Investigation of CMA Substrate Interactions with Hsc70 Reveals Potential Auxiliary Binding Sites

Devid Sahu[1], Nidhi Malhotra[1]*

[1] Department of Chemistry, School of Natural Sciences, Shiv Nadar Institution of Eminence, Gautam Buddha Nagar, Uttar Pradesh, India

*Corresponding author

Email: nidhi.malhotra@snu.edu.in; ORCID: 0000-0002-0218-3440

Short title: ***CMA Substrate Binding with Hsc70***

Abstract

Chaperone-mediated autophagy (CMA) is a selective lysosomal degradation pathway crucial for cellular homeostasis. Its dysfunction is linked to cancer and neurodegenerative disorders. Hsc70 mediates CMA selectivity by recognizing proteins with 'KFERQ-type' sequences, but the structural basis for substrate recognition and the substrate-binding mechanism of Hsc70 remain unclear. This study integrates literature, bioinformatics, AlphaFold 3 predictions, and 44 μs of molecular dynamics (MD) simulations to elucidate the lid domain interactions of Hsc70 with fifteen experimentally verified CMA substrate segments. Structural and thermodynamic residue-level analyses suggest stable binding of substrates in two distinct sites. The first site resides in a hydrophobic cavity formed by alpha helix α1 and α3 in the lid domain of Hsc70, with complex stabilization primarily driven by hydrophobic residues within 'KFERQ-type' motif and its flanking regions. The second stable site was observed to be canonical hydrophobic cleft in Hsc70, with complex stabilization primarily driven by hydrophobic residues within 'KFERQ-type' motif as well as the through water-mediated interactions by the lid domain. The robustness of our findings is confirmed by consistent convergence across force fields, substrates, and supporting literature, all identifying the same site. We propose these sites may facilitate upstream or downstream recognition and activation events.

Author Summary

Cells must constantly remove damaged or unnecessary proteins in order to stay healthy. One important pathway that performs this task is chaperone-mediated autophagy. When this pathway fails, it has been linked to conditions such as cancer and neurodegenerative diseases. A key player in this process is the protein Hsc70, which recognizes short sequence signatures in target proteins and directs them for degradation. However, how Hsc70 physically recognizes these signatures has

remained poorly understood. In this study, we investigated how Hsc70 interacts with a range of confirmed target protein segments. We combined information from prior experimental studies, with computational structure prediction and extensive molecular simulations to explore these interactions in detail. Our results indicate that Hsc70 can bind target proteins at two distinct regions, each stabilized mainly by hydrophobic amino acids within and/or around the recognition signal, with additional contribution from water-mediated interactions in the latter site. Notably, these binding modes were consistent across different targets and computational conditions. Together, our findings provide a clearer structural picture of how selectivity is achieved in this degradation pathway. This improved understanding may help guide future studies on how chaperone-mediated autophagy is regulated and how it becomes disrupted in disease.

Introduction

Chaperone-Mediated Autophagy (CMA) is a selective degradation process with diverse roles in maintaining cellular homeostasis, regulating immune response, influencing aging, and contributing to the pathogenesis of various diseases, including cancer[1–5]. The unique feature of selectivity distinguishes it from other types of autophagy. The selectivity is rendered by a heat shock cognate protein, Hsc70, that recognizes only a cohort of selective proteins called substrates and only such proteins are amenable to degradation by CMA. Despite its well-established function in recognizing and binding substrate proteins, the precise molecular details governing the interaction between Hsc70 and its substrate proteins remain largely unknown. Sequence-based designation of the substrate requires presence of 'KFERQ-like' (CMA-targeting) motifs in them[6–9]. This motif typically includes one negatively charged residue (D or E), one or two positively charged residues (K or R) and one or two hydrophobic residues (I, L, V, or F) in any

order along with a glutamine (Q) at either end of the motif. The “KFERQ-like” motif can be turned on or off by chemical changes to proteins i.e post-translational modifications (ubiquitination, phosphorylation or acetylation). KFERQ-like motifs without any modifications are called canonical motifs[10–13]. Currently, ~98 protein substrates are known for CMA, of which 76 substrates have been identified in humans[13–20]. The challenge lies in deciphering how Hsc70 achieves selectivity for substrate proteins in a cellular environment where multiple proteins bear similar recognition motifs. Particularly, what are the structural fingerprints that label a protein as a CMA substrate, how does Hsc70 recognize those fingerprints, what is the substrate binding site in Hsc70? This gap in understanding has significant implications, as substrate recognition is the cornerstone of CMA's selectivity, influencing various cellular processes and disease outcomes.

Hsc70 is a constitutive member of the Hsp70 family that shares a conserved structural architecture comprising of a nucleotide binding domain (NBD), and a substrate binding domain (SBD). The SBD is sub-divided into SBDβ, SBDα (lid) and a C-terminal extension[21,22]. Most of the existing literature exclusively utilized short peptides as model substrates to develop the current understanding of chaperone-substrate binding[23–26]. Crystal structure of a model peptide NRLLLTG (NR) in complex with DnaK suggested that the peptide binds in an extended conformation in a canonical hydrophobic cleft within SBDβ[26]. Peptides constituting CMA-targeting motifs of substrate protein ‘Tau’ were experimentally shown to bind independently and did not compete with model peptides for binding at SBDβ[27]. In the same study, NMR experiments have shown that these KFERQ-like peptide motifs were unable to bind Hsc70’s SBD that contained only SBDβ. This suggests that either the substrate protein containing KFERQ-like motifs bind at a site other than the canonical hydrophobic binding groove or the binding of KFERQ-like motifs is weaker than the model peptide at the same site with possibilities of existence

of secondary binding sites. It also points to the role of SBDα in binding. Additionally, SBDα is the sequentially divergent region amongst Hsp70 orthologs and residues being involved in binding at this site can account for the differential binding of tau with Hsc70 and Hsp72[28].

There are several bottlenecks towards gaining a comprehensive understanding of the interactions between Hsc70-substrate complex. Firstly, the atomistic structure of Hsc70 is unknown and the exact binding site is unidentified. Additionally, consistent with the principle applicable to chaperones in general, its binding with the client protein must be strong enough to prevent the spontaneous detachment of the bound protein yet not so strong to enable its detachment for subsequent downstream events. Thus, these chaperone-client complexes are considered as "fuzzy" complexes. The binding interfaces constitute formation of transient, dynamic, and nonspecific interactions, capable of modulating the conformational equilibrium, which are difficult to capture experimentally. The groundbreaking deep learning approach of AlphaFold 3 predicts structures of proteins and their complexes by leveraging evolutionary and structural data[29]. Combined with MD simulations, which model atomic interactions and dynamic motions, it provides a powerful framework to analyze these transient and dynamic interactions, overcoming experimental limitations. In the current study, we have utilized an integrated approach combining information from existing literature, molecular modelling, AlphaFold 3 and molecular dynamics simulations to predict Hsc70-substrate binding interactions. Our findings suggest that the residues adjacent to the CMA-targeting motifs in the same secondary structure element stabilize the binding interactions between SBDα of Hsc70 and the substrate. The analysis on these extended CMA-targeting motifs in fifteen human substrate proteins reveal common binding site on SBDα of Hsc70 encompassing residues Tyr545, Asn548, Thr552, Leu558, Lys561, Glu600, Ile607, Leu610, and Tyr611. The site consists of hydrophobic interior with hydrophilic residues flanking the central

113 pocket. The use of various force fields, different substrates, and supporting evidence from existing
114 literature-all converging on the same site, validates the robustness and success of our study. The
115 functional implications of the binding may not be to render selectivity to the CMA pathway since
116 the KFERQ pentapeptide is necessary and sufficient for binding[30], but they might contribute to
117 facilitating upstream or downstream recognition or activation events. The study opens the door to
118 further exploration in this direction. Additionally, we observed stable and selective binding of
119 CMA-targeting motifs to the canonical site in SBDβ, with major contributions from the
120 hydrophobic amino acids of the motif, along with significant interactions involving residues in
121 SBDα of Hsc70.

122 Results and Discussion

123

124 Adjacent Residues contribute to the Stability of 'KFERQ-like' Motif 125 Binding to Hsc70's lid domain

126

127 The binding of Hsc70 to CMA substrates has primarily been studied at the sequence level to date,
128 with the well-known recognition of 'KFERQ-like' motifs in substrates involved in binding. At the
129 structural level, evidence from multiple sources in literature suggests potential role of residues at
130 the lid domain/SBDα of Hsc70[27,28,31–33]. To understand the structural basis for substrate
131 recognition and the substrate-binding mechanism to SBDα of Hsc70 (if any), we compiled a list
132 of CMA substrates and analyzed their interaction with SBDα of Hsc70, focusing on the recognition
133 of 'KFERQ-like' motif.

134 A comprehensive list of experimentally validated CMA substrates in humans was collated. Using
135 the KFERQ Finder bioinformatics tool, both canonical KFERQ motifs and KFERQ-like motifs

136 generated by post-translational modifications (PTMs) were identified[13]. Fifteen substrates
137 containing a single canonical motif were selected for analyzing the recognition and binding site in
138 Hsc70 SBDα. The selected substrates are: RND3(Rho-related GTP-binding protein RhoE),
139 Abeta(Amyloid-beta precursor protein), Annexin A1, Annexin A2, Annexin A6, GBA1(Lysosomal
140 acid glucosylceramidase), Cyclin D1(G1/S-specific cyclin-D1), PEA15(Astrocytic
141 phosphoprotein PEA-15), HBB(Hemoglobin subunit beta), CHK1(Serine/threonine-protein
142 kinase Chk1), α-Synuclein, Annexin A4, Prion protein(Major prion protein), PKM(Pyruvate
143 kinase PKM), and PARK7(Parkinson disease protein 7). The list along with their canonical motifs
144 are enlisted in S1 Table. AlphaFold 3 was utilized to predict the structure(s) of the complex(es)
145 between substrate's 'KFERQ-like' motif and SBDα of human Hsc70. Fig 1A depicts the residue-
146 level interaction profiles of Hsc70 with different CMA substrate segments. Majorly, three distinct
147 sites were observed (site-1 is colored light steel blue, site-2 as pink and site-3 as peru), Fig 1B.
148 Few representative protein complexes from each site (RND3, Abeta and Annexin A1 from site-1,
149 CHK1, and PKM from site-2 and PARK7 from site-3) were subjected to all-atom MD simulations.
150 The stability of the binding site was checked by plotting the minimum distance between the
151 substrate and the receptor, Fig 1C and S1A Fig. In all cases, the predicted binding site demonstrated
152 structural instability during the simulations, failing to maintain its integrity throughout the
153 trajectory. Consequently, the substrate exhibited a loss of association with the putative binding
154 pocket and migrated away from the predicted site. This dislocation suggests a potential need to
155 refine the molecular environment to accurately replicate the native binding dynamics.

