Reliable repurposing of antibody interactome inside the cell

Caitlin M. O'Shea[1], Rushba Shahzad[1], Kimia Aghasoleimani[1], Stuart Newman[1], Jiraporn Panmanee[2], Leonard C. Schalkwyk[1], Greg N. Brooke[1], Fiona E. Benson[3], James S. Trimmer[4], Daryl A. Bosco[5], Takao Fujisawa[6,7], Hidenori Ichijo[6,7], Neil R. Cashman[8], Stanislav Engel[9], Gareth S. A. Wright[1]*

1. School of Life Sciences, University of Essex, Colchester CO4 3SQ, United Kingdom.
2. Research Center for Neuroscience, Institute of Molecular Biosciences, Mahidol University, Nakhon Pathom 73170, Thailand.
3. Division of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster, LA1 4YQ, United Kingdom.
4. UC Davis/NIH NeuroMab Facility, Department of Physiology and Membrane Biology, University of California School of Medicine, Davis, CA, United States.
5. UMass Chan Medical School, 368 Plantation Street, Worcester, MA 01605, United States.
6. Laboratory of Cell Signalling, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan.
7. Cell Signalling and Stress Responses Laboratory, Advanced Research Initiative, Institute of Integrated Research, Institute of Science Tokyo, Tokyo, Japan.
8. Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, V6T 2B5, Canada.
9. Department of Clinical Biochemistry and Pharmacology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel.

*Correspondence to gareth.wright@essex.ac.uk

Abstract

In biology proximity is paramount and eighty-five percent of the human proteome has at least one documented interacting monoclonal antibody. These molecules penetrate the cytoplasm poorly and are very often non-functional within the cell. Sequence analysis of 10^6 antibody variable domains alongside the cytoplasmic human proteome shows charge and isoelectric point are characteristics ill adapted to intracellular monodispersity. Characterisation of forty-five single-chain variable fragment (scFv) intrabodies expressed in human cells confirmed charge to have the greatest impact on solubility. We created new interdomain linkers, optimised scFv domain orientation and found variable heavy domain framework sites to be generally positively charged, and promote insolubility, but be amenable to optimisation. This is applied in combination to reduce the search space and refine the products of AI-led inverse folding to create highly soluble, abundant and thermally stable intrabodies that maintain parent antibody epitope recognition. Over six hundred intrabody sequences are described targeting sixty cytoplasmic proteins with linear, conformational, post-translational modification or oligomeric state specificity. Interactions were validated for p53, α-synuclein, SOD1, polyQ, FUS/TLS, UCHL1 and GFP. This approach removes obstacles hindering intracellular repurposing of the vast sequenced antibody interactome with applications relevant to many human disease states.

Introduction

Antibodies are the precision arm of the immune system, as well as valuable research tools and medicines. More than 10^9 antibodies have been sequenced[1] and over 10^5 are defined with a known antigen[2]. Antibodies are composed of heavy and light chains that are translated into the endoplasmic reticulum. After assembly they are transported through the Golgi, during which they are glycosylated and form intra- and inter-domain disulphide bonds, before secretion. Extracellular native and therapeutic antibodies can be internalised through endocytic pathways but are degraded in lysosomal compartments with few accounts of effective direction to the cytoplasm[3,4]. This limits our ability to deploy the antibody interactome inside the cell.

Pairs of antibody variable domains housing complementarity determining regions (CDRs) can be peptide linked to form single-chain variable fragments (scFvs)[5,6] (**Fig 1a & b**) and expressed within cells as intrabodies. However, scFv intrabodies also have notoriously low solubility leading to aggregation and loss of function[7]. The development of new synthetic biology applications could be propelled by the ability to generate custom intracellular biological interactors. Grafting CDR loops onto variable domain framework regions with desired characteristics[8], engineering physicochemical properties by rational mutagenesis[9], removing cysteines that form disulphide bonds and directed evolution[10] have been applied to this problem but no strategy consistently facilitates antibody to intrabody reformatting. Maintenance of specificity and affinity from parent antibody to derived scFv intrabody is important but paratope–epitope interactions can only operate if a correctly structured scFv can exist in and diffuse through the cell. High solubility, thermal stability and demonstrable epitope binding are therefore the key characteristics of deployable scFv intrabodies and should be the focus of a broadly applicable reformatting strategy.

Here we initially define the intracellular solubility of forty-five scFvs that have applications in the study and treatment of neurodegenerative diseases including Alzheimer's, Parkinson's, Huntington's diseases and amyotrophic lateral sclerosis. We use these data to discover the determinants of scFv intrabody solubility and develop a system to predict scFv solubility based on molecular charge. In addition, we describe new interdomain linkers that contribute to intracellular solubility. From this work emerges a rationale for scFv intrabody design that ensures first construct success making the use of these molecules reliable and economical. Applying our approach to the products of AI-led molecular redesign allowed *in silico* reformatting of 672 non-redundant antibodies as scFv intrabodies targeting intracellular proteins including tau, α-synuclein, SOD1, TDP-43, p53, HIF-1α, histones and ubiquitin along with protein phosphorylation, citrullination and acetylation post-translational modifications. The sequences of all intrabodies described are made available following the principles of open source science[11]. This should aid design and implementation of synthetic molecular pathways based on collocation of component macromolecules and therapeutics able to interact with protein sub-populations known to co-segregate with disease all within the living cell.

Results

Variable domain charge drives intrabody solubility

The immunoglobin (Ig) protein fold superfamily is widely found throughout nature. It predates the existence of vertebrate antibodies and its members are often found at high concentrations within cells[12–14]. For example, Cu,Zn-superoxide dismutase has an Ig fold, is near ubiquitous throughout life, abundant and very often found in the cytoplasm where it maintains a disulphide bond[14,15]. Conversely, scFv expression in the cytoplasm often results in poor solubility and functionality. To find why variable fragments (Fv) are poorly adapted to the intracellular environment we compared the physicochemical properties of the complete and cytoplasmic human proteomes with 10^6 paired variable light (VL) and heavy (VH) domains from mice, rats and humans. Variable domains exist within the bounds of hydrophobicity and stability common in the cytoplasm (**Fig. S1a, b, c & d**). **Fig. 1c** shows extracellular proteins, including full-length antibodies and variable domains, are likely to have an isoelectric point (pI) above 7.4. Seventy percent of Fvs are found in this region and are ill adapted to a role as an scFv intrabody. Charge strongly influences pI and only 0.4 % of the human proteome with a charge of zero or less has a pI over 7.4 (**Fig. S1e**). Direct comparison of net charge indicates antibody variable domains are generally more electropositive than cytoplasmic proteins (**Fig. 1d**).

We defined the solubility of forty-five scFv intrabodies and compared that experimental data to predictions made by nine sequence-based solubility calculators (**Fig. S2**). The ProteinSol algorithm predicted scFv solubility well (R^2 0.715) using a weighted combination of protein features focusing on charge and hydrophobicity[16]. We calculated 79 individual physicochemical characteristics for each protein aiming to unpick the drivers of scFv intrabody solubility and find a single, easily modifiable characteristic (**Supplementary File S1**). Hydrophobicity, aliphatic index, charge or isoelectric point of CDR loops, individual variable domains or their framework regions in addition to predicted thermal stability[17] and whole molecule predicted local difference distance test (pLDDT) score calculated from AlphaFold3 structural models were poor predictors of solubility (**Fig. 1e**, **Supplementary table S1**). The latter was included because AlphaFold pLDDT score is a good predictor of disorder[18,19]; the antithesis of stability. Whole molecule net charge at physiological pH has a negative linear correlation with solubility over the range +3 to -20 (R^2 0.75) (**Fig. 1f**), exceeding correlations with related characteristics including charge at pH 5.5[20] (R^2 0.69), pI (R^2 0.39), and all solubility calculators (**Fig. S2**). This relationship is implemented in the bioinformatics website scFvright which isolates Fv protein sequences from full-length antibodies or takes full-length scFv with fusion tags and linkers then predicts the percentage likely to be found in the soluble fraction (https://scfvright.essex.ac.uk).

Large scale prediction of scFv intrabody solubility

Using the relationship presented in **Figure 1f**, we predicted the intracellular solubility of 68,551 Fv and scFv proteins. Only 0.02 % of unmodified Fvs would have high solubility within the cell, which we have defined as over 70 % partitioning in the soluble mammalian cell fraction (**Fig. 1g**). Addition of canonical Gly-Gly-Gly-Gly-Ser (G_4S) linkers and a C-terminal HA tag yields 0.03 % with high solubility reflecting the historical problem of scFv expression in the cytoplasm. Kabayama *et al.* 2020 introduced the combination of 3xFLAG

and HA tags to improve scFv solubility and monodispersity in the cell[20] by reducing net charge by -9, but still only 3.9 % are highly soluble.

