaneously to multiple hot spots could result in higher binding affinities and better specificity. In the case of intrinsically disordered peptides and proteins, such as the Aβ42 peptide considered here, the additional complication is that these systems do not populate a small number of specific conformations, but rather experience conformational fluctuations of large amplitude. In these cases, the existence of hot spots requires careful examination, and so the aim of this work has been to explore this idea and identify compounds capable of binding to specific pockets in particular conformations. We have therefore carried out screens of small molecular fragments by identifying a series of representative highly populated clusters of conformations within the Aβ42 structural ensemble.

The concept that intrinsically disordered proteins are potentially druggable is relatively recent.[6–10] Our results indicate that the conformational space populated by the Aβ42 peptide may contain specific structures with significant statistical weights, and that such conformations may contain binding pockets that can be targeted by small molecules. The overall dimension of the Aβ42 peptide in its monomeric form in solution is rather more compact than a random coil, and transient hydrophobic pockets exist most likely as a result of the high propensity of certain regions to form turns and of interactions between hydrophobic regions in the amino acid sequence of the peptide. The compactness of the structures and the long-range contacts within them are important for forming potential binding pockets because they provide the environment for favourable hydrogen bond, electrostatic, hydrophobic or van der Waals interactions as well as shape complementarity with a ligand. One could also expect the type of binding pockets that we have identified here for the monomeric form of the Aβ42 peptide to appear in more structured assemblies formed by this peptide, including oligomeric, membrane-bound, and fibrillar conformations. It is possible that small molecules designed to bind the monomeric form would also, and less transiently, bind such assemblies, as their more ordered nature could reduce the entropic penalty of binding.

We found that hot spot formation is assisted particularly by the N-terminal and CHC regions, as illustrated in Figure 9. Phe4 and Tyr10 are closely involved in hot spot formation, together with Leu17, Phe19, Ile31, and Met35, which interact with each other to form favourable pockets that could be suitable for binding small molecules. Two of the residues that we identified as involved in hot spots formation, Phe19 and Met35, are known to be particularly important in the aggregation process of the Aβ42 peptide, as mutation of either of the two residues has been shown to inhibit oligomer and fibril formation.[104,105] It is intriguing to speculate that finding brain-penetrable small molecules that could bind to a pocket formed by these residues may have a significant effect in inhibiting the aggregation of the Aβ42 peptide.

From a methodological point of view, the identification of clusters of conformations populated by the Aβ42 peptide in its monomeric form in solution allows the number of structures to be searched to find potential binding sites to be significantly reduced. Such reduction in search space is crucial since the screening procedure is computationally costly.

The work presented here represents an initial step toward targeting the Aβ42 peptide in its monomeric form, by demonstrating that it exhibits potential small molecule binding sites. One development of this approach will be to improve the accuracy of the Aβ42 structural ensemble by incorporating experimental data in the molecular dynamics simulations in a way similar to that used for example for α-synuclein.[18,19,51] Using the potential binding sites identified on different representative structures, the next step of this drug design strategy will be to conduct a structure based high-throughput docking screening of small-molecules to these, and verification of *in silico* hits by *in vitro* and *in vivo* studies.

ACKNOWLEDGMENTS

This work was supported by grants from Alzheimer's Research UK (M.Z.) and the Biotechnology and Biological Sciences Research Council (BBSRC) (C.M.D. and M.V.). We are grateful to Tobin Sosnick and his group for providing the SCM structures of the Aβ42 peptide.