156 Since the 'KFERQ-like' motif alone does not maintain sufficient stability in complex with Hsc70,
157 we investigated the potential contribution of adjacent residues in stabilizing the motif's interaction
158 with the chaperone, as their presence may be essential for effective binding. Extensive stabilization

159 by additional residues outside the consensus sequence motifs have also been previously observed
160 for LIR motifs[34,35]. LIR motif is flanked by C-helix residues that cause additional stabilization.
161 To define 'adjacent' residues in our study, the subset of shortlisted substrate proteins is analyzed
162 for the existence of secondary structural elements in 'KFERQ-like' sequence motifs (Fig 1D). It
163 was observed that the motif forms a part of alpha-helical conformation in ten proteins. In only one
164 and four proteins, on the other hand, the motif exists in beta-sheet and loop respectively. Those
165 existing in the loop have flanking residues in alpha-helical conformation. Adjacent residues are
166 defined as variable-length stretches of residues flanking the 'KFERQ-like' motif on either side that
167 adopts same (in case of alpha-helix and beta-sheet) or adjacent (in case of loop) secondary structure
168 component as residues of the motif. These residues along with the motif, hereafter referred as
169 'extended motifs' were used for exploring the binding site between the substrate proteins and
170 Hsc70 SBDα.

171 AlphaFold 3 was utilized to predict the structure of the complex between the peptides with
172 extended motifs and the lid domain of Hsc70. To check the stability of these complexes, each of
173 them were subjected to 1 μs of all-atom MD simulations, generating a total of 15 μs of data.
174 Interestingly, compared to the previous case wherein only 'KFERQ-like' motifs were explored, the
175 extended motifs demonstrated structural stability of the complex. The minimum distance between
176 the substrate and the receptor protein remained below 2.5Å, indicating stable binding during the
177 timescale of MD simulations (Fig 1E top panel and S1B Fig.). The residues of Hsc70 involved in
178 the interaction with the residues of the substrate during the last 200 ns of the simulations are
179 illustrated in Fig 1E (bottom panel). It was observed that the simulations stabilize all the proteins
180 to site 1 of Hsc70. To investigate whether the observed binding stability was an artifact of a
181 particular force field or a general phenomenon, we performed simulations of the two of the

182 complexes generated by AlphaFold 3 (PARK7 and α-Synuclein) using multiple force fields
183 (CHARMM36, AMBER, and OPLS) with varying parameterization schemes. It was observed that
184 the substrates are bound throughout the trajectory (Fig 1F). Overall, across all tested force fields,
185 the binding remained intact, suggesting that the interaction is robust and not specific to any single
186 force field. This consistency enhances confidence in the accuracy of the binding site prediction
187 and its relevance under physiological conditions. Furthermore, the site is in agreement with the
188 previous proposed model based on cross-linking experiments between human Hsc70, yeast ssa1p
189 and α-Synuclein[33].

190 To evaluate the contribution of adjacent residues to the stability of the 'KFERQ-like' motifs in the
191 bound state, we removed the adjacent residues after they had participated in forming a complex
192 with the motif and the chaperone i.e. adjacent residues flanking the motif were removed from the
193 bound complexes in four of the previous runs (RND3, CHK1, PKM and PARK7) and the resulting
194 complexes were subjected to MD simulations. The time evolution of the distance between substrate
195 and receptor was monitored thereafter to check the stability of the complexes, Fig 1G. None of the
196 complexes remained stable at the same site. The substrate disengaged from the proposed binding
197 region and relocated to alternate positions. These findings indicate that the binding interactions
198 may not have been sufficiently robust. Taken together, our data suggests that additional residues
199 adjacent to 'KFERQ-like' motif appear to play a crucial role in stabilizing the complex with the
200 lid domain of Hsc70 under the conditions studied.

201 Since it is well-established that the pentapeptide motif in substrates is both necessary and sufficient
202 for performing its CMA activity[30], we believe that the binding of CMA substrates in this region
203 may not have direct implications in rendering selectivity to the process. Instead, the observed
204 interactions point toward the presence of an additional binding site, which may play a role in

upstream events such as substrate recruitment or in downstream processes like substrate transfer to LAMP2A. Whether and how the site could impact substrate triage, retention, or handoff to LAMP2A requires further experimental investigation. Nonetheless, the identified site remains robust, with supporting experimental evidence from the literature[32,33].

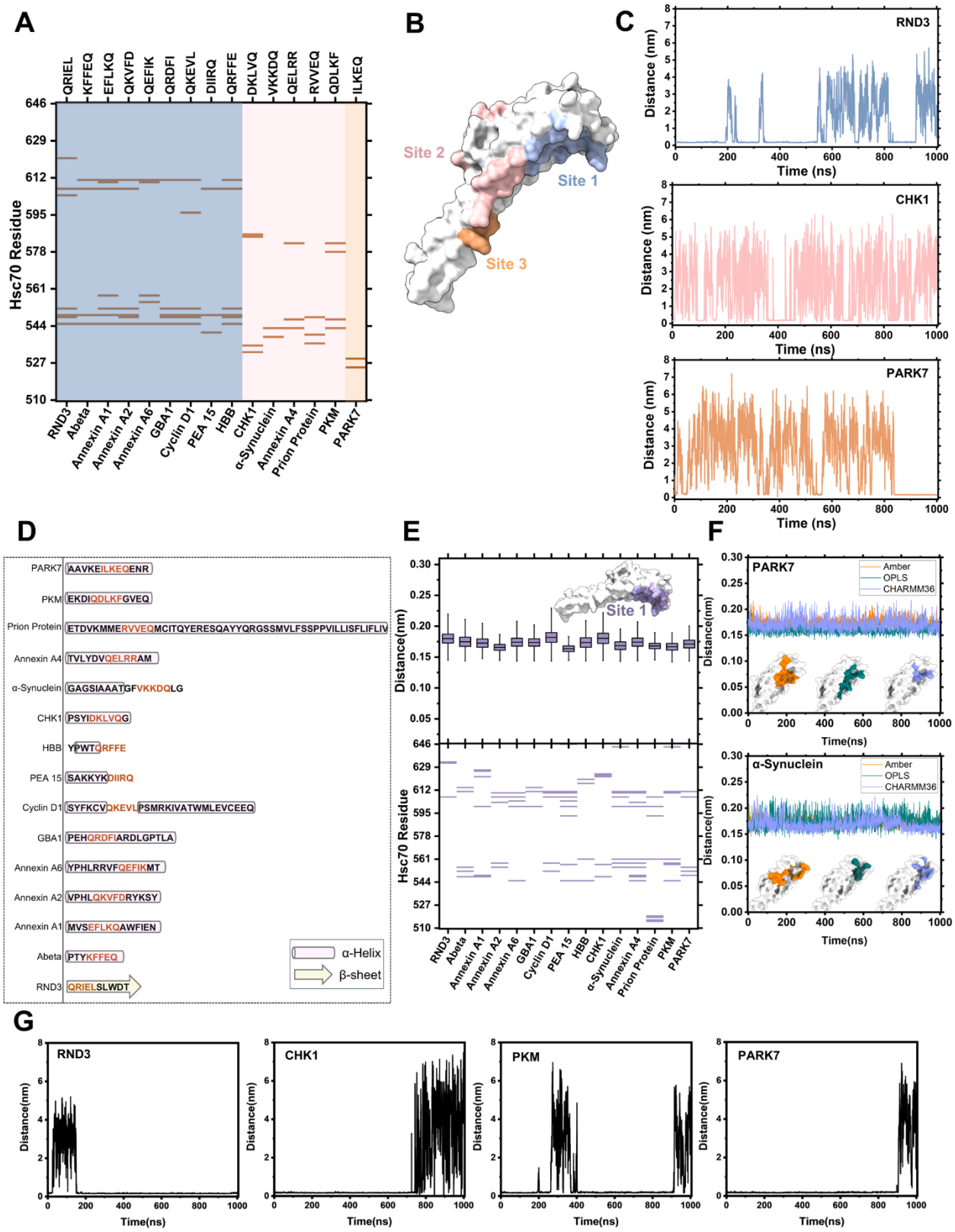

A
QRIEL
KFFEQ
EFLKQ
QKVFD
QEFIK
QRDFI
QKEVL
DIIRQ
QRFFE
DKLVQ
VKKDQ
QELRR
RVVEQ
QDLKF
ILKEQ
Hsc70 Residue
RND3
Abeta
Annexin A1
Annexin A2
Annexin A6
GBA1
Cyclin D1
PEA 15
HBB
CHK1
α-Synuclein
Annexin A4
Prion Protein
PKM
PARK7
B
Site 1
Site 2
Site 3
C
RND3
CHK1
PARK7
Distance (nm)
Time (ns)
D
PARK7 AAVKEILKEQENR
PKM EKDIQDLKFGVEQ
Prion Protein ETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSSPPVILLISFLIFLIV
Annexin A4 TVLYDVQELRRAM
α-Synuclein GAGSIAAATGFVKKDQLG
CHK1 PSYIDKLVQG
HBB YPWTQRFFE
PEA 15 SAKKYKDIIRQ
Cyclin D1 SYFKCVQKEVLPSMRKIVATWMLEVCEEQ
GBA1 PEHQRDFIARDLGPTLA
Annexin A6 YPHLRRVFQEFIKMT
Annexin A2 VPHLQKVFDRYKSY
Annexin A1 MVSEFLKQAWFIEN
Abeta PTYKFFEQ
RND3 QRIELSLWDT
α-Helix
β-sheet
E
Site 1
Distance(nm)
Hsc70 Residue
F
PARK7
α-Synuclein
Amber
OPLS
CHARMM36
Time(ns)
G
RND3
CHK1
PKM
PARK7
Distance(nm)
Time(ns)

Fig 1. Residues adjacent to the CMA-targeting motif stabilize its binding with the lid domain of truncated Hsc70.