Electronegative linkers improve scFv intrabody solubility

High scFv solubility requires a net charge of less than -15 (**Fig 1e**). However, the pre-eminence of distributed negative charge means Asp/Glu residues can be incorporated or Arg/Lys residues removed in any location that does not affect target binding or structural stability; ideally in solvent exposed linker sites, fusion tags or framework sites outside the paratope. Ninety-seven percent of scFvs contain interdomain linkers that do not contribute to, or increase, net charge (**Fig 2a**). Ninety-two percent are G_4S linkers[6] or closely related sequences that are not optimised for expression of intracellular fusion proteins. Modification of the G_4S sequence to include aspartate or glutamate in place of serine to yield $(G_4D)_4$ or $(G_4E)_4$ and contribute to solubility by adding four negative charges (**Fig. 2b & c**). Linkers $(G_2D)_7$, $(G_2E)_7$, $(GD)_{10}$ and $(GE)_{10}$ also add negative charge (**Fig. 2c**). Each provides sufficient flexibility to maintain target binding to FUS/TLS protein by an scFv intrabody derived from anti-FUS antibody N152B/23 (**Fig. 2d**). Assessment of scFv linker variants indicate each confers different properties with $(G_2E)_7$ having the highest segregation into, and abundance in, the soluble fraction (**Fig. 2e**). Therefore, different combinations of serine, aspartate and glutamate in the linker would allow tuning of scFv negative charge from zero to -10 in addition to native Fv and fusion tag charge. These linkers require little expertise to implement and in combination with 3xFlag and HA tags are predicted to confer greater than 70 % and 50 % solubility on 22 % and 84 % of scFv intrabodies respectively when using $(G_4D)_4$ or $(G_4E)_4$ versions (**Fig. 1g**). This is illustrated by the high solubility of scFv intrabodies with charged linkers derived from antibodies targeting tau (**Fig. 2f & g**).

Electronegative modifications to framework sites improve solubility

Aspartate scanning through CDR loops has been shown to reduce variable domain aggregation[21] but is unfavourable given the possibility that CDR mutations may change epitope specificity or affinity. To find an alternative approach we analysed the charge of framework and CDR (**Fig. 1b**) sequences from 166,179 Fv molecules (**Fig. 3a**). VH domain CDR and framework sequences have characteristic, non-overlapping charge frequency distributions. VH CDR loops are generally electronegative while VH framework regions are electropositive. VH di-polarity marginally reduces as antibodies mature but is absent in VL domains (**Fig. 3a**). Arginine and lysine residues are not found within the core Ig-fold β-sandwich but very often project into solvent, except for a ubiquitous salt-bridge between VL domain Arg/Lys61-Asp82 and VH domain Arg/lys66-Asp86 (Chothia numbering) found at the N-terminus of β-strands D and F (**Fig. 3b**). Mutations at this site in an scFv intrabody derived from the 3B5H10 poly-glutamine specific antibody[22] reduce solubility likely by reducing thermal stability (**Fig 3c**). With this exception, positively charged amino acids in variable domain framework sites, particularly VH, may be good sites to introduce negative charge.

scFv domain order effects intrabody solubility

scFvs are typically designed with VH domain N-terminal to the VL domain (VHVL)[6]. However, variable heavy domains are known to have heterogenous physical and chemical properties, including thermal stability[23], in comparison with VL domains. Analysis of fifteen scFv intrabodies in both VHVL and VLVH orientations shows they have similar solubility when the net charge of the VH domain is zero or less (74.1 ± 17.3 % and 72.8 ± 17.2 % respectively, mean ± one standard deviation). However, VHVL scFv intrabodies with positively charged VH domains have reduced solubility with respect to the VLVH form (49.3 ± 21.9 % and 58.2 ± 3.5 % respectively, mean ± one standard deviation) (**Fig. S3**). Variable heavy domain charge is negatively correlated with overall VHVL scFv solubility (R^2 0.38) but poorly corelates with VLVH solubility (R^2 0.15) (**Fig S3b**). A positively charged VH domain often seems to require molecular chaperoning by its accompanying VL domain which cannot happen if it is translated first. VLVH scFv intrabodies have overall better solubility and are more reliable therefore, but this effect can be overcome by charge swap mutations that increase net charge of the VH domain (**Fig. S3c**).

scFv intrabody construction using inverse folding with solubility prediction

Data point residuals in **Figure 1f** indicate charge is not the only factor that influences scFv intrabody solubility. We used ProteinMPNN_{sol}[24] to increase scFv intrabody thermal stability and solubility while implementing domain orientation and charge rules to restrict design options and filter outputs respectively for poly-glutamine-specific antibody 3B5H10-derived[22] scFv intrabodies (**Fig. 4a**). The inverse folding approach creates protein sequences that are able to satisfy the geometric constraints of a given protein backbone structural model[25,26] and ProteinMPNN_{SOL} is known to increase surface hydrophilicity[24]. All inverse folded scFv intrabodies display near complete solubility and high thermal stability in comparison with native and human redesigned constructs (**Fig. 4b & c**). However, sequences produced with model structure coordinate noise of 0.20 Å in comparison with 0.02 Å have higher abundance in the soluble fraction (**Fig. 4d**). Increased abundance and thermal stability has been noted previously on inverse folded proteins using 0.20 Å model noise expressed in bacteria[27]. Increased model noise is inversely proportional to sequence recovery[28] which may facilitate beneficial changes. Minimally restricting fixed residue positions to only CDR loops or allowing redesign of the interdomain linker did not produce protein capable of interaction with co-expressed GFP-polyQ (**Fig. 4e,** constructs 6-9). Variable domain interfaces affect antigen binding[29]. This may be as part of an extended epitope-paratope interaction surface[30] or through domain, and therefore CDR loop orientation, given two thirds of the β-sheets N- and C-terminal to each CDR loop are found in the interface. As CDR loops very often lack structure until interaction with an epitope we chose to focus on maintenance of domain orientation. Fixing CDR loops, linker and domain interface β-sheets again produced highly soluble, abundant and stable proteins but only one strongly immunoprecipitated GFP-polyQ (**Fig. 4e**, construct 5). Sequences of each construct can be found in **Supplementary File S2**.

To test reproducibility, we reformatted the misfolding-specific anti-SOD1 antibody MS785[31] as a scFv intrabody. A traditional $(G_4S)_4$+HA tag construct, or a human redesigned variant with charge swap mutations, were both insoluble (**Fig. 4f**). Inverse folding with 0.02 Å or 0.20 Å model noise produced highly soluble proteins with the latter abundant in the soluble

fraction; >4000-fold abundance increase and approximately 10 ^{o}C increase in melting temperature in comparison with the best human design (**Fig. 4f & g**). This MS785 scFv variant reliably pulls down A4V SOD1 with an optimum at physiological pH and NaCl concentration (**Fig. 4h & i**).

Impact of inverse folding with ProteinMPNN$_{sol}$ on scFv intrabodies

Effective scFv construction and redesign using the rules established above (**Fig. 5a**) can be applied to any antibody. Accordingly, scFv intrabodies were produced by inverse folding from 672 Fv sequences with specificity for cytoplasmic human proteins and oligomeric or post-translationally modified states thereof (**Fig. 5b**). Across 672,000 derivative scFv sequences no position was consistently substituted in all inverse folding outputs but some are predisposed to changes, including domain N- and C-termini (**Fig. 5c**) and surface sites, for example, small hydrophobic amino acids totalling 98.1 % of original sequences at highly solvent exposed VH position 11 are substituted for polar or charged amino acids in 93.9 % of inverse folded variants. Conversely, disulphide bonding cystines and salt-bridge charged residues are strictly maintained with also VL amino acids 61-69 and VH 149, 151-154 showing strong conservation through inverse folding. The consensus sequences of parent and inverse folded progeny show a preference for substitutions in VH domains using 0.02 Å model (16 % and 32 % of non-fixed VL and VH positions respectively) but this discrepancy is reduced when using the 0.20 Å model (31 % and 34 % respectively). Furthermore, changes incorporated by model 0.20 Å reduce mean net charge of VH framework sites from +2.1±1.8 to -1.6±1.6 (mean ± one standard deviation). This reduces the number of VH domains with net positive charge framework regions from 83.9 % to 9.3 % (**Fig. 5d**). This effect is not seen for VL domain framework sites that are natively better adapted to the cytoplasmic environment (**Fig. 5d & 3a**). As a result of inverse folding with ProteinMPNN$_{sol}$ using 0.20 Å model, consensus VH domain net charge is reduced from +1 to -1 and VL maintained at +2. Conversely, model 0.02 Å increases VL domain net charge from +2 to +5 and maintains VH charge. This effect observed using 0.20 Å model is mediated by a reduction of the total number of charged residues but disproportionately reducing the number of positively charged amino acids (**Fig. 5c**) and is mediated by changes to only forty-four (15 %) of the total amino acid positions. Of those forty-four, forty substitutions (90 %) are to amino acids with side chains largely or entirely solvent exposed sites. Two of the remaining four substitutions are found on a loop (VL Met4Leu) or the end of a β-strand away from the core β-sandwich (VL Leu78Val), and two maximise hydrophobic packing around the β-sandwich core (VH Glu6Val and VH Ile69Leu). These may contribute to increased VH domain thermal stability.