[1] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[2] C. Haass and D. J. Selkoe, Nat. Rev. Mol. Cell Biol. **8**, 101 (2007).
[3] M. Prince, R. Bryce, and C. Ferri, Alzheimer's Disease International: World Alzheimer Report 2011.
[4] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, Science **319**, 916 (2008).
[5] C. M. Dobson, Science **304**, 1259 (2004).
[6] Y. Cheng, T. LeGall, C. J. Oldfield, J. P. Mueller, Y. Y. J. Van, P. Romero, M. S. Cortese, V. N. Uversky, and A. K. Dunker, Trends Biotech. **24**, 435 (2006).
[7] M. Citron, Nat. Rev. Drug Disc. **9**, 387 (2010).
[8] A. K. Dunker and V. N. Uversky, Curr. Opin. Pharmacol. **10**, 782 (2010).
[9] S. J. Metallo, Curr. Opin. Chem. Biol. **14**, 481 (2010).
[10] P. Tompa, Curr. Opin. Struct. Biol. **21**, 419 (2011).
[11] G. Yamin, K. Ono, M. Inayathullah, and D. B. Teplow, Curr. Pharm. Des. **14**, 3231 (2008).
[12] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. S. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani, Nature (London) **416**, 507 (2002).
[13] N. Cremades, S. I. A. 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. J. Knowles, C. M. Dobson, and D. Klenerman, Cell **149**, 1048 (2012).
[14] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky, J. Mol. Biol. **362**, 1043 (2006).
[15] Z. Dosztanyi, B. Meszaros, and I. Simon, Bioinformatics **25**, 2745 (2009).
[16] M. R. Arkin and J. A. Wells, Nat. Rev. Drug Disc. **3**, 301 (2004).
[17] P. J. Whittle and T. L. Blundell, Annu. Rev. Biophys. Biomol. Struct. **23**, 349 (1994).
[18] C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, and M. Zweckstetter, Proc. Natl. Acad. Sci. U.S.A. **102**, 1430 (2005).