(A**)** Heatmap representing the residues of Hsc70 lid domain interacting with 'KFERQ-like' motifs in AlphaFold 3 predicted complexes for 15 CMA substrates. Three different sites were observed (site-1 in light steel blue, site-2 in pink, and site-3 in peru). (B**)** 3D surface representation of the lid domain of Hsc70 depicting the three predicted binding sites. The loop region in the lid domain is not shown for clarity. (C) Minimum distance between Hsc70 lid domain and CMA substrate peptides with only 'KFERQ-like' motif. It represents structural instability of the complex during the course of MD simulations. (D) Schematic representation of secondary structure of CMA-targeting motifs along with adjacent amino acid residue of 15 CMA substrates that were taken for binding site predictions (termed as 'extended CMA-targeting motifs'). (E) The minimum distance plot between the Hsc70 lid domain and the extended CMA-targeting motifs (Top panel), along with heatmap representing the interacting residues of the two (Bottom panel), averaged over last 200 ns. Surface representation of Hsc70 lid domain with the interacting side highlighted in purple (loop region is not shown for clarity). The binding site remained stable throughout the simulations. (F) Minimum distance plots of Hsc70 lid domain and extended CMA-targeting motifs in PARK7 and α-Synuclein using AMBER, OPLS, and CHARMM 36 force fields. A value below 2.5 Å indicates stability in each force field. (G) Minimum distance between Hsc70 and RND3, CHK1, PKM and PARK7 peptide in the complexes. The complexes were generated by removing the adjacent residues in CMA substrates after they had participated in forming a complex with the motif and the chaperone. None of the complexes remained stable.

209

Extended CMA-targeting motifs in substrate proteins have selective affinity to hydrophobic cavity formed by helix α1 and helix α3 in SBDα of Hsc70

To delve deeper into the binding site and interaction pattern, the binding pocket in Hsc70 and the nature of interacting residues were identified. Fig 2A depicts a chord plot representing the residues involved in interaction. Following an extensive analysis of the trajectories, we identified a set of receptor residues that consistently participated in interactions across majority of the simulated complexes. These residues include Tyr545, Asn548, Met549, Asp555, Lys557, Leu558, Lys561, Glu600, Asn604, Ile607, Leu610, Tyr611, which were involved in stabilizing the binding through van der Waals interactions. Notably, Ile607 and Tyr611 emerged as critical hotspots, forming persistent interactions with the peptide in over 8 of the complexes. The conservation of these interaction patterns across the majority of systems underscores their potential significance in maintaining the binding and highlights their role in the structural integrity of the receptor-substrate complex. Overall, it was observed that interacting residues belong to helix *α1* and helix *α3* of SBD*α* in Hsc70 (Fig 2B).

The contribution of each Hsc70 residue towards stabilizing the binding interactions were analyzed by performing free energy calculations using MM-GBSA (Molecular Mechanics-General Born Surface Area) approach. Fig 2C showcases the binding energies of these residues with values more negative than -2.5 kcal/mol. Most of the residues are consistent with our occupancy analysis above. The common residues of occupancy and binding energy analysis include Tyr545, Asn548, Thr552, Leu558, Lys561, Glu600, Ile607, Leu610, and Tyr611. We propose the significance of these residues in rendering stabilization to the complex. Further, the identified residues in Hsc70 encompassing the site contribute to the formation of a cavity with hydrophobic interior and

hydrophilic exterior. The site can accommodate the side chains of hydrophobic amino acids of the peptide segment.

Extended CMA-targeting motifs in substrate proteins display promiscuous orientation in the binding hydrophobic cavity of SBDα

The stabilizing interactions between the substrate and chaperone within the selective binding site were examined in greater detail. Interestingly, despite the substrate consistently associated with the designated binding pocket, no common binding orientation of the substrate was identified. Our simulations suggested that the substrate(s) adopt different binding modes in each simulation, leading to variable contact patterns with the binding site, S2A Fig. A comprehensive analysis of interaction patterns revealed a lack of reproducible stabilizing contacts across simulations of different substrates. Quantitative analysis, including calculating the contact frequencies and interaction energy underscores this observation (Fig 2A and S2B Fig. respectively). Nevertheless, in all the simulations, hydrophobic amino acids were observed to have more negative interaction energy. The minimum distance between key interacting pairs of Hsc70 and the substrate(s) is depicted in S3 Fig. These interactions were stable with the value ranging within 4Å. However, no consistent pair was identified in all the simulations. This variability indicates that, while the substrate binds selectively to the hydrophobic pocket, the mode of binding demonstrates considerable flexibility. The absence of a consistent interaction pattern suggests a promiscuous nature of substrate-chaperone binding to this site.

When simulations were performed using different force fields, this dual characteristic of binding was consistently observed: while the substrate exhibited high specificity for the binding site across

all force fields, the nature and dynamics of individual interactions varied significantly (S4 Fig.). Certain force fields emphasized stronger hydrophobic contacts, while others resulted in a more diverse range of transient interactions. These differences highlight the role of force field parameters in capturing the flexibility of substrate binding modes, suggesting that the observed promiscuity may be an inherent property of the system rather than an artifact of a specific simulation setup. Overall, the data suggest that substrate binding to the observed site in SBDα is promiscuous and does not by itself account for substrate selectivity. Instead, our data reveals a potential secondary site in the lid domain. The functional implication of this site requires further experimental investigation. The observed site is however consistent with previous cross-linking experiments[32,33].

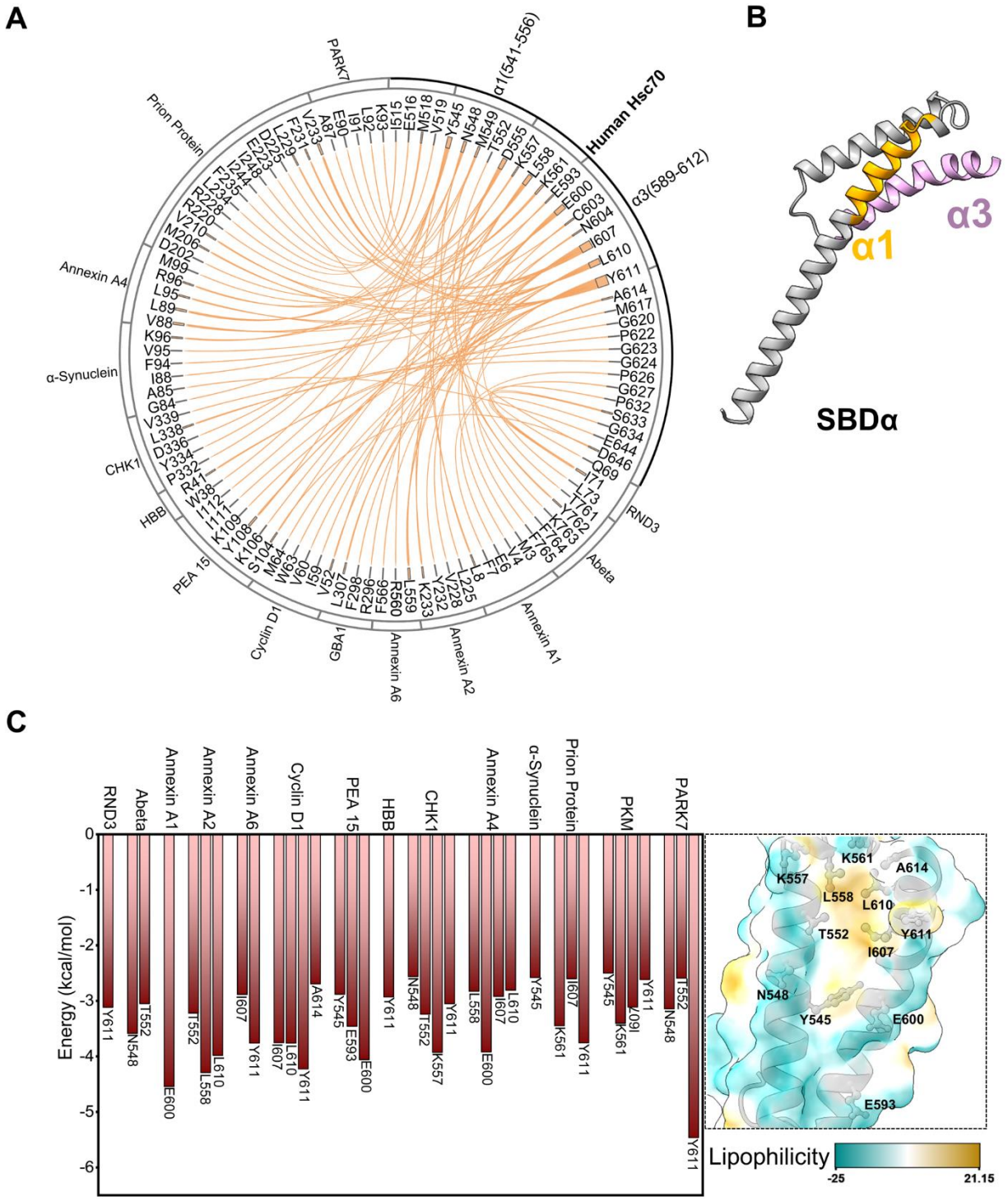


Fig 2. Extended CMA-targeting motifs selectively bind to the hydrophobic cavity formed by helix α1 and helix α3 in the lid domain of truncated Hsc70.