Two intrabodies derived from antibodies 11D3 and 421 targeting the C-terminus of p53 tumour suppressor protein were tested to validate the redesign process. Unmodified scFv derived from 11D3 and 421 antibodies are predicted to have 40 % and 30 % solubility respectively based on **Figure 1f**. Previous work showed scFv 421 to express well but with staining resembling aggregation whereas no expression was observed for scFv 11D3 [32]. By contrast, redesign yields good expression with very high solubility (>99 %) (**Fig. 5e**) and both intrabodies maintain parent antibody interactions with p53 (**Fig. 5f**). As further validation, we created three new highly soluble intrabodies for ubiquitin carboxy-terminal hydrolase 1

(UCHL1) (**Fig. 5g**) and one for GFP (**Fig. 5h**) and validated their target interactions. Sequences are reported in **Supplementary File S4.**

State specific recognition

Repurposing well characterised conformation and post-translational state specific antibodies that interact with structures that co-segregate with neurodegenerative disease is an advantage of our approach. **Figure 6a** shows interaction assessments for five conformational specific scFv intrabodies derived from parent antibodies shown to be specific for forms of SOD1 found in familial or sporadic ALS. C4F6 scFv intrabody has reactivity to ALS-associated G93A and A4V mutants but not wild-type or wild-type-like D90A replicating the specificity of the parent antibody[33,34]. Intrabodies derived from AMF7-63, α-miSOD, SE21 and MS785 antibodies follow the same pattern[31,35–37]. This is despite each intrabody interacting with a distinct and separate loop epitope within the SOD1 molecule (**Fig. 6b**). Inverse folded intrabody specificity was determined using the non-SOD1 specific Z-FUS intrabody with $(G_4E)_4$ linker which was found not to pull down A4V SOD1 (**Fig. 6c**). Sequences are reported in **Supplementary File S5**.

Antibodies with a proclivity for binding to the different quaternary structures presented by α-synuclein[38–42] are reactive to epitopes found in Parkinson's disease and other α-synucleinopathy tissues but have also been the subject of many clinical trials[43,44] (**Fig. 6d**). Reformatting as intrabodies (**Supplementary File S6**) and comparison of α-synuclein interaction by co-immunoprecipitation following co-expression indicates the monomeric form prevails in our HEK293T expression system with oligomer or fibril states either absent or below the limit of detection (**Fig. 6e**). This work highlights the ability of these interactors to differentiate an intracellular protein subpopulation and will surely expand to recognition of post-translational state and heterocomplex species.

Discussion

The relationship between protein charge and solubility is well characterised[45] and introducing negatively charged mutations or fusion tags has been shown to improve scFv behaviour[20,46,47]. We take this concept to its conclusion by showing scFv intrabody solubility has a negative, linear relationship with net charge (**Fig. 1f**) and to achieve ≥80 % solubility an scFv intrabody should have charge less than -15. Calculated integer net charge is likely a proxy for surface charge. The size of antibody variable domains, their folding around a hydrophobic core and association across a hydrophobic interface mean all variable domain charged residues are at least partially solvent accessible, but solubility predictions may be improved by greater granularity in surface charge calculations.

It is very unlikely a charged protein will have an isoelectric point equivalent to physiological pH (**Fig. S1e**). That equivalence is avoided in the human proteome (**Fig. 1c & d**) due to the likelihood of self-association and aggregation[48]; any charge is better than no charge. However, positively charged proteins interact with nucleic acids and cell membranes thereby inhibiting free diffusion and monodispersity. DNA and RNA are negatively charged and it may be that most cytoplasmic proteins have evolved a similar charge profile to prevent unwanted interactions. In the extracellular space that selective pressure is absent but is replaced by the necessity to diffuse through the charged mesh of proteoglycans. This manifests as very different isoelectric point and charge distribution profiles between intracellular and variable domain protein sequences (**Fig. 1c & d**). Nevertheless, charge is not the only contributor to protein solubility. High thermal stability prevents exposure of hydrophobic, aggregation-prone residues normally sequestered in the molecular core. This is well articulated by inverse folded 3B5H10-derived scFv variants with high thermal stability, solubility and near neutral charge, but these molecules do not retain epitope interactions (**Fig. 4**, constructs 8 & 9). We found that introducing positive to negative charge swap mutations had a cumulative effect on solubility but compounding five or six mutations reduced protein abundance. This is indicative of reduced thermal stability and resulting propensity for degradation. By contrast, inverse folding with ProteinMPNN$_{SOL}$ changed ~30 % of non-fixed positions, increased thermal stability and solubility, while maintaining function. Evidently it is possible to break out of strictures imposed by the inverse relationship between stability and function.

The scFv intrabodies with experimentally defined solubility described here (**Supplementary File S1**) were created from monoclonal antibody sequences found in public databases or retrieved by hybridoma DNA or mass spectrometry protein sequencing. They comprise a range of molecules with native net charge -6 to +6 representative of 95 % of the charge distribution presented in **Figure 1d**. The documented antibody interactome can now be effectively repurposed inside the cell. While that interactome is large, it is the result of *in vivo* generation; antigen presentation in animals, hybridoma technology, B-cell sequencing and two-hybrid screening that are limited by scale. *De novo* antibody design, specifically prediction of CDR-epitope interactions is being revolutionised[49–51], but currently requires production and testing of a prohibitively large numbers of putative interactors. This process will certainly become more efficient and accessible. The scFv design rules established here ensure first construct design success and provide an access route for either *in vivo* or *in silico* generated antibodies targeting any molecule within the cell. That may be interactions with conformationally plastic or intrinsically disordered proteins that are difficult to address with small molecules, protection of functional sites, inhibition of protein-protein interactions,

molecular chaperoning, enzyme inhibition, providing substrate recognition for biological protein editors like recently described for tau and SOD1[52,53], or the molecular interactions to collocate elements of new enzymatic pathways. The intrabodies described are derived from antibodies used in research but also those that have been tested in clinical trials that did not meet end point goals. Neurodegenerative protein misfolding disease immunotherapies in particular are plagued by this limited efficacy. Bringing those interactions inside the cell where disease related changes occur may expedite medicine development especially given the burgeoning effectiveness, acceptance and safety of gene therapies. This approach also capitalises on the large amount of research that has already been performed on those molecules in a manner analogous to small molecule drug repurposing. Introduction of synthetic molecules into humans and other animals will require careful consideration and testing of their immunogenicity, however we note a recent report of variant of ProteinMPNN that has been tuned to reduce MHC-I visibility[54].

Methods

Expression plasmid construction and site directed mutagenesis

All intrabody coding sequences were cloned into Twist Bioscience's CMV BetaGlobin WPRE Neo plasmid vector without coding for sequences for N-terminal localisation signals but including, where stated, N- and C-terminal tags and interdomain linkers. Site directed mutagenesis was performed using Platinum SuperFi II DNA Polymerase and sequence verified.

Solubility assay

HEK293T cells were cultured in DMEM medium with 10 % foetal calf serum supplemented with 100 units/ml penicillin and 100 μg/ml streptomycin in six-well plates. Cells were transfected with intrabody expression plasmids using Turbofect reagent and harvested after 24 hours. Media was removed, 300 μl phosphate buffered saline (PBS) with protease inhibitor cocktail (Roche EDTA-free, cOmplete) was added and cells lifted with gentle pipetting. Cells were lysed by sonication at 4 °C. Soluble and insoluble cell fractions separated by centrifugation at 20,000 g for 45 minutes at 4 °C. The insoluble component was resuspended in 300 μl PBS. Reducing and denaturing SDS-PAGE was performed and proteins transferred to nitrocellulose membrane by dry western blotting. Membranes were probed with anti-HA (Proteintech 81290-1, 66006-2), anti-FLAG (Proteintech HRP-66008) or anti-myc (Merck 05-724), anti-FUS/TLS (Proteintech 68262-1), anti-SOD1 (Proteintech 67480-1), anti-α-synuclein (Proteintech 66412-1), anti-UCHL1 (Proteintech 66230-1) antibodies with appropriate secondary antibodies (LI-COR IRDye 680RD goat anti-rabbit 926-68071, IRDye 800CW goat anti-mouse 926-32210, Proteintech HRP-conjugated goat anti-mouse SA00001-1 or Invitrogen HRP-conjugated goat anti-rabbit A16104).

Immunoprecipitation

Immunoprecipitation was performed at room temperature. HEK293T cells were cultured and transfected as above. Cells were lifted in 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 % NP-40,

1 mM EDTA, 5 % glycerol with protease inhibitor, incubated for 30 minutes and centrifuged at 12,000 g for 10 minutes. Supernatant was applied to pre-equilibrated HA-Trap magnetic agarose beads (Proteintech ATMA-200) and washed with PBS with protease inhibitor cocktail then eluted by heating to 95 ^{o}C for 5 minutes in SDS-PAGE sample buffer. SDS-PAGE and western blot was performed on the samples.

Statistical analysis

Unless otherwise stated, two-tailed, two-sample equal variance Student's t-test was applied.

Cellular thermal shift assay

HEK293T cells were cultured and transfected as above then resuspended in 400 μl PBS with protease inhibitor cocktail and 40 μl aliquoted into PCR tubes. Cell suspensions were heated at 35, 40, 45, 50, 55, 60, 65, 70, 75 ^{o}C for 5 minutes then frozen at -80 ^{o}C. Cells were then thawed and sonicated and centrifuged at 20,000 g for 45 minutes at 4 °C and the supernatant heated with SDS-PAGE loading buffer before SDS-PAGE and western blot.

Hybridoma DNA and mass spectrometry protein sequencing

Mouse anti-TDP43 C-terminus DB9 antibody and rat anti-SOD1 MS785 antibody expressing hybridomas were sequenced at The Pirbright Institute, UK. Briefly, 5´-Rapid amplification of cDNA ends-ready cDNA was prepared followed by PCR amplification of the heavy and light chain Ig variable regions. The products were subsequently cloned into a sequencing vector and Sanger sequenced. C4F6 SOD1 conformational specific antibody was sequenced by tryptic digest mass spectrometry as described previously[55].