[19]M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, J. Am. Chem. Soc. **127**, 476 (2005).
[20]M. D. Mukrasch, P. Markwick, J. Biernat, M. von Bergen, P. Bernado, C. Griesinger, E. Mandelkow, M. Zweckstetter, and M. Blackledge, J. Am. Chem. Soc. **129**, 5235 (2007).
[21]L. M. Hou, H. Y. Shao, Y. B. Zhang, H. Li, N. K. Menon, E. B. Neuhaus, J. M. Brewer, I. J. L. Byeon, D. G. Ray, M. P. Vitek, T. Iwashita, R. A. Makula, A. B. Przybyla, and M. G. Zagorski, J. Am. Chem. Soc. **126**, 1992 (2004).
[22]S. Zhang, K. Iwata, M. J. Lachenmann, J. W. Peng, S. Li, E. R. Stimson, Y. Lu, A. M. Felix, J. E. Maggio, and J. P. Lee, J. Struct. Biol. **130**, 130 (2000).
[23]O. S. Makin and L. C. Serpell, FEBS J. **272**, 5950 (2005).
[24]R. Tycko, Annu. Rev. Phys. Chem. **62**, 279 (2011).
[25]P. J. Hajduk and J. Greer, Nat. Rev. Drug Disc. **6**, 211 (2007).
[26]M. Congreve, G. Chessari, D. Tisi, and A. J. Woodhead, J. Med. Chem. **51**, 3661 (2008).
[27]C. W. Murray and D. C. Rees, Nat. Chem. **1**, 187 (2009).
[28]C. W. Murray and T. L. Blundell, Curr. Opin. Struct. Biol. **20**, 497 (2010).
[29]K. A. Ball, A. H. Phillips, P. S. Nerenberg, N. L. Fawzi, D. E. Wemmer, and T. Head-Gordon, Biochemistry **50**, 7612 (2011).
[30]Y. Chebaro, N. Mousseau, and P. Derreumaux, J. Phys. Chem. B **113**, 7668 (2009).
[31]M. Convertino, A. Vitalis, and A. Caflisch, J. Biol. Chem. **286**, 41578 (2011).
[32]M. G. Krone, A. Baumketner, S. L. Bernstein, T. Wyttenbach, N. D. Lazo, D. B. Teplow, M. T. Bowers, and J. E. Shea, J. Mol. Biol. **381**, 221 (2008).
[33]A. R. Lam, D. B. Teplow, H. E. Stanley, and B. Urbanc, J. Am. Chem. Soc. **130**, 17413 (2008).
[34]S. Mitternacht, I. Staneva, T. Hard, and A. Irback, Proteins **78**, 2600 (2010).
[35]N. G. Sgourakis, M. Merced-Serrano, C. Boutsidis, P. Drineas, Z. M. Du, C. Y. Wang, and A. E. Garcia, J. Mol. Biol. **405**, 570 (2011).
[36]N. G. Sgourakis, Y. L. Yan, S. A. McCallum, C. Y. Wang, and A. E. Garcia, J. Mol. Biol. **368**, 1448 (2007).
[37]B. Urbanc, L. Cruz, S. Yun, S. V. Buldyrev, G. Bitan, D. B. Teplow, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **101**, 17345 (2004).
[38]A. Vitalis and A. Caflisch, J. Mol. Biol. **403**, 148 (2010).
[39]F. E. Herrera, A. Chesi, K. E. Paleologou, A. Schmid, A. Munoz, M. Vendruscolo, S. Gustincich, H. A. Lashuel, and P. Carloni, PLoS ONE **3**, e3394 (2008).
[40]R. Cuchillo and J. Michel, Biochem. Soc. Trans. **40**, 1004 (2012).
[41]J. H. Lin, A. L. Perryman, J. R. Schames, and J. A. McCammon, J. Am. Chem. Soc. **124**, 5632 (2002).
[42]D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, Nat. Rev. Drug Disc. **3**, 935 (2004).
[43]G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head, J. Med. Chem. **49**, 5912 (2006).
[44]R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, J. Med. Chem. **47**, 1739 (2004).
[45]T. J. A. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz, J. Comp. Aid. Mol. Des. **15**, 411 (2001).
[46]R. Thomsen and M. H. Christensen, J. Med. Chem. **49**, 3315 (2006).
[47]W. Sinko, S. Lindert, and J. A. McCammon, Chem. Biol. Drug Des. **81**, 41 (2013).
[48]K. W. Kaufmann and J. Meiler, PLoS ONE **7**, e50769 (2012).
[49]A. K. Jha, A. Colubri, K. F. Freed, and T. R. Sosnick, Proc. Natl. Acad. Sci. U.S.A. **102**, 13099 (2005).
[50]V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernado, C. Charavay, and M. Blackledge, Bioinformatics **28**, 1463 (2012).
[51]J. R. Allison, P. Varnai, C. M. Dobson, and M. Vendruscolo, J. Am. Chem. Soc. **131**, 18314 (2009).
[52]A. Cavalli, C. Camilloni, and M. Vendruscolo, J. Chem. Phys. **138**, 094112 (2013).
[53]U. H. E. Hansmann and Y. Okamoto, Curr. Opin. Struct. Biol. **9**, 177 (1999).
[54]D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comp. Chem. **26**, 1701 (2005).
[55]A. De Simone and P. Derreumaux, J. Chem. Phys. **132**, 165103 (2010).
[56]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).
[57]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[58]R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009).
[59]S. A. Showalter and R. Bruschweiler, J. Am. Chem. Soc. **129**, 4158 (2007).
[60]A. H. Elcock and J. A. McCammon, Bioph. J. **80**, 613 (2001).
[61]H. Park and S. Lee, J. Am. Chem. Soc. **125**, 16416 (2003).