(A) A chord diagram representing the residues of Hsc70 lid domain interacting with extended CMA-targeting motifs from 15 CMA substrates, where the thickness of the node for each respective residue represents the frequency of interactions. Residues of helices α1 and α3 are extensively participating in interactions with the substrate segment. (B) 3D cartoon representation of the lid domain of Hsc70 depicting the helices α1 and α3 which are involved interactions with the substrate segment. The loop region in the lid domain is not shown for clarity. (C) Residue-wise decomposition of binding energies of Hsc70 residues with the CMA substrate peptides is shown in the left panel. The bar plot of binding energies of residues with values less than -2.5 kcal/mol (calculated using MMGBSA) are shown for clarity. Right panel showcases lipophilic representation of the binding pocket in Hsc70. Key interacting residues are depicted in ball-and-stick representation.

Stable recognition of CMA-targeting motifs requires the closed conformation of full-length Hsc70

Owing to the obtained promiscuous nature of binding of CMA-targeting motifs with SBDα of Hsc70, we next investigated the binding of the motif with full-length Hsc70 (along with ADP) to reason its selective behavior in CMA. AlphaFold 3 predicted the binding site to be the canonical hydrophobic cleft in SBDβ for all the 15 tested substrates. Interestingly, 4 of them were predicted to be bound to the ADP-bound closed form of Hsc70 while remaining 11 were bound to its open form (Fig 3A-B and S5 Fig.). All the closed-form complexes (Abeta, Annexin A1, PEA 15 and HBB), along with six representative open-form (RND3, α-Synuclein, CHK1, PKM, Annexin A6 and PARK7) complexes were subjected to MD simulations of 1 μs each to check their stability. The analysis of the time evolution of the minimum distance between the substrate and Hsc70

suggested that complexes bound to the closed form were stable throughout the time scale of our simulations, while the open forms exhibited pronounced fluctuations (Fig 3C and D). This suggests that efficient recognition of KFERQ-like motifs requires conformational coupling between the nucleotide-binding domain (NBD) and SBDβ, and/or the presence of additional structural elements such as the interdomain linker or SBDα/lid. Consistent with our observations, the importance of ATP allostery of Hsc70 in regulating the interaction has also been shown experimentally[36].

The unstable open form of Hsc70-substrate complexes in our MD simulations raises the question of whether the instability arose from weak substrate-chaperone interactions themselves or from the absence of the lid domain. To disentangle these possibilities, we preserved the substrate-chaperone contacts from the three of the representative AF3-predicted open forms (RND3, α-synuclein, PARK7) and converted the complexes into closed conformations. Each system was then subjected to 1 μs of MD simulations. In all three cases, the substrate remained stably bound throughout the trajectories (Fig 3E), supporting the view that lid closure is critical for stable substrate engagement.

Having established that lid closure promotes stable substrate engagement, we next examined the effect of lid/SBDα removal. To this end, we truncated the SBDα from two representative closed-form complexes (Abeta and Annexin A1), thereby mimicking isolated SBDβ bound to the KFERQ-like motif while retaining the previously established interactions. Both systems were subjected to 1 μs of all-atom MD simulations. In both cases, the CMA-targeting peptide failed to remain in a stable pose over the course of the simulation (Fig 3F). These results, together with our earlier observation that full-length Hsc70 stably accommodates the motif, are consistent with experimental reports that KFERQ motifs fail to bind isolated SBDβ constructs[27]. Collectively,

302 these results highlight the importance of investigating substrate recognition in the context of the

303 full-length chaperone.

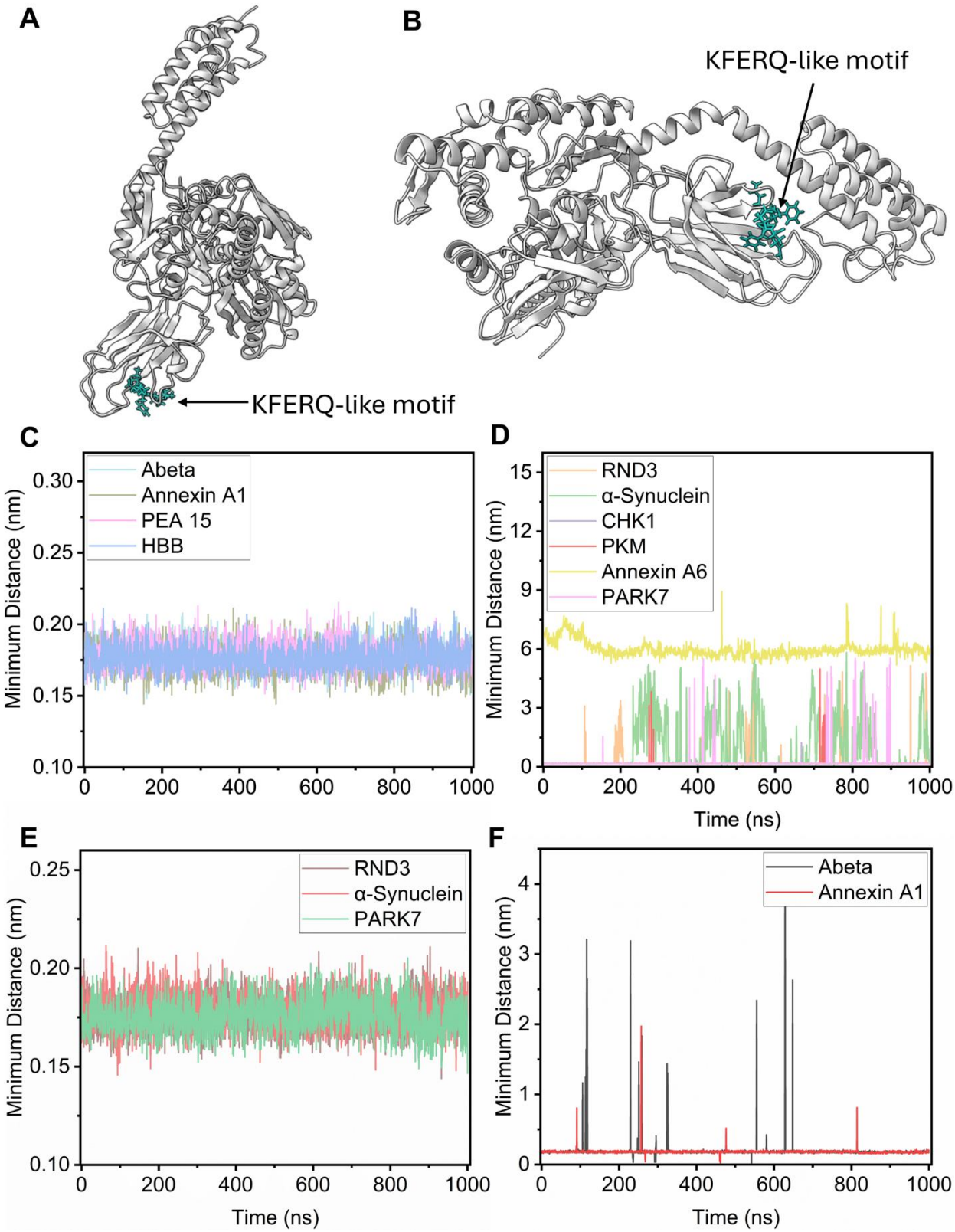

A
KFERQ-like motif
B
KFERQ-like motif
C
Abeta
Annexin A1
PEA 15
HBB
Minimum Distance (nm)
D
RND3
α-Synuclein
CHK1
PKM
Annexin A6
PARK7
Minimum Distance (nm)
Time (ns)
E
RND3
α-Synuclein
PARK7
Minimum Distance (nm)
Time (ns)
F
Abeta
Annexin A1
Minimum Distance (nm)
Time (ns)

Fig 3. CMA-targeting motifs bind to full-length Hsc70 in its closed conformation, where the lid domain plays an essential role in stabilizing the interaction.

Representative snapshot of AlphaFold 3 (AF3) predicted conformation of Hsc70 (dark gray)-substrate (light sea green) complex with Hsc70 in (A**)** open conformation and (B**)** closed conformation. Time evolution of the minimum distance between CMA-targeting motif and Hsc70 in AF3 predicted substrate-Hsc70 complexes wherein Hsc70 is in (C**)** closed formation, (D) open conformation, (E) open conformation initially but were converted to the closed conformation keeping initial contacts intact, F) closed conformation initially but the lid domain (residues 510–646) was removed before subjecting it to MD simulations.

304

305 **CMA-targeting motifs selectively bind the canonical hydrophobic**
306 **cavity of SBDβ, preserving backbone alignment despite side-chain**
307 **registry shifts**

308

309 To gain mechanistic insights into how Hsc70 closed conformation stabilizes substrate binding, we
310 analyzed the residues and interaction sites in full-length Hsc70 involved in recognizing CMA-
311 targeting motifs (Fig 4A). Ala 406, and Thr 429 in Hsc70 were observed to have occupancies
312 exceeding 90% of the simulation time, across all the tested substrates. All of the stable CMA
313 substrate complexes in our study were bound to the canonical hydrophobic cavity within SBDβ
314 (residues 394-509 of Hsc70), Fig 4B. NRLLLTG peptide has also been previously reported to be
315 nestled in the same groove of DnaK[37] and Hsp70[38]. The binding site in these proteins is
316 organized into five distinct pockets (-2,-1,0,1,2) that accommodate the extended peptide backbone,
317 while the helical lid of the SBD folds over to enclose the substrate. The central pocket designated
318 as position "0", is highly specific and primarily accommodates hydrophobic residues. The