***In silico* protein physicochemical analysis**

In silico physicochemical analysis was performed on 1,073,713 paired variable domain sequences retrieved from the Observed Antibody Space (OAS) database[56]; 843 proteins annotated as cytoplasmic only localisation by the Human Protein Atlas[57] and 20,590 proteins representing the complete, but non-isoformic, human proteome (UP000005640) downloaded from Uniprot[58]. The isoelectric point of 127 full-length clinically approved IgG antibodies of multiple subtypes was also calculated. Isoelectric point and peptide mass were calculated with standalone Isoelectric Point Calculator[59] taking the average value for pI. Charge at physiological pH was calculated as the difference between the number of arginine and lysine residues and aspartate and glutamate; (K+R)-(D+E). Charge at pH 5.5 was calculated with the addition of the number of histidines; (K+R+H)-(D+E). Grand average of hydropathy (GRAVY) was calculated as the ratio of the count of each amino acid scaled using Kyte & Doolittle paramaterisation[60] to the total number of amino acids. Aliphatic index was calculated as the mole percent fraction of alanine, valine, isoleucine and leucine amino acids[61]. Mean pLDDT was calculated as the mean of all atom pLDDT scores from variable domains and linkers. DeltaG IF and DeltaG (Kcal/Mol) were predicted with using the method of Cagiada *et al* 2025[17].

scFvright website

scFvright predicts scFv solubility based on the formula: percent solubility = -4.6237 × net charge + 8.2469, derived from **Figure 1f**. The operation can be performed on Fv protein sequences isolated from larger Fab or full-length antibody sequences, which uses AbRSA to isolate variable domain sequences, or scFv protein sequences including linkers and fusion tags. The website is available at https://scfvright.essex.ac.uk.

CDR and framework charge analysis

CDR and framework sequences were isolated using AbRSA[62] from 68,551 non-identical variable domain sequences taken from the Patent & Literature Antibody Database (PLAbDab)[2], 8993 antibody-coding memory B cells[63] and 19303 naïve B cells[64]. Chothia numbering[65] was used to determine CDR-framework boundaries throughout using AbRSA[62], but IMGT and Kabat procedures were also applied to memory B cell data to ensure observed charge distribution differences were not an artefact of a particular numbering scheme. Mean CDR and framework charge for all three numbering schemes are presented in **Supplementary table S2**.

scFv interdomain linker charge analysis

The NCBI databases were searched using the term 'scFv' yielding 9688 sequences. Those with a pair of variable domains and a linker length longer than six amino acids, 1840 sequences, were accepted for analysis. Linker sequences were extracted with AbRSA[62] and checked to ensure no framework residues were incorporated at N and C termini. Charge at neutral pH was calculated as above.

Human redesign of scFv to maximise net charge

Sequences of 3B5H10 and MS785 derived scFv targeting polyQ, SOD1 and FUS/TLS were used to predict structural models with AlphaFold3. They were redesigned by incorporating positive to negative charge swap mutations at sites where lysine and arginine residues were clearly solvent exposed and deemed to not make side chain interactions with any other part of the molecule. 3B5H10 scFv mutations VL K20D, VL K42D, VH K13D, VH K19D, VH K23E and MS785 scFv mutations VL K18D, VL K45D, VH K15D (Chothia numbering).

High throughput *in silico* scFv intrabody construction

The PLAbDab was filtered to remove antibodies targeting non protein targets, bacterial, viral, extracellular, membrane proteins, bi- or tri-specific antibodies, those with no defined target, unusually long CDR loops or where AbRSA[62] failed to assign three CDRs and four framework regions per variable domain. Remaining sequences were then sifted for CDR loop sequence redundancy with the lowest charge paired variable domain chosen over other isoforms. This yielded 672 sequences. Net charge was calculated as above along with predicted solubility of scFv including triple FLAG tag, HA tag and $(G_4D)_4$ linkers. scFv protein structures were predicted with AlphaFold3[66] on an Nvidia A100 40GB GPU machine running

CUDA 12.6. Inverse folding was performed with ProteinMPNN$_{SOL}$[24,26] using 0.20 Å and 0.02 Å models directed to design 100 sequences while fixing amino acids in VL and VH domains corresponding to CDR1, framework region 2, CDR2 and CDR3 extended by three residues at both N and C termini defined using AbRSA[62] and corresponding to Chothia numbering VL24-56, VL86-100, linker, VH26-56, VH92-105. Inverse folded outputs were then sorted based on predicted solubility calculated using the formula % solubility = -4.6237 × net charge + 8.2469 in combination with ProteinMPNN rank (described in **Figure 5a**).

Analysis of ProteinMPNN$_{SOL}$ outputs

To generalise the effect of inverse folding using ProteinMPNN$_{sol}$ on this large amount of structurally similar proteins all sequences were aligned, substitution mutations were analysed and compared between outputs using 0.20 Å and 0.02 Å models. Six hundred and seventy-seven variable domains including linkers were aligned with Clustal Omega with Pearson/FASTA output to create a 294-character alignment. One hundred ProteinMPNN$_{sol}$ inverse folded output sequences for each of the 672 parent scFv were then padded using the aligned parent as a template. This allowed quantification of the number of positions conserved between ProteinMPNN$_{sol}$ output sequences in comparison with each parent scFv. Consensus sequences for 673 parent scFv and 672,000 inverse folded sequences, using 020 and 002 models, were calculated with Jalview.

Data Availability

All scFv intrabody sequences described can be found in **Supplementary Files 1 – 6**.

Acknowledgments

We thank Prof. Yang Cao (Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610065, People's Republic of China) for the use of AbRSA.

Author Contributions (CRediT Classification)

C.M.O'S, Formal Analysis, Investigation, Visualisation, Writing – Original Draft Preparation.
R.S, Investigation, Writing – Review & Editing.
K.A, Investigation, Writing – Review & Editing.
S.N, Software, Writing – Review & Editing.
J.P, Resources, Writing – Review & Editing.
L.C.S, Supervision, Writing – Review & Editing.
G.N.B, Supervision, Writing – Review & Editing.
F.E.B, Resources, Funding Acquisition.
J.S.T, Resources, Funding Acquisition, Writing – Review & Editing.
D.A.B, Resources.
T.F, Resources.
H.I, Resources, Funding Acquisition.
N.R.C, Resources, Funding Acquisition, Writing – Review & Editing.
S.E, Resources, Funding Acquisition, Writing – Review & Editing.

G.S.A.W, Conceptualisation, Formal Analysis, Funding Acquisition, Investigation, Visualisation, Methodology, Project Administration, Software, Supervision, Writing – Original Draft Preparation.

Funding

J.P. National Research Council of Thailand (NRCT) (N42A670122).
J.S.T. National Institutes of Health (U24 NS109113, U24 NS119916).
H.I. JSPS (25H00971), JST (JPMJMS2022-18), AMED (JP21gm5010001, 22gm0010009s0101, 23gm0010009s0102 and 23gm1710001s0102).
S.E. Israel Science Foundation (grant no. 221/22).
G.S.A.W. Royal Society (RGS\R1\231005), Academy of Medical Sciences (SBF008\1028), Motor Neurone Disease Association (Wright/Oct18/969-799, Wright/Apr25/2472-791).

Competing Interest

N.R.C. is Chief Science Officer of ProMIS Neurosciences, and a Professor Emeritus at the University of British Columbia.