[62]T. ten Brink and T. E. Exner, J. Chem. Inf. Mod. **49**, 1535 (2009).
[63]F. Newby, Ph.D. thesis, University of Cambridge, 2012.
[64]K. H. Lim, G. L. Henderson, A. Jha, and M. Louhivuori, ChemBioChem **8**, 1251 (2007).
[65]K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, and M. Vendruscolo, J. Am. Chem. Soc. **131**, 13894 (2009).
[66]R. Brenke, D. Kozakov, G. Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos, and S. Vajda, Bioinformatics **25**, 621 (2009).
[67]A. Ivetac and J. A. McCammon, Chem. Biol. Drug Des. **76**, 201 (2010).
[68]L. Hoffer, J. P. Renaud, and D. Horvath, Comb. Chem. High Throughput Screening **14**, 500 (2011).
[69]K. W. Lexa and H. A. Carlson, Q. Rev. Bioph. **45**, 301 (2012).
[70]A. Kumar, A. Voet, and K. Y. J. Zhang, Curr. Med. Chem. **19**, 5128 (2012).
[71]P. J. Goodford, J. Med. Chem. **28**, 849 (1985).
[72]A. Miranker and M. Karplus, Proteins **11**, 29 (1991).
[73]J. Meiler and D. Baker, Proteins **65**, 538 (2006).
[74]K. W. Lexa and H. A. Carlson, J. Am. Chem. Soc. **133**, 200 (2011).
[75]M. McGann, J. Chem. Inf. Mod. **51**, 578 (2011).
[76]G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, J. Mol. Biol. **267**, 727 (1997).
[77]S. M. Boyd, A. P. Turnbull, and B. Walse, WIREs Comput. Mol. Sci. **2**, 868 (2012).
[78]M. R. McGann, H. R. Almond, A. Nicholls, J. A. Grant, and F. K. Brown, Biopolymers **68**, 76 (2003).
[79]Y. Deng and B. Roux, J. Phys. Chem. B **113**, 2234 (2009).
[80]E. Gallicchio, M. Lapelosa, and R. M. Levy, J. Chem. Theory Comp. **6**, 2961 (2010).
[81]D. L. Mobley and K. A. Dill, Structure **17**, 489 (2009).
[82]H. S. Muddana, C. D. Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe, and M. K. Gilson, J. Comp. Aid. Mol. Des. **26**, 475 (2012).
[83]J. Wereszczynski and J. A. McCammon, Q. Rev. Biophys. **45**, 1 (2012).
[84]E. O. Purisima and H. Hogues, J. Phys. Chem. B **116**, 6872 (2012).
[85]T. A. Halgren, J. Comp. Chem. **20**, 720 (1999).
[86]S. Wlodek, A. G. Skillman, and A. Nicholls, J. Chem. Theory Comp. **6**, 2140 (2010).
[87]A. L. Hopkins, C. R. Groom, and A. Alex, Drug. Disc. Today **9**, 430 (2004).
[88]C. H. Reynolds and M. K. Holloway, ACS Med. Chem. Lett. **2**, 433 (2011).
[89]G. G. Ferenczy and G. M. Keseru, J. Chem. Inf. Mod. **52**, 1039 (2012).
[90]See supplementary material at http://dx.doi.org/10.1063/1.4811831 for the results on the potential ligand efficiency values.
[91]W. L. DeLano, *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, USA., 2002).
[92]J. Kirchmair, G. Wolber, C. Laggner, and T. Langer, J. Chem. Inf. Mod. **46**, 1848 (2006).
[93]J. Danielsson, J. Jarvet, P. Damberg, and A. Graslund, Mag. Res. Chem. **40**, S89 (2002).
[94]I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman, Proc. Natl. Acad. Sci. U.S.A. **96**, 9997 (1999).
[95]B. Y. Ma, T. Elkayam, H. Wolfson, and R. Nussinov, Proc. Natl. Acad. Sci. U.S.A. **100**, 5772 (2003).
[96]O. Keskin, B. Y. Ma, and R. Nussinov, J. Mol. Biol. **345**, 1281 (2005).
[97]C. Lendel, B. Bolognesi, A. Wahlstrom, C. M. Dobson, and A. Graslund, Biochemistry **49**, 1358 (2010).
[98]F. S. Yang, G. P. Lim, A. N. Begum, O. J. Ubeda, M. R. Simmons, S. S. Ambegaokar, P. P. Chen, R. Kayed, C. G. Glabe, S. A. Frautschy, and G. M. Cole, J. Biol. Chem. **280**, 5892 (2005).
[99]M. O. Pedersen, K. Mikkelsen, M. A. Behrens, J. S. Pedersen, J. J. Enghild, T. Skrydstrup, A. Malmendal, and N. C. Nielsen, J. Phys. Chem. B **114**, 16003 (2010).
[100]Y. Porat, A. Abramowitz, and E. Gazit, Chem. Biol. Drug Des. **67**, 27 (2006).

[101] D. E. Ehrnhoefer, J. Bieschke, A. Boeddrich, M. Herbst, L. Masino, R. Lurz, S. Engemann, A. Pastore, and E. E. Wanker, Nat. Struct. Mol. Biol. **15**, 558 (2008).
[102] X. Y. Zheng, M. M. Gessel, M. L. Wisniewski, K. Viswanathan, D. L. Wright, B. A. Bahr, and M. T. Bowers, J. Biol. Chem. **287**, 6084 (2012).
[103] T. Clackson and J. A. Wells, Science **267**, 383 (1995).
[104] S. L. Bernstein, T. Wyttenbach, A. Baumketner, J. E. Shea, G. Bitan, D. B. Teplow, and M. T. Bowers, J. Am. Chem. Soc. **127**, 2075 (2005).
[105] G. Bitan, B. Tarus, S. S. Vollers, H. A. Lashuel, M. M. Condron, J. E. Straub, and D. B. Teplow, J. Am. Chem. Soc. **125**, 15359 (2003).
[106] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).