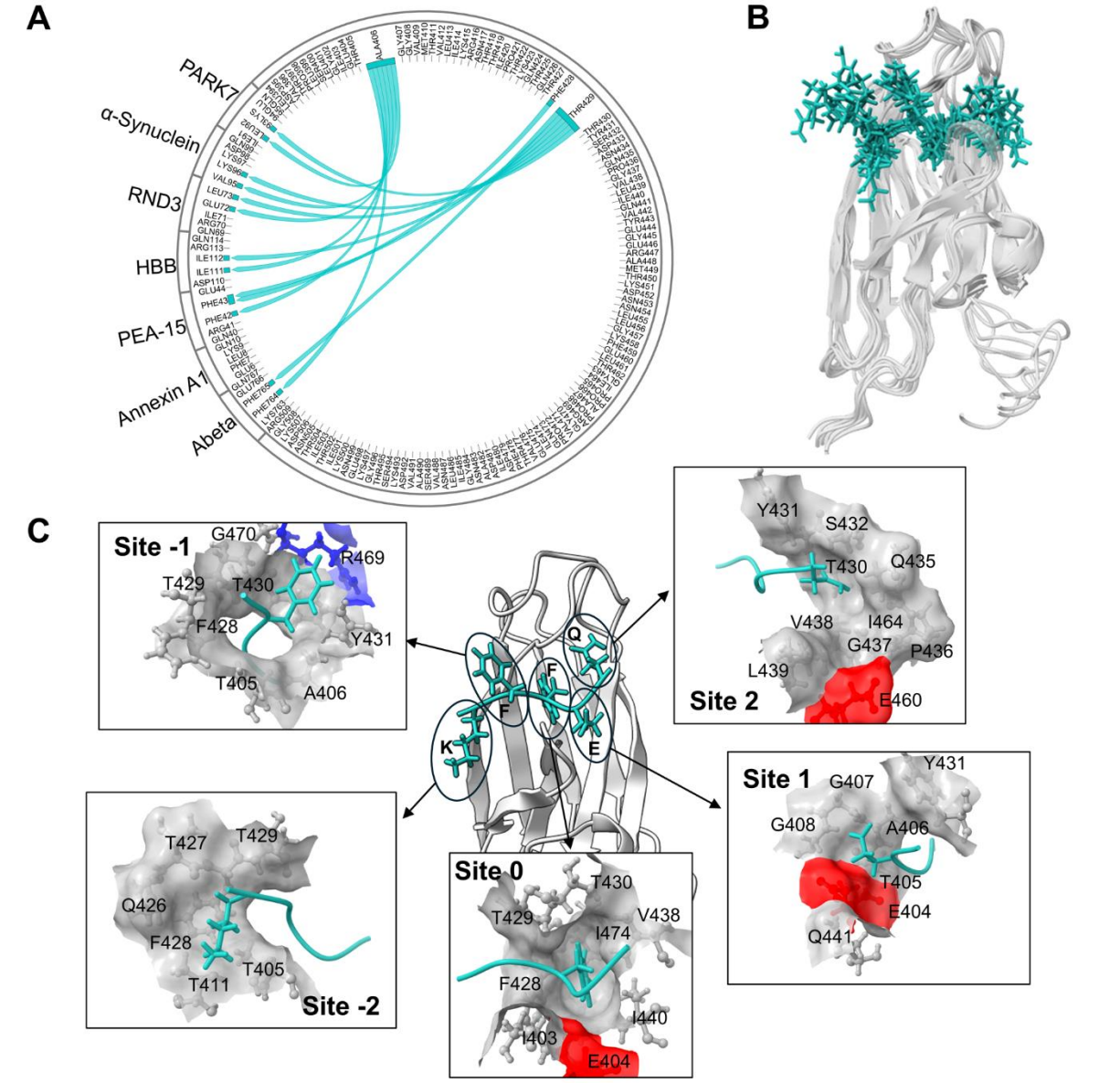
319 remaining pockets are numbered relative to the central pocket. For the NRLLLTG peptide, it is
320 reported that the central leucine residue fits optimally into the pocket 0. The flanking pockets are
321 less specific and can accommodate diverse substrate side chains, thereby conferring binding
322 promiscuity.

323 In line with these observations, we analyzed the placement of the pentapeptide residues within the
324 five pockets. A representative example of a KFERQ-like motif from Abeta interacting with pocket
325 residues is shown in Fig 4C. Our results suggest that the central binding site (site 0) in Hsc70
326 comprising residues Ile 403, Glu 404, Phe 428, Thr 429, Thr 430, Val 438, Ile 440, and Ile 474
327 serves as the most selective site stabilizing hydrophobic residue(s) of CMA-targeting motifs in all
328 seven stable complexes in MD simulations (Closed form of Hsc70 with CMA-targeting motifs of
329 (i) Abeta, (ii) Annexin A1, (iii) PEA-15, (iv) HBB, along with AF3-predicted open form Hsc70
330 complexes that were converted into complex with Hsc70 in closed conformations: (v) RND3, (vi)
331 α-Synuclein, and (vii) PARK7). The remaining pockets, by contrast, are less specific,
332 accommodating both hydrophobic and polar residues of CMA-targeting motifs and contributing
333 only minor interactions. Binding free energy analysis using MM-GBSA further supports
334 significant contribution of residues involved in site 0, as depicted in Fig 4D. Collectively, our
335 findings that nonpolar interactions provide the major energetic contribution to binding are
336 consistent with prior reports for DnaK[39].

337 The canonical binding groove of Hsp70 is well known to exhibit registry plasticity and a lack of
338 strict directional bias, as both forward and reverse orientations of peptide binding have been
339 reported previously[40]. Additionally, comparative structural studies have revealed a one-residue
340 registry shift between bacterial DnaK and human Hsp70/BiP bound to similar NR peptides,
341 indicating that even canonical motifs can adopt alternate alignments. In this context, our structural

342 modeling of the KFERQ-like motifs revealed a registry nearly identical to that observed for Hsp70-
343 bound NRLLLTG, with the hydrophobic residue fitting into the 0 subsite, with a specific
344 preference to leucine residue. Other hydrophobic residues like isoleucine, and phenylalanine were
345 also observed to bind at site 0 (S2 Table). Energy analysis suggests that these hydrophobic residues
346 are the major contributors of the interaction energy (Fig 4E).

347 Interestingly, even though side-chain positioning within the substrate-binding groove may vary
348 i.e., different residues occupy the hydrophobic pockets at subsites 0, ± 1, and ± 2, but the backbone
349 registry remains conserved, (Fig 4B, S6 Fig.). Irrespective of the substrate, backbone residues
350 Ala406 and Thr429 were observed to form hydrogen bonds with the backbone of the CMA-
351 targeting motif. This structural feature was robust across independent simulations and aligns with
352 prior structural observations in both bacterial DnaK and mammalian Hsp70, where registry shifts
353 at the side-chain level did not disrupt the backbone hydrogen bonding pattern[26,40]. These
354 findings suggest that the Hsp70 groove accommodates a degree of registry plasticity at the side-
355 chain level while enforcing a conserved backbone geometry, providing a mechanistic basis for
356 how diverse CMA-targeting motifs can be stably recognized without requiring identical residue
357 alignments.


A
PARK7
α-Synuclein
RND3
HBB
PEA-15
Annexin A1
Abeta
B
C
Site -1
G470
R469
T429
T430
F428
Y431
T405
A406
Site 2
Y431
S432
T430
Q435
V438
I464
G437
P436
L439
E460
Site 1
Y431
G407
G408
A406
T405
E404
Q441
Site -2
T427
T429
Q426
F428
T405
T411
Site 0
T430
T429
V438
I474
F428
I440
I403
E404
Q
F
F
K
E

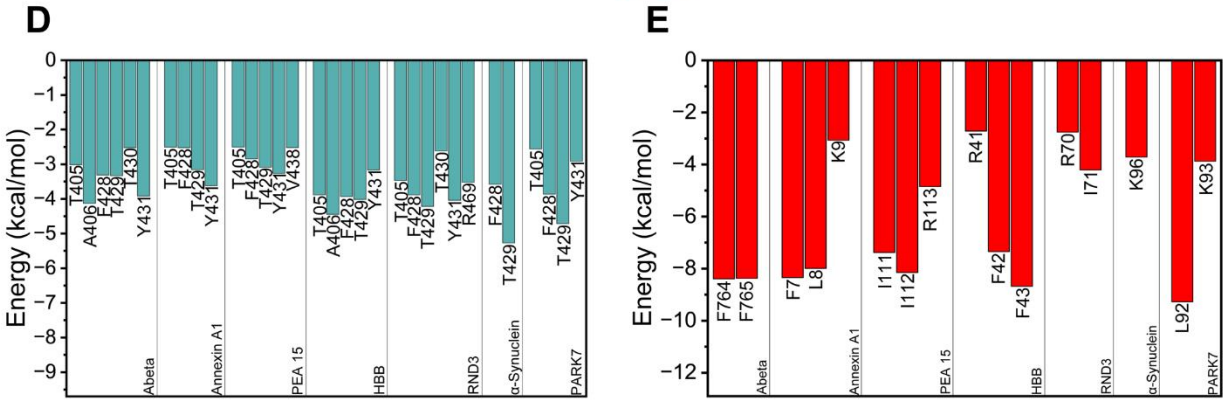

D
Energy (kcal/mol)
T405
A406
F428
T429
T430
Y431
Abeta
T405
F428
T429
Y431
T429
Y431
Annexin A1
T405
F428
T429
Y431
T429
Y431
T429
Y431
PEA 15
T405
A406
F428
T429
HBB
T405
F428
T429
T430
Y431
R469
RND3
F428
T429
α-Synuclein
T405
F428
T429
Y431
PARK7
E
Energy (kcal/mol)
F764
F765
Abeta
F7
L8
K9
Annexin A1
I111
I112
R113
PEA 15
R41
F42
F43
HBB
R70
I71
RND3
K96
α-Synuclein
L92
K93
PARK7

Fig 4. CMA-targeting motifs bind to the hydrophobic cleft of the SBDβ domain in the closed conformation of Hsc70, maintaining the backbone registry and being stabilized primarily by hydrophobic residues.