Figures

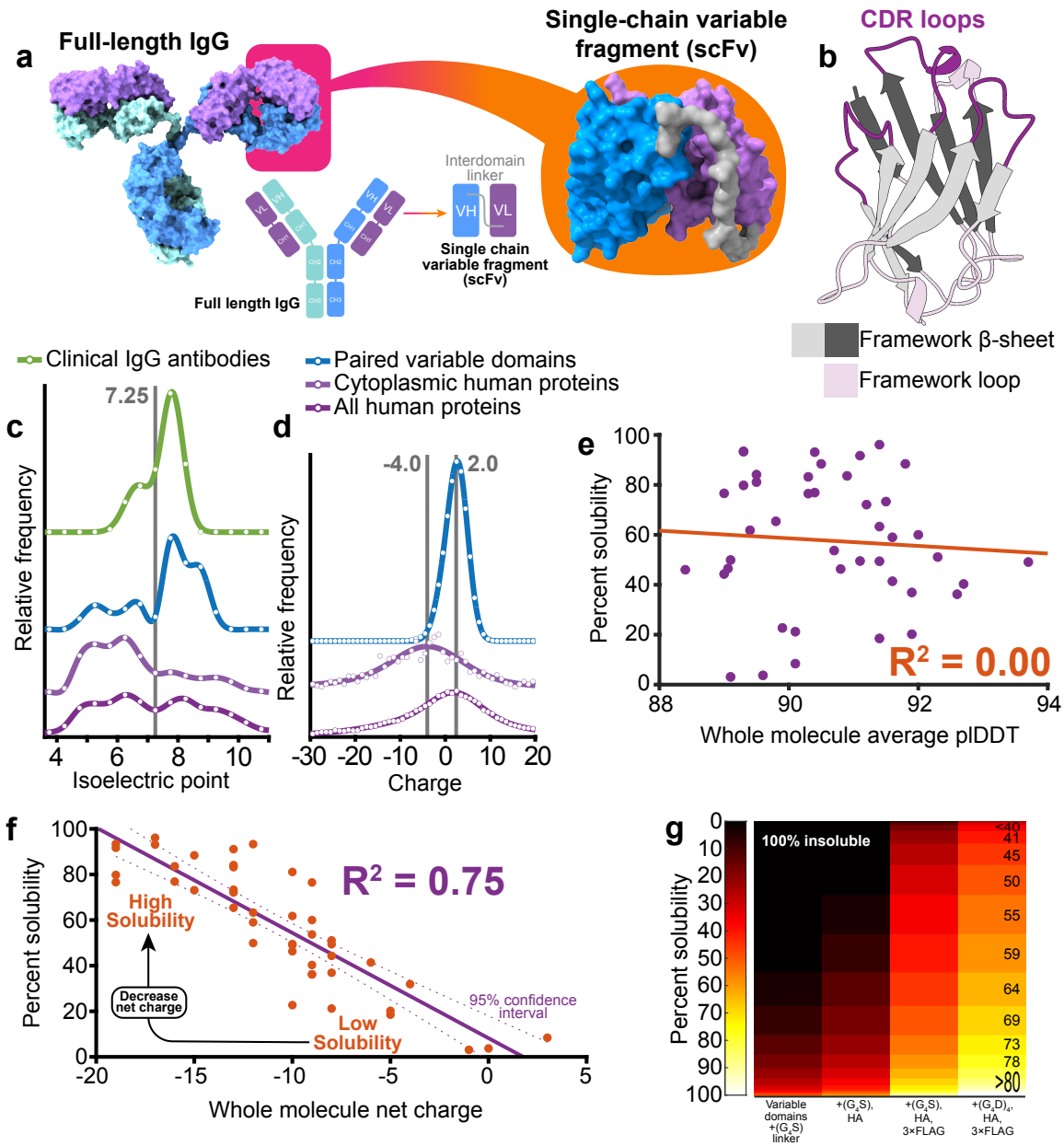

Figure 1. Whole molecule net charge is a determinant of scFv intrabody solubility. **a.** Structural and schematic representations of antibody to scFv reformatting with variable domains (VL and VH) joined by a peptide linker (grey). **b.** Cartoon representation of an antibody variable domain showing CDR and framework regions. **c.** Isoelectric point of full-length clinical IgG antibodies, paired variable domains, human cytoplasmic proteins and the complete human proteome. A characteristic minimum at physiological pH is highlighted in grey. Extracellular proteins predominate in the 7.5-11 region while intracellular proteins are found 4-7.5. **d.** Paired variable domains have a different charge profile in comparison with cytoplasmic proteins. Mode 2.0 and -4.0 (highlighted grey), mean 1.8±2.6 and -8.7±28.3 respectively, mean ± one standard deviation **e.** Average pLDDT scores calculated from AlphaFold3 structural models have no correlation with scFv intrabody solubility. **f.** scFv solubility correlates with net charge. Data points are mean solubility values n≥2, for n=45

scFv intrabodies, P <0.0001 (F-test) with 95% confidence level indicated. **g.** Intracellular solubility prediction for 68,551 scFv with various modifications to the interdomain linker, N and C-termini.

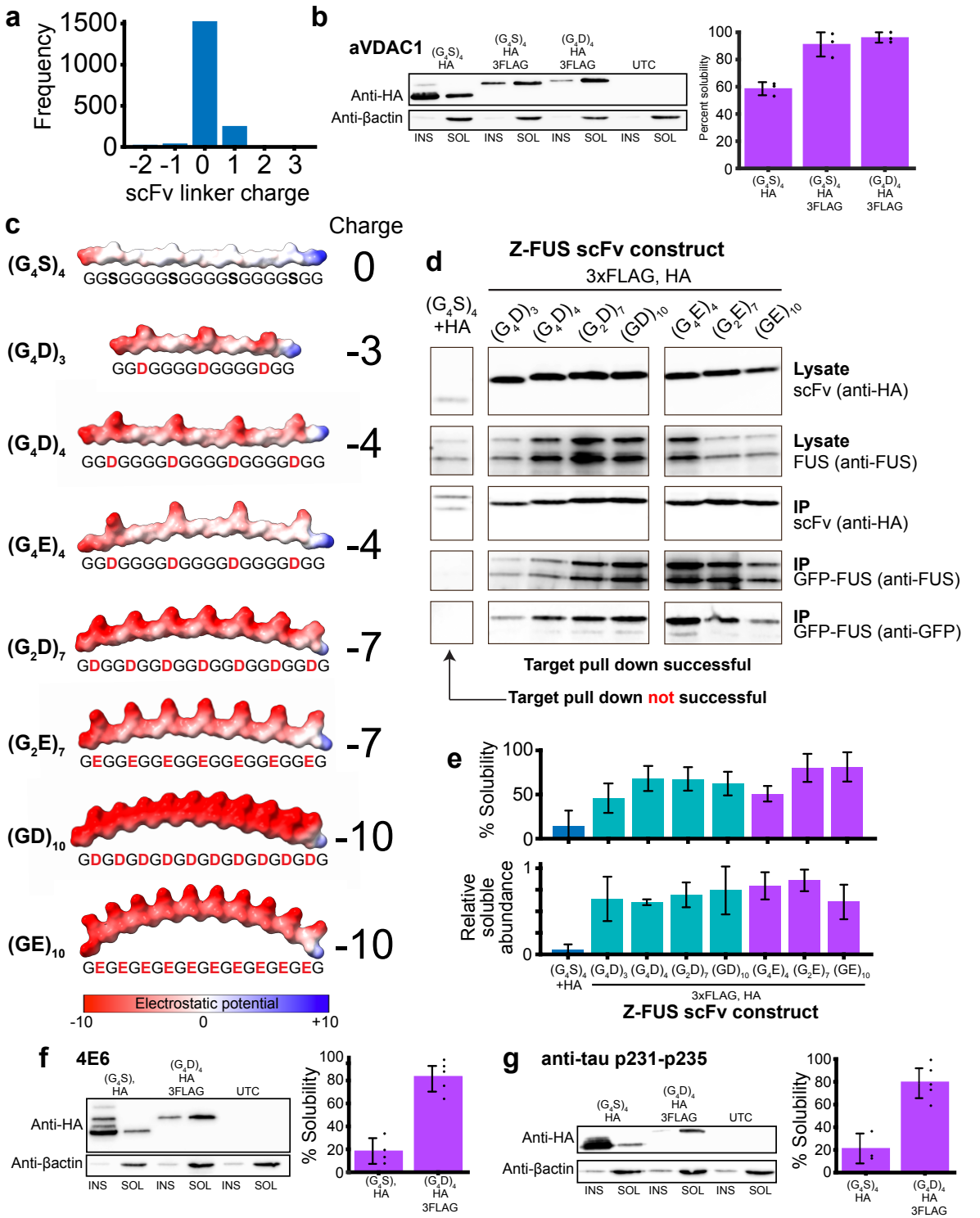

Figure 2. Charged interdomain linkers improve scFv intrabody solubility. a. Frequency bar chart showing the charge of 1840 scFv interdomain linker sequences in the NCBI database. **b.** Increased solubility of an anti-VDAC1 scFv with charged interdomain linker. **c.** Coulombic electrostatic potential colouring of negatively charged interdomain linker structural representations, n=3, $(G_4S)_4$/HA vs $(G_4S)_4$/HA/3FLAG P=0.001, $(G_4S)_4$/HA/3FLAG vs $(G_4D)_4$/HA/3FLAG P=0.43. **d.** Co-immunoprecipitation of GFP-FUS/TLS protein by Z-FUS-derived scFv intrabodies with varying charge inter-domain linkers bound by immobilised anti-HA beads and visualised by SDS-PAGE/western blot. Antibodies used to probe western

membranes are stated to the right. **e.** Percent solubility and abundance in the soluble fraction of scFv intrabodies shown in **d**, n=3, $(G_4S)_4$/HA vs $(G_4D)_4$/HA/3FLAG solubility P=0.0007, abundance P=0.0002. Blue – serine linker, green – aspartate linkers, purple – glutamate linkers. **f** and **g**. High solubility engineered into scFv intrabodies, derived from 4E6 (P=0.0001) and anti-p231-235 (P=0.0016) antibodies, targeting tau with combined N- and C-terminal tags and linker modifications, $n \geq 3$. All error bars represent one standard deviation.

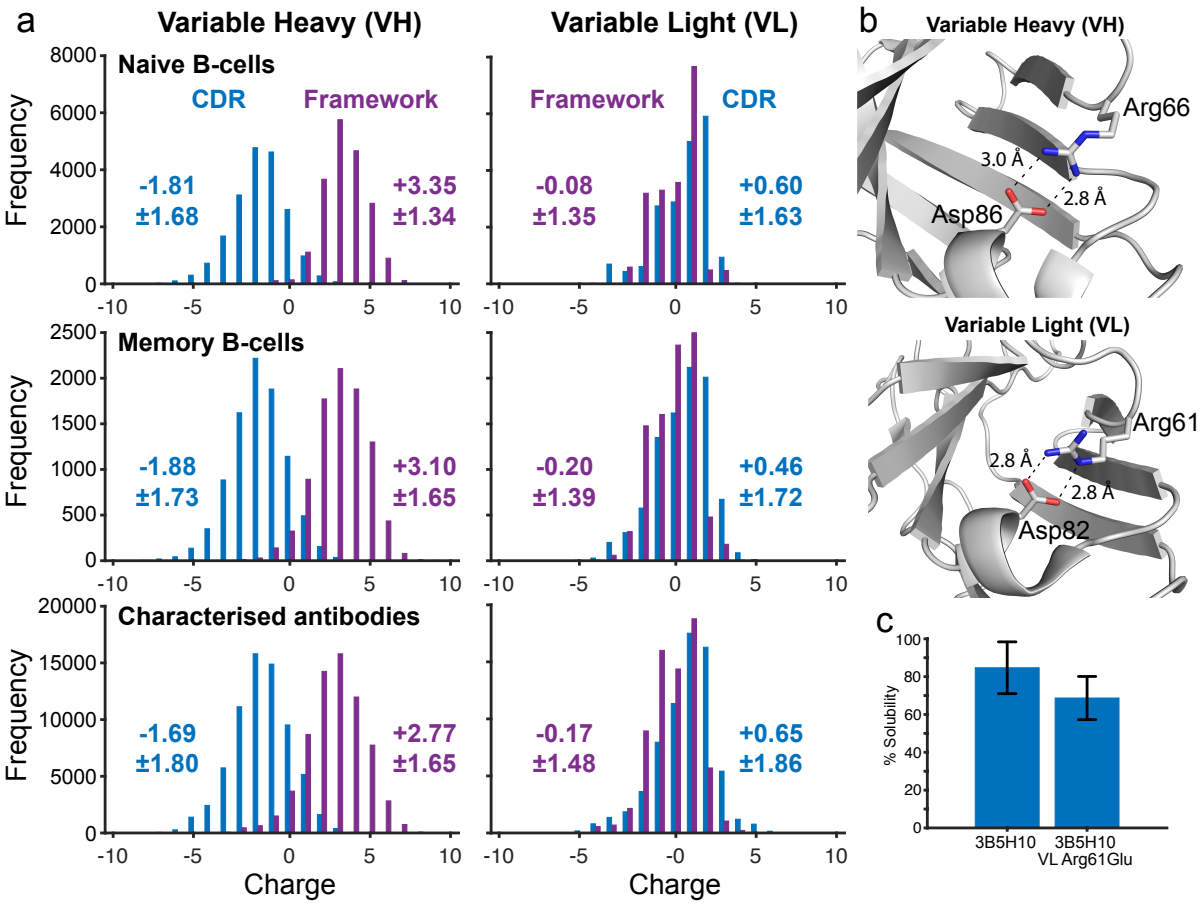

Figure 3. Antibody variable domain CDR loop and framework region charge. a. The majority of variable heavy domain CDR loops are negatively charged and normally distributed around -1.81 to -1.69 dependent on maturation state. VH framework regions are predominantly positively charged and distributed around +2.77 to +3.10. This effect is observed irrespective of variable domain numbering scheme (**Supplementary table S2**). Conversely, light chain CDR and framework charge display overlapping distributions. A slight skew to positive and negative respectively is conserved between Chothia and Kabat but not IMGT numbering schemes (**Supplementary table S2**). **b.** Salt-bridge bonding between β-strands D and F of the 3B5H10 antibody variable domains. Bonding distances annotated. **c.** Charge substitution mutation VL Arg61Glu reduces 3B5H10-derived scFv intrabody solubility by breaking the salt-bridge in **b**. Error bars represent one standard deviation, n≥5, P=0.058.

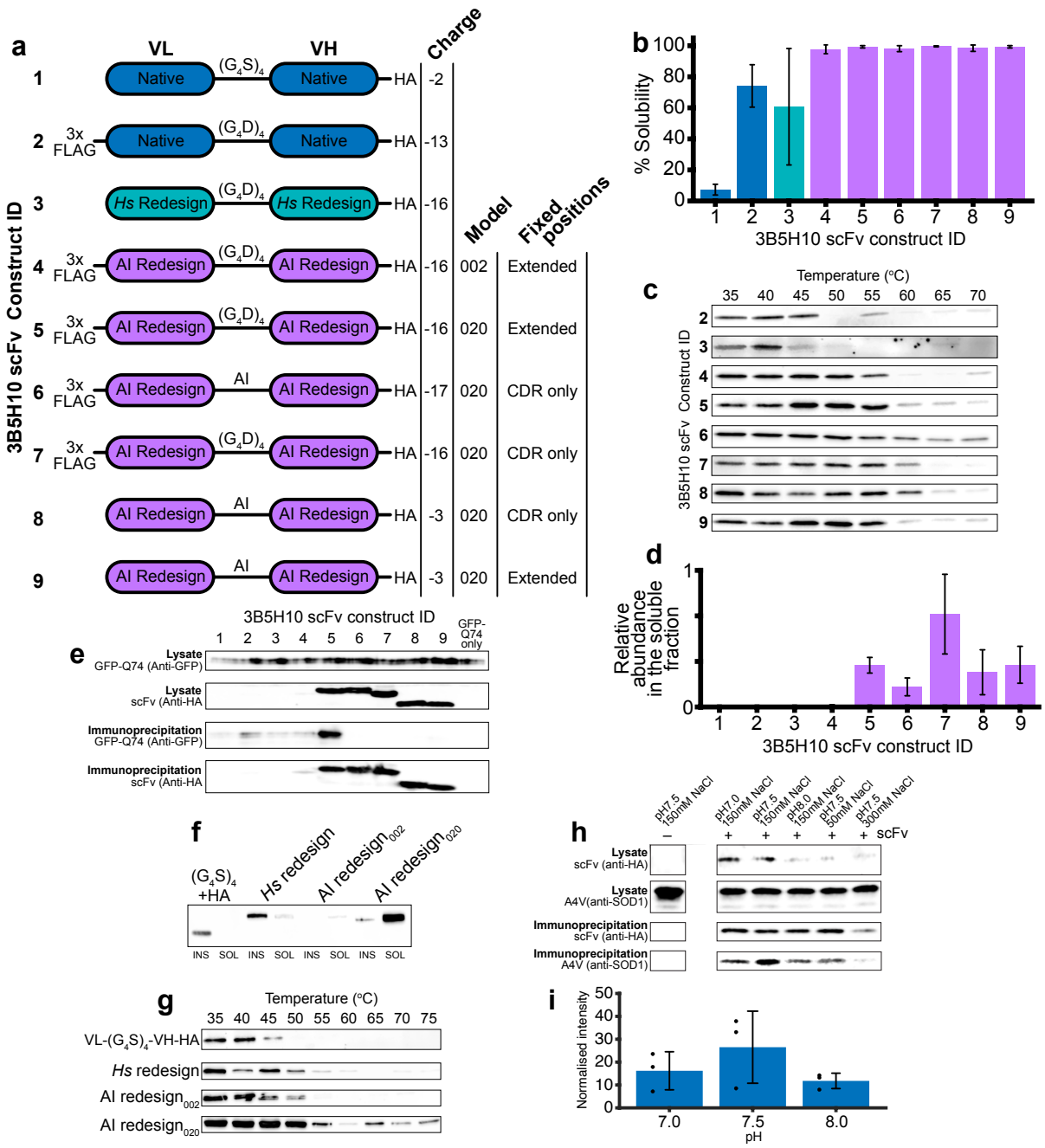

Figure 4. Inverse folding improves scFv intrabody solubility and thermal stability. a. 3B5H10-derived scFv intrabodies created using native, human redesigned (VL K20D, VL K42D, VH K13D, VH K19D, VH K23E), and ProteinMPNN$_{SOL}$ inverse folded variable domains with stated tags, linkers and model noise. Fixed residue positions CDR only: VL 24-34, 50-56, 89-97 and VH 25-32, 52-56, 95-102. Extended: VL 24-36, 47-56, 86-97 and VH 26-59, 91-105. **b.** Percent of 3B5H10-derived scFv found in the soluble cell fraction. n≥3. **c.** In-cell thermal stability assay of 3B5H10-derived scFv intrabodies. Stability for construct 1 could not be measured. n=2. **d.** Abundance of 3B5H10-derived scFv in the soluble fraction. n=3. **e.** Co-immunoprecipitation of GFP-polyQ74 by 3B5H10-derived scFv intrabodies bound by immobilised anti-HA beads and visualised by SDS-PAGE/western blot. Antibodies used to probe western membranes are stated to the left. **f.** SDS-PAGE/western blot showing the solubility and **g.** Intracellular thermal stability of MS785-derived scFv intrabodies. **h.** Co-

immunoprecipitation of A4V SOD1 by MS785-derived scFv (AI redesign$_{020}$) intrabody bound by immobilised anti-HA beads and visualised by SDS-PAGE/western blot. **i.** Quantification of A4V SOD1 immunoprecipitation by MS785 showing robust binding across pH 7.0 to 8.0 with a non-significant preference for physiological pH, n=3, P=0.284 (one way ANOVA). Antibodies used to probe western membranes are stated to the left. Chothia numbering throughout. Error bars represent one standard deviation. *Hs* – human, AI – ProteinMPNN$_{SOL}$ inverse folding.