(A) Chord plot representing the interacting residues of Hsc70 SBDβ domain and the CMA-targeting motif of seven CMA substrates in which Hsc70-substrate complex is stable during 1μs of MD simulations. The thickness of the node for each respective residue represents the frequency of interactions For Clarity, residues with interaction occupancy more than 90% are shown. (B) Overlay of the CMA-targeting motifs (light sea green) in these seven CMA substrates, when bound to closed conformation of full-length Hsc70 (dark gray). For clarity, only SBDβ of Hsc70 is shown. (C) Snapshot of CMA-targeting motif of Abeta (light sea green in stick representation) bound to Hsc70 (dark grey in cartoon representation) complex at 1μs of MD simulation data. Again, for clarity, only SBDβ of Hsc70 is shown. Surface representation of five distinct pockets within hydrophobic cleft of SBDβ in Hsc70 accommodating five residues of the motif are shown alongside. Pocket residues are explicitly labeled and colored gray, blue, or red to represent hydrophobic, basic, and acidic amino acids, respectively. (D) Residue-wise decomposition of binding energies of Hsc70 residues with the CMA substrate peptides. (E) Residue-wise decomposition of binding energies CMA substrate peptides with Hsc70. For clarity, only the bar plots of binding energies of residues with values less than -2.5 kcal/mol (calculated using MMGBSA) are shown in panels (D) and (E).

358

359 **Transient water molecules mediate interactions between CMA-**
360 **targeting motifs and the SBDβ-lid interface of Hsc70.**
361

362 To delve deeper into the reason behind the stable binding of KFERQ-like peptides only in the
363 presence of SBDα/lid, we evaluated the contribution of the surrounding water molecules to the
364 binding stability. Our simulations suggested that water molecules mediate interactions between
365 CMA-targeting motifs and the lid interface of Hsc70 via the formation of hydrogen bonds (Fig 5A,
366 S7 Fig.). The time evolution of the number of water molecules bridging the substrate peptide and
367 the lid domain indicates the involvement of ~1-7 water molecules depending on the substrate.
368 While the residues of substrate peptide and/or lid domain of Hsc70 involved in hydrogen bond
369 formation is not consistent among different substrates and/or different time frames, interestingly,
370 the water bridge was observed to be formed between the hydrophilic residues of lid domain with
371 hydrophilic amino acid of KFERQ-like motifs (S8 Fig.).

372 In addition to the formation of a direct water bridge, extended water chains (two concurrent water
373 bridges, or in some cases, three or more water molecules) where a substrate-water hydrogen bond
374 is relayed through additional water molecules before reaching the lid domain, were also observed
375 between the substrate peptide and the lid domain of Hsc70 (Fig 5). For instance, residue E766 of
376 Abeta and residue K539 of SBDα were connected by two water molecules via hydrogen bonds
377 (Fig 5B), whereas three water molecules mediate interactions between residue Q767 of Abeta and
378 residue E543 of SBDα (Fig 5C). On similar lines, four water molecules bridges residue K763 of
379 Abeta and residue G635 of SBDα (Fig 5D), further reinforcing the interaction network. Similar to
380 the above observations, these bonds are not consistent in different substrates and/or in different
381 time frames. We speculate that these transient hydrogen bonds dynamically stabilize the closed
382 conformation of Hsc70.

383 Furthermore, we explored the contribution of water in stabilizing the interactions between the
384 substrate peptide and SBDβ of Hsc70. The time evolution of the number of direct water bridges

between the substrate peptide and the SBDβ indicates the involvement of ~1-13 water molecules, (S9 Fig.). Overall, these results suggest that the role of water in CMA substrate recognition may be indirect but functionally significant, particularly in shaping the conformational landscape of the binding groove.

To summarize, the current work proposed two potential binding sites in Hsc70 for CMA substrates. The first site resides in the hydrophobic cavity inside the lid domain of Hsc70 encompassing residues in α1 and α3 helix. Notably, CMA-targeting motifs alone are insufficient to stabilize the interactions and that adjacent hydrophobic residues play a crucial role in reinforcing it. The second site was observed to be the canonical hydrophobic cavity inside the SBDβ domain of Hsc70. The binding in the site remained stable exclusively in the closed conformation of Hsc70, where transient water molecules mediated interactions between hydrophilic residues of the CMA-targeting motifs and the lid domain. Both identified sites have supporting interaction evidence in the literature[32,36]. However, whether our simulation results reflect a metastable intermediate or an alternative, physiologically relevant interaction pose remains to be validated. Further structural and biochemical analyses will be required to determine whether this modeled conformation corresponds to a functional state of CMA substrate recognition.

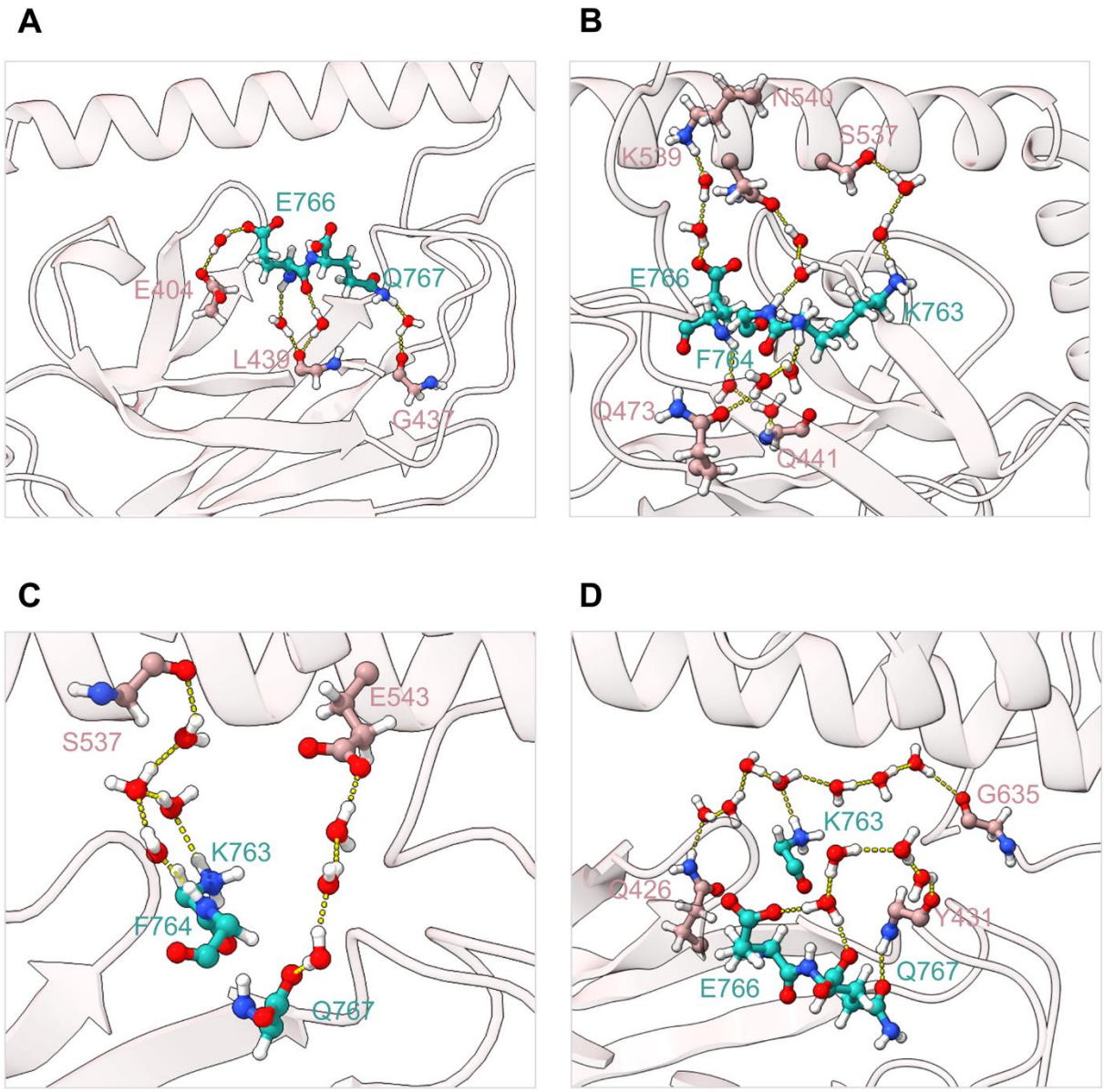


Fig 5. Water molecules mediate interactions between CMA-targeting motifs and the SBDβ-lid interface of Hsc70.

Representative examples showcasing CMA-targeting motifs in Abeta interacting with the lid domain of Hsc70 through following number of water molecules: (A) one (B) two (C) three and (D) four. Hsc70 is represented in rosy brown and the interacting residues of CMA-targeting motifs in light sea green color.

Methods

Preparation of Starting Inputs for Hsc70-Substrate Complex Generation

Fifteen experimentally verified CMA substrates were selected based on the following criteria: i) they are of human origin and ii) each contains a single canonical KFERQ-like motif. The presence and sequence of KFERQ-like motifs in these substrates were identified using the KFERQ finder V0.8 tool[13]. The sequences of these canonical motifs along with the full-length Hsc70 sequence and ADP, were used as an input to AlphaFold 3 (AF3)[29] to generate models of the binding complexes between Hsc70 and their respective CMA substrates. For truncated Hsc70 models comprising only the substrate-binding domain α (SBDα), residues 510-646 were used as input. Among the five models generated for each complex, the one with the highest interface predicted TM-score (iPTM) was selected for further analysis (S3 Table). The secondary structures of the substrate proteins were determined using the STRIDE[41] program, applied to structures obtained from the Protein Data Bank (PDB) and AlphaFold database[42–44]. To construct complexes of extended 'KFERQ-like' motifs with Hsc70, the following steps were performed. For each substrate protein, residues constituting the 'KFERQ-type' motif were first located in the sequence and mapped onto their corresponding positions in the 3D structure. The residues immediately flanking the motif on both N- and C-terminal sides were then analyzed for their secondary structure annotation. Adjacent residues were defined as variable-length stretches of amino acids contiguous to the KFERQ-type motif that (i) adopt the same secondary structure as the motif (in cases where the motif resides within an α-helix or β-sheet) or (ii) belong to an adjacent secondary structure element (in cases where the motif is located within a loop region). These residues, together with the KFERQ-type motif, were grouped and termed as 'extended motifs'. The sequence of these

425 extended motifs was fed as an input to AF3, along with truncated Hsc70 sequence (only SBDα
426 domain, residues 510 to 646) to generate complex models. As before, the models with the highest
427 iPTM scores were selected for further analysis.