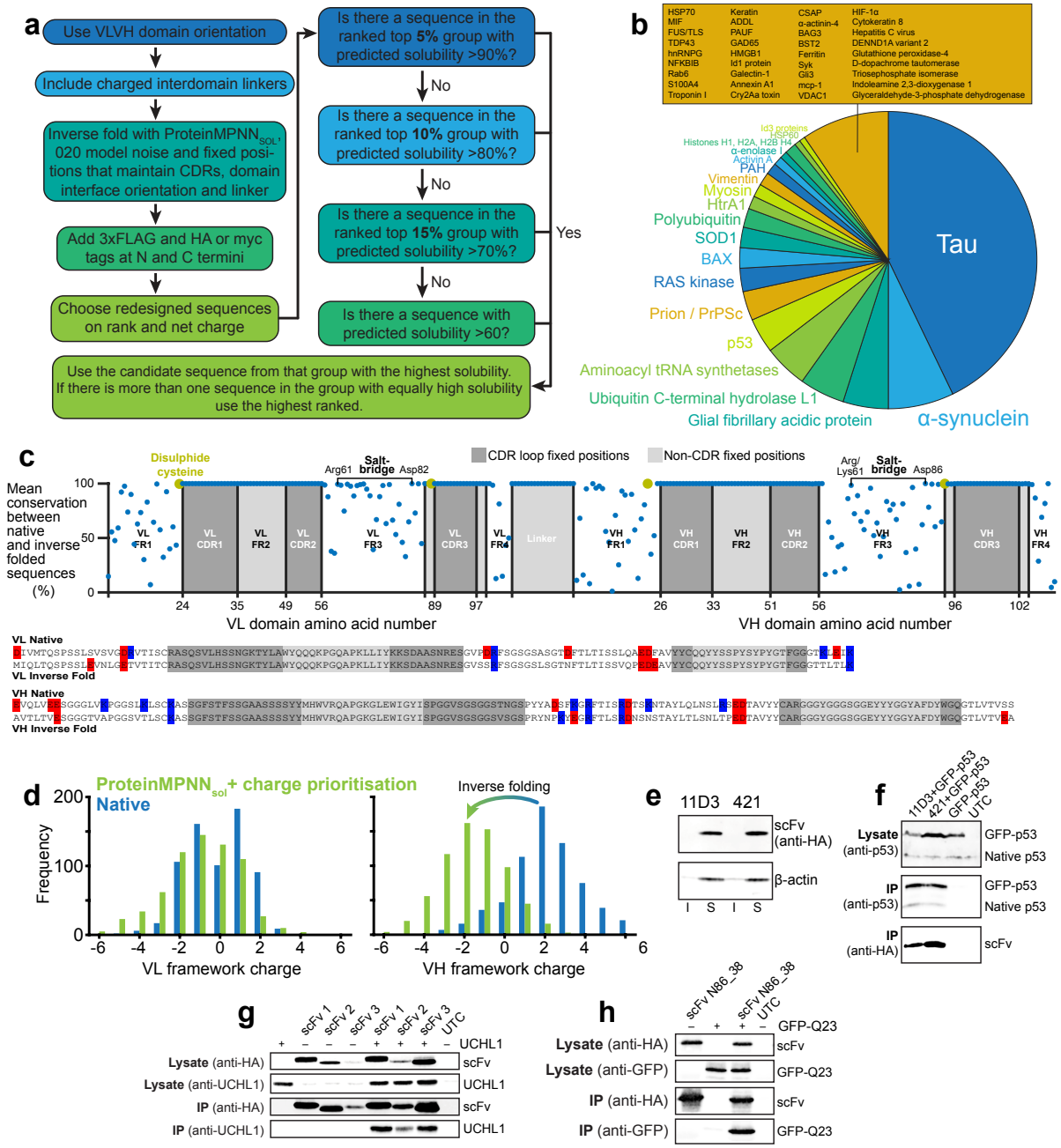

Figure 5. Reliable *in silico* construction of scFv intrabodies. a. Design rules for reformatting an antibody as an optimised scFv intrabody. **b.** Intracellular target specificity of 672 scFv constructed using the principles in **a. c.** Mean amino acid conservation for native to inverse folded scFv intrabody and charged residues in non-fixed positions within variable domain consensus sequences. **d.** Charge frequency histograms for VL and VH framework sites before and after inverse folding. **e.** Anti-p53 scFv 11D3 and 421 constructed using rules in **a** are highly soluble. **f.** Co-immunoprecipitation of p53 by anti-p53 scFv intrabodies. **g.** Co-immunoprecipitation of UCHL1 by anti-UCHL1 scFv intrabodies. **h.** Co-immunoprecipitation of GFP by anti-GFP scFv intrabody derived from Neuromabs antibody N86_38. All immunoprecipitation visualised by SDS-PAGE/western following immobilisation with anti-HA beads. Antibodies used to probe western membranes are stated to the left. IP – Immunoprecipitation. Chothia numbering throughout.

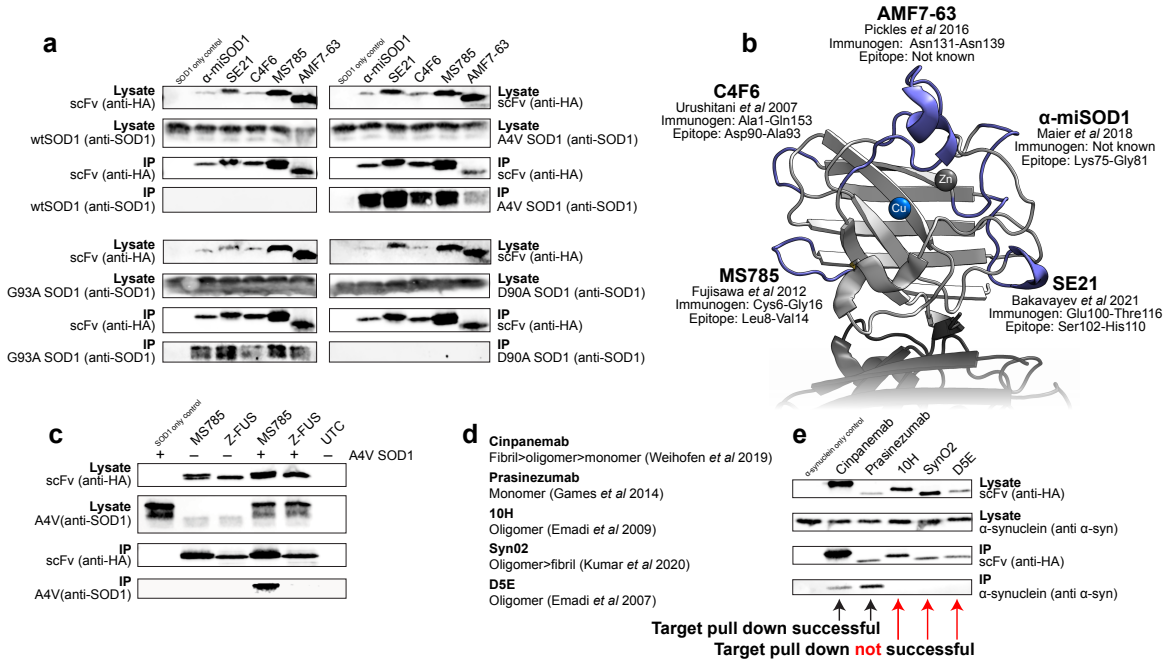

Figure 6. State specific intrabody interactions with SOD1 and α-synuclein. a. Assessment of the interaction between scFv intrabodies derived from anti-SOD1 conformation specific antibodies with wild-type, A4V, G93A and D90A SOD1 by co-immunoprecipitation on anti-HA beads and visualised by SDS-PAGE/western blot. Antibodies used to probe western membranes are stated on the left and right. **b.** SOD1 conformation specific intrabodies target five distinct epitopes in the SOD1 monomer here shown in the fully metalated disulphide intact form. **c.** Immunoprecipitation of A4V SOD1 by MS785-derived scFv but not Z-FUS-derived scFv. **d.** Oligomeric state specificity of five anti-α-synuclein antibodies. **e.** Co-immunoprecipitation of α-synuclein by scFv intrabodies derived from antibodies in **d** on anti-HA beads and visualised by SDS-PAGE/western blot. Antibodies used to probe western membranes are stated. IP – Immunoprecipitation.

Supplementary Information

Reliable repurposing of antibody interactome inside the cell

Caitlin M. O'Shea[1], Rushba Shahzad[1], Kimia Aghasoleimani[1], Stuart Newman[1], Jiraporn Panmanee[2], Leonard C. Schalkwyk[1], Greg N. Brooke[1], Fiona E. Benson[4], James S. Trimmer[5], Daryl A. Bosco[6], Takao Fujisawa[7,8], Hidenori Ichijo[7,8], Neil R. Cashman[9], Stanislav Engel[10], Gareth S. A. Wright[1]*

Supplementary Figures S1 – S3

Figure S1. Physicochemical properties of complete and cytoplasmic human proteomes in comparison with antibody variable domains.

Figure S2. Comparison of the relationship between experimentally defined solubility for forty-five scFv intrabodies and predicted solubility by nine solubility predictors

Figure S3. Impact of variable heavy domain charge on scFv intrabody solubility.

Supplementary Tables S1 & S2

Supplementary table S1. Coefficient of determination (r^2) for predicted physicochemical and biophysical properties of scFv molecules broken down into regions of the molecule with experimentally determined solubility.

Supplementary table S2. Comparison of variable domain CDR and framework mean charge of 8993 antibody-coding memory B cells[63] using IMGT, Chothia and Kabat amino acid numbering schemes.

Supplementary Files

Supplementary File 1 – Sequence, description, predicted physicochemical characteristics and solubility of forty-five scFv intrabodies with various tags, linkers and domain orientations.