428 Complexes of the KFERQ-like motifs from RND3, α-synuclein, and PARK7 with the closed
429 conformation of full-length Hsc70 were constructed using a hybrid modeling approach. Initially,
430 the interaction interfaces between each KFERQ-type motif and the AlphaFold 3 (AF3)-predicted
431 open conformation of Hsc70 (which includes ADP though) were identified and used as reference
432 contact maps. These contact interfaces were then preserved while fitting the respective motifs into
433 the closed conformation of Hsc70, obtained from the AF3-predicted complex of Hsc70 with the
434 KFERQ-like motif of Abeta. This procedure ensured that the substrate-Hsc70 interactions
435 observed in the open state were retained within the closed-state framework of Hsc70 for
436 subsequent analyses.

437 Molecular Dynamics Simulations

438

439 To explore the stability of protein-protein interactions in the Hsc70-substrate complexes predicted
440 using AlphaFold 3, all-atom MD simulations were performed using the GROMACS 2020.4
441 package for each model of Hsc70-CMA substrate complexes[45]. CHARMM36 force field was
442 used to simulate protein complexes[46,47]. Preparation of initial systems was done by assigning
443 standard protonation states to the ionizable residues corresponding to pH 7 using PropKa[48].
444 Cubic box with a gap of 10 Å between the surface of the protein complex and the edge was defined.
445 Systems were solvated using TIP3P water model and neutralized by NaCl[49]. Energy
446 minimization was done using Steepest Descent algorithm until the energy difference between steps
447 was less than 1000 kJ mol^{-1} nm^{-1}[50]. Two-step equilibration was performed, first using the NVT

ensemble, followed by the NPT ensemble. The temperature was controlled to 300 K by a V-rescale thermostat[51] and pressure to 1 bar using the Parrinello–Rahman barostat[52]. A time step of 2 fs was used and at every 100 steps coordinates were written. Amber as well as OPLS force fields were used for two complexes (α-Synuclein and PARK7) to analyze the effect of force fields[53,54]. The analysis was carried out either using in-build gromacs utilities or in-house scripts.

Acknowledgements

N.M. acknowledges the financial support from the Shiv Nadar Foundation and the DST Inspire Faculty Grant [DST/INSPIRE/04/2022/001599]. D.S. acknowledges the junior research fellowship from the Shiv Nadar Foundation. We are grateful to the supercomputing facility, Magus02, at SNIoE for the compute support.

References

1. Massey A, Kiffin R, Cuervo AM. Pathophysiology of chaperone-mediated autophagy. Int J Biochem Cell Biol. 2004;36: 2420–2434. doi:10.1016/j.biocel.2004.04.010

2. Vélez EJ, Schnebert S, Goguet M, Balbuena-Pecino S, Dias K, Beauclair L, et al. Chaperone-mediated autophagy protects against hyperglycemic stress. Autophagy. 2024;20: 752–768. doi:10.1080/15548627.2023.2267415

3. Massey AC, Zhang C, Cuervo AM. Chaperone-Mediated Autophagy in Aging and Disease. Curr. Top. Dev. Biol. 2006;73: 205–235. doi:10.1016/S0070-2153(05)73007-6

4. Koga H, Cuervo AM. Chaperone-mediated autophagy dysfunction in the pathogenesis of neurodegeneration. Neurobiol Dis. 2011;43: 29–37. doi:10.1016/j.nbd.2010.07.006

5. Wu J, Xu W, Su Y, Wang G, Ma J. Targeting chaperone-mediated autophagy in neurodegenerative diseases: mechanisms and therapeutic potential. Acta Pharmacol Sin. 2024;46: 816-828. doi:10.1038/s41401-024-01416-3

475 6. Dice JF. Peptide sequences that target cytosolic proteins for lysosomal proteolysis.
476 Trends Biochem Sci. 1990;15: 305–309. doi:10.1016/0968-0004(90)90019-8

477 7. Chiang H-L, Terlecky SR, Plant CP, Dice JF. A Role for a 70-Kilodalton Heat Shock
478 Protein in Lysosomal Degradation of Intracellular Proteins. Science 1989;246: 382–
479 385. doi:10.1126/science.2799391

480 8. Cuervo AM, Terlecky SR, Dice JF, Knecht E. Selective binding and uptake of
481 ribonuclease A and glyceraldehyde-3-phosphate dehydrogenase by isolated rat liver
482 lysosomes. J Biol Chem. 1994;269: 26374–80.

483 9. Koga H, Martinez-Vicente M, Macian F, Verkhusha V V., Cuervo AM. A
484 photoconvertible fluorescent reporter to track chaperone-mediated autophagy. Nat
485 Commun. 2011;2: 386. doi:10.1038/ncomms1393

486 10. Lv L, Li D, Zhao D, Lin R, Chu Y, Zhang H, et al. Acetylation targets the M2 isoform
487 of pyruvate kinase for degradation through chaperone-mediated autophagy and
488 promotes tumor growth. Mol Cell. 2011;42: 719–730.
489 doi:10.1016/j.molcel.2011.04.025

490 11. Bonhoure A, Vallentin A, Martin M, Senff-Ribeiro A, Amson R, Telerman A, et al.
491 Acetylation of translationally controlled tumor protein promotes its degradation
492 through chaperone-mediated autophagy. Eur J Cell Biol. 2017;96: 83–98.
493 doi:10.1016/j.ejcb.2016.12.002

494 12. Kaushik S, Cuervo AM. Degradation of lipid droplet-associated proteins by
495 chaperone-mediated autophagy facilitates lipolysis. Nat Cell Biol. 2015;17: 759–
496 770. doi:10.1038/ncb3166

497 13. Kirchner P, Bourdenx M, Madrigal-Matute J, Tiano S, Diaz A, Bartholdy BA, et al.
498 Correction: Proteome-wide analysis of chaperone-mediated autophagy targeting
499 motifs. PLoS Biol. 2022;20: e3001550. doi:10.1371/journal.pbio.3001550

500 14. Kaushik S, Cuervo AM. The coming of age of chaperone-mediated autophagy. Nat
501 Rev Mol Cell Biol. 2018;19: 365–381. doi:10.1038/s41580-018-0001-6

502 15. Majeski AE, Fred Dice J. Mechanisms of chaperone-mediated autophagy. Int J
503 Biochem Cell Biol. 2004;36: 2435–2444. doi:10.1016/j.biocel.2004.02.013

504 16. Wang D, Peng Z, Ren G, Wang G. The different roles of selective autophagic protein
505 degradation in mammalian cells. Oncotarget. 2015;6: 37098–37116.
506 doi:10.18632/oncotarget.5776

507 17. Robert G, Jacquel A, Auberger P. Chaperone-Mediated Autophagy and Its Emerging
508 Role in Hematological Malignancies. Cells. 2019;8: 1260. doi:10.3390/cells8101260

509 18. Cuervo AM, Gomes A V., Barnes JA, Dice JF. Selective Degradation of Annexins by
510 Chaperone-mediated Autophagy. J. Biol. Chem. 2000;275: 33329–33335.
511 doi:10.1074/jbc.M005655200

19. Gomes LR, Menck CFM, Cuervo AM. Chaperone-mediated autophagy prevents cellular transformation by regulating MYC proteasomal degradation. Autophagy. 2017;13: 928–940. doi:10.1080/15548627.2017.1293767

20. Tang Y, Wang X-W, Liu Z-H, Sun Y-M, Tang Y-X, Zhou D-H. Chaperone-mediated autophagy substrate proteins in cancer. Oncotarget. 2017;8: 51970–51985. doi:10.18632/oncotarget.17583

21. Qi R, Sarbeng EB, Liu Q, Le KQ, Xu X, Xu H, et al. Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. Nat Struct Mol Biol. 2013;20: 900–907. doi:10.1038/nsmb.2583

22. Bertelsen EB, Chang L, Gestwicki JE, Zuiderweg ERP. Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proc Natl Acad Sci U S A. 2009;106: 8471–8476. doi:10.1073/pnas.0903503106

23. Cupp-Vickery JR, Peterson JC, Ta DT, Vickery LE. Crystal Structure of the Molecular Chaperone HscA Substrate Binding Domain Complexed with the IscU Recognition Peptide ELPPVKIHC. J Mol Biol. 2004;342: 1265–1278. doi:10.1016/j.jmb.2004.07.025

24. Zahn M, Berthold N, Kieslich B, Knappe D, Hoffmann R, Sträter N. Structural Studies on the Forward and Reverse Binding Modes of Peptides to the Chaperone DnaK. J Mol Biol. 2013;425: 2463–2479. doi:10.1016/j.jmb.2013.03.041

25. Tapley TL, Cupp-Vickery JR, Vickery LE. Sequence-Dependent Peptide Binding Orientation by the Molecular Chaperone DnaK. Biochemistry. 2005;44: 12307–12315. doi:10.1021/bi051145r

26. Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, et al. Structural Analysis of Substrate Binding by the Molecular Chaperone DnaK. Science 1996;272: 1606–1614. doi:10.1126/science.272.5268.1606

27. Taylor IR, Ahmad A, Wu T, Nordhues BA, Bhullar A, Gestwicki JE, et al. The disorderly conduct of Hsc70 and its interaction with the Alzheimer's-related Tau protein. J Biol Chem. 2018;293: 10796–10809. doi:10.1074/jbc.RA118.002234

28. Nachman E, Wentink AS, Madiona K, Bousset L, Katsinelos T, Allinson K, et al. Disassembly of Tau fibrils by the human Hsp70 disaggregation machinery generates small seeding-competent species. J Biol Chem. 2020;295: 9676–9690. doi:10.1074/jbc.RA120.013478

29. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630: 493–500. doi:10.1038/s41586-024-07487-w

30. Dice JF, Chiang HL, Spencer EP, Backer JM. Regulation of catabolism of microinjected ribonuclease A. Identification of residues 7-11 as the essential pentapeptide. J Biol Chem. 1986;261: 6853–6859.

31. Johnson OT, Gestwicki JE. Multivalent protein-protein interactions are pivotal regulators of eukaryotic Hsp70 complexes. Cell Stress Chaperones. 2022;27: 397–415. doi:10.1007/s12192-022-01281-1

32. Nury C, Redeker V, Dautrey S, Romieu A, van der Rest G, Renard P-Y, et al. A Novel Bio-Orthogonal Cross-Linker for Improved Protein/Protein Interaction Analysis. Anal Chem. 2015;87: 1853–1860. doi:10.1021/ac503892c

33. Redeker V, Pemberton S, Bienvenut W, Bousset L, Melki R. Identification of Protein Interfaces between α-Synuclein, the Principal Component of Lewy Bodies in Parkinson Disease, and the Molecular Chaperones Human Hsc70 and the Yeast Ssa1p. J Biol Chem. 2012;287: 32630–32639. doi:10.1074/jbc.M112.387530

34. Fox-Erlich S. Structural conservation of a short, functional, peptide-sequence motif. Front Biosci. 2009;14: 1143-1151. doi:10.2741/3299

35. Popelka H, Klionsky DJ. Structural basis for extremely strong binding affinity of giant ankyrins to LC3/GABARAP and its application in the inhibition of autophagy. Autophagy. 2018;14: 1847–1849. doi:10.1080/15548627.2018.1522884

36. Seike T, Terasawa K, Iwata T, Guan J-L, Watabe T, Yokoyama S, et al. Site-specific photo-crosslinking of Hsc70 with the KFERQ pentapeptide motif in a chaperone-mediated autophagy and microautophagy substrate in mammalian cells. Biochem Biophys Res Commun. 2024;736: 150515. doi:10.1016/j.bbrc.2024.150515

37. Nordquist EB, English CA, Clerico EM, Sherman W, Gierasch LM, Chen J. Physics-based modeling provides predictive understanding of selectively promiscuous substrate binding by Hsp70 chaperones. PLoS Comput Biol. 2021;17: e1009567. doi:10.1371/journal.pcbi.1009567

38. Zhang P, Leu JI-J, Murphy ME, George DL, Marmorstein R. Crystal Structure of the Stress-Inducible Human Heat Shock Protein 70 Substrate-Binding Domain in Complex with Peptide Substrate. PLoS One. 2014;9: e103518. doi:10.1371/journal.pone.0103518

39. Kasper P, Christen P, Gehring H. Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK. Proteins Struct Funct Genet. 2000;40: 185–92. doi:10.1002/(sici)1097-0134(20000801)40:2<185::aid-prot20>3.0.co;2-x

40. Yang J, Nune M, Zong Y, Zhou L, Liu Q. Close and Allosteric Opening of the Polypeptide-Binding Site in a Human Hsp70 Chaperone BiP. Structure. 2015;23: 2191–2203. doi:10.1016/j.str.2015.10.012

41. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. 2004;32: W500–W502. doi:10.1093/nar/gkh429

42. Burley SK, Bhatt R, Bhikadiya C, Bi C, Biester A, Biswas P, et al. Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. Nucleic Acids Res. 2025;53: D564–D574. doi:10.1093/nar/gkae1091

43. Berman HM. The Protein Data Bank. Nucleic Acids Res. 2000;28: 235–242. doi:10.1093/nar/28.1.235

44. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589. doi:10.1038/s41586-021-03819-2

45. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2: 19–25. doi:10.1016/j.softx.2015.06.001

46. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J Comput Chem. 2013;34: 2135–2145. doi:10.1002/jcc.23354

47. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31: 671–690. doi:10.1002/jcc.21367

48. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Science. 2018;27: 112–128. doi:10.1002/pro.3280

49. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79: 926–935. doi:10.1063/1.445869

50. Debye P. Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. Math Ann. 1909;67: 535–558. doi:10.1007/BF01450097

51. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126. doi:10.1063/1.2408420

52. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981;52: 7182–7190. doi:10.1063/1.328693

53. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

54. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B. 2001;105: 6474–6487. doi:10.1021/jp003919d

Supplementary information Captions

S1 Table. List of experimentally verified CMA substrates in humans along with the canonical CMA-targeting motif in them. The superscript numbers indicate the starting position of the CMA-targeting motif.

S1 Fig. CMA-targeting motifs in substrate proteins stably binds to the lid/SBDα of Hsc70 only in the presence of extended/adjacent flanking residues.

(A) Time evolution of the minimum distance between Hsc70 SBDα and the KFERQ-like motif of 3 different substrates. (B) Time evolution of the minimum distance between HSC70 SBDα and the extended KFERQ-like motif of 15 different CMA substrates.

S2 Fig. Extended CMA-targeting motifs in substrate proteins display promiscuous orientation in the binding hydrophobic cavity of Hsc70, with major energetic contributions from hydrophobic residues in substrate protein.

643 (A) Orientation of CMA substrate peptides (CMA-targeting motifs along with flanking residues or
644 extended CMA-targeting motifs) within the hydrophobic cavity of Hsc70 at 1μs of MD
645 simulations, with the substrate segment depicted in gray cartoon representation. (B) Residue-wise
646 decomposition of binding energies of extended CMA substrates peptides. Only values less than -
647 2.5 kcal/mol are shown for clarity.

648

649 **S3 Fig. Interacting residue pairs between Hsc70 and extended CMA-targeting motifs differ**
650 **across substrates, suggesting a flexible and promiscuous mode of binding.** The plots depict the
651 minimum distance between representative residue pairs with binding energies less than -2.5
652 kcal/mol.

653

654 **S4 Fig. Promiscous mode of binding of extended CMA-targeting motifs in the hydrophobic**
655 **cavity of Hsc70's lid domain was observed in presence of different force fields.**

656 (A) Snapshot at 1μs of MD simulations depicting orientation of extended CMA-targeting motifs
657 in two substrate peptides within the hydrophobic cavity of Hsc70 using three different force fields
658 (Amber: rosy brown, OPLS:dark khaki, and CHARMM36:cornflower blue), residues with binding
659 energies less than -2.5 kcal/mol are highlighted in the ball and stick model. Residues which are
660 part of CMA-targeting motif(s) are labelleled in red color while the flanking residues are labelled
661 in black. (B) Residue-wise decomposition of binding energies of PARK7 and α-Synuclein using
662 three different force fields (Amber, OPLS, and CHARMM36), with different values in each force
663 field indicating the flexibility of substrate binding modes.

664 **S5 Fig. AlphaFold 3 predicted complexes of Hsc70 and CMA-targeting motifs of fifteen**
665 **shortlisted CMA substrates suggests SBDβ to be the possible binding site.**KFERQ-like motifs
666 (light sea green) are bound to hydrophobic cleft of SBDβ (dark gray) in two hsc70 conformations
667 i.e. open and closed(*). For Clarity, only the SBDβ domain is shown.

668

669 **S2 Table. Segregation of residues within CMA-targeting motifs into distinct binding pockets**
670 **(labelling as per previous literature**[37,38]**) based on their stable interaction sites.** Data are
671 shown for seven closed Hsc70–substrate complexes that remained stable during the MD
672 simulations.

673

674 **S6 Fig. Backbone of CMA-targeting motifs in substrate proteins are involved in interaction**
675 **with backbone of residues in SBDβ of Hsc70, suggesting preservation of backbone alignment**
676 **in the interaction mode of the substrate.** Figure depicts hydrogen bonding (yellow color)
677 between backbone of CMA-targeting motif in seven CMA substrates (Abeta, AnnexinA1, PEA-
678 15, HBB, RND3, α-Synuclein and PARK7) with backbone of two key residues (having >90%
679 occupancy) in SBDβ viz. Ala406 and Thr429. The interacting residues are explicitly labelled.

680

681 **S7 Fig. Water-mediated hydrogen bonds stabilize the binding between the lid domain of**
682 **Hsc70 and the CMA-targeting motif in the substrate protein.**

683 **(A)** The time evolution of the number of water molecules forming hydrogen bonds between the
684 CMA-targeting motif in different substrate proteins and the lid domain of Hsc70 for four stable

complexes in the MD simulations with Hsc70 in the closed conformation. (**B)** The time evolution of the number of water molecules forming hydrogen bonds between the CMA-targeting motif in different substrate proteins and the lid domain of Hsc70 for three stable closed complexes in the MD simulations with Hsc70 which were initially predicted in the open conformation by AF3.

S8 Fig. Transient water molecules stabilizes the binding of the substrate peptides with the lid domain of Hsc70 at different time frames. Snapshots of seven closed stable complexes forming water mediated hydrogen bonding between lid domain and CMA-targeting motifs at three different time frames. Interacting residues of Hsc70 lid domain (rosy brown) and CMA-targeting motifs (light sea green), along with water molecules are represented in ball-stick form.

S9 Fig. Water-mediated hydrogen bonds stabilize the binding between the SBDβ domain of Hsc70 and the CMA-targeting motif in the substrate protein.

(A) Time evolution of the number of water molecules forming hydrogen bonds between the CMA-targeting motif in different substrate proteins and SBDβ domain of Hsc70 for four stable complexes in the MD simulations, with Hsc70 in the closed conformation. (B) Time evolution of the number of water molecules forming hydrogen bonds between the CMA-targeting motif in different substrate proteins and the SBDβ domain of Hsc70 for three stable complexes in the MD simulations with Hsc70 predicted in the open conformation and converted to closed conformation.

705 **S3 Table. List of complexes generated using AlphaFold 3 utilizing segments of Hsc70 and**
706 **substrate proteins, along with their confidence scores (iptm and ptm).**

707 Data availability statement

708 The data files related to this study have been made available at:
709 https://github.com/CSBLabNM/Project_HSC70.git