Supplementary File 2 – 3B5H10-derived scFv intrabodies including sequence, construct ID, fusion tags, linker, net charge and ProteinMPNN$_{SOL}$ inverse folding settings**.**

Supplementary File 3 – Sequence and description of 672 non-redundant scFv intrabodies targeting cytoplasmic proteins with native and inverse folded sequences.

Supplementary File 4 – Anti-p53 scFv intrabody sequences with reference.

Supplementary File 5 – Anti-SOD1 conformational specific scFv intrabodies with sequence and references.

Supplementary File 6 – Anti-α-synuclein state specific scFv intrabodies with sequences and references.

Supplementary Figures

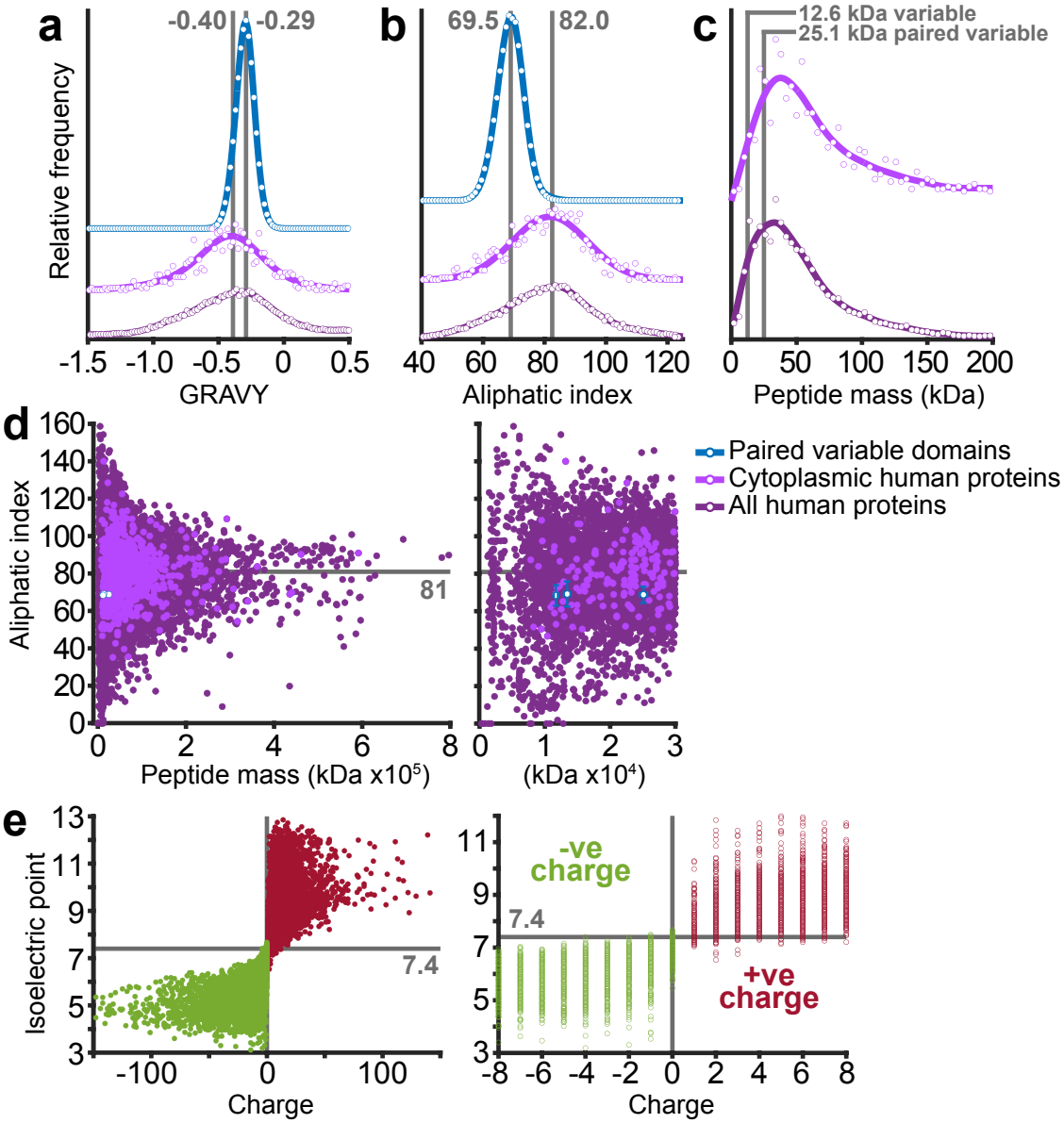

Figure S1. Physicochemical properties of complete and cytoplasmic human proteomes in comparison with antibody variable domains. a. Fv grand average of hydropathy (GRAVY) and **b.** Aliphatic index. Modes distinguishing Fv and cytoplasmic proteins derived from the peak of the fitted curve are shown in grey. **c.** Molecular mass of single and paired variable domains in comparison with total and cytoplasmic human proteomes. **d.** Comparison of peptide mass and aliphatic index showing single variable and paired domains occupy a region within the main body of total and cytoplasmic human proteins synonymous with folded and stable proteins. Mean aliphatic index for total (80.9 ± 18.1) and cytoplasmic (81.2 ± 13.6), highlighted with grey line, can be compared with single VH (69.2 ± 6.7), VL (68.4 ± 5.5) and paired VLVH (68.8 ± 4.4). Error bars show one standard deviation. **e.** Relationship between net charge and isoelectric point across the human proteome, R^2 0.42. Data in **a**, **b** and **c** are normalised to area under the curve.

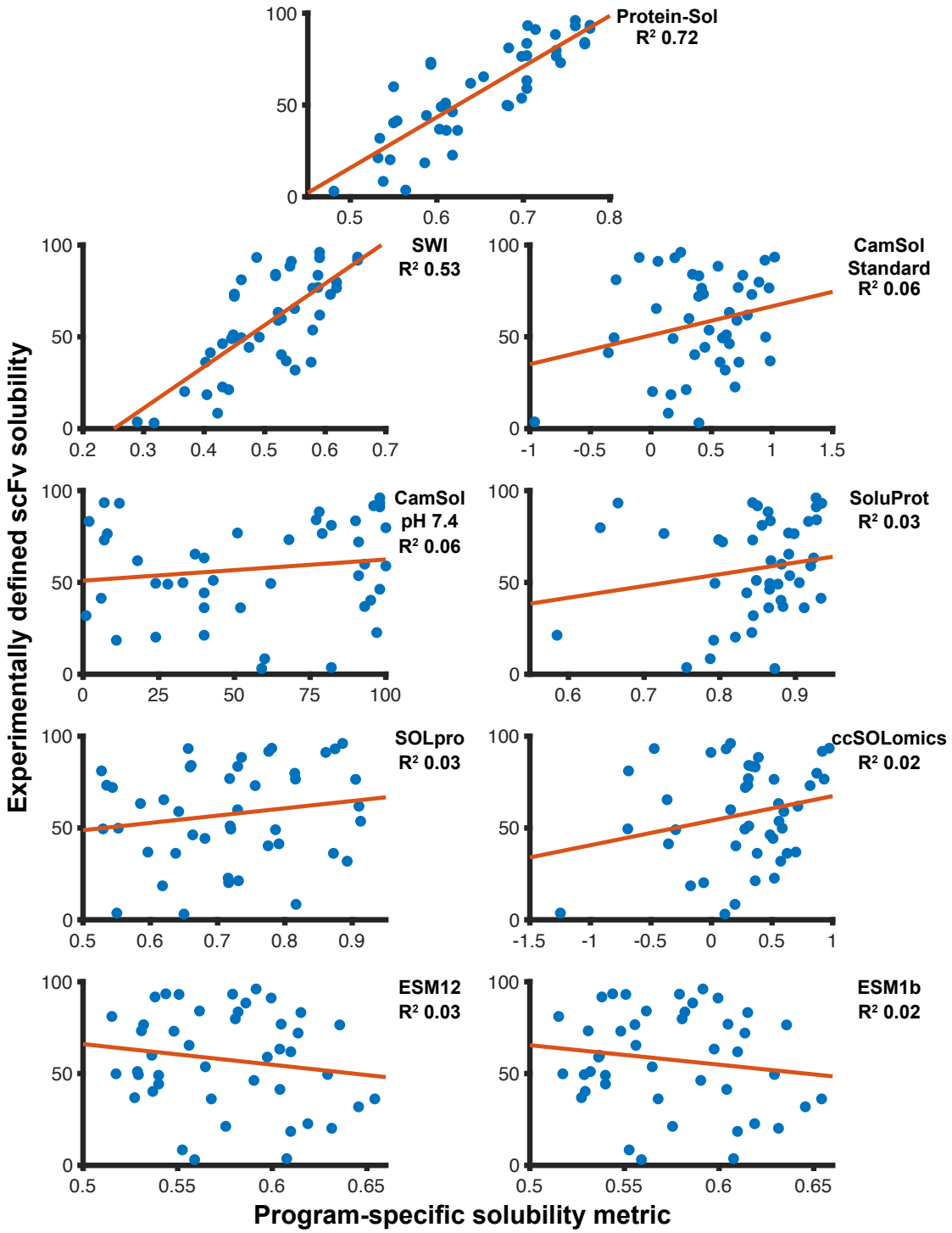

Figure S2. Comparison of the relationship between experimentally defined solubility for forty-five scFv intrabodies and predicted solubility by nine solubility predictors[16,67–72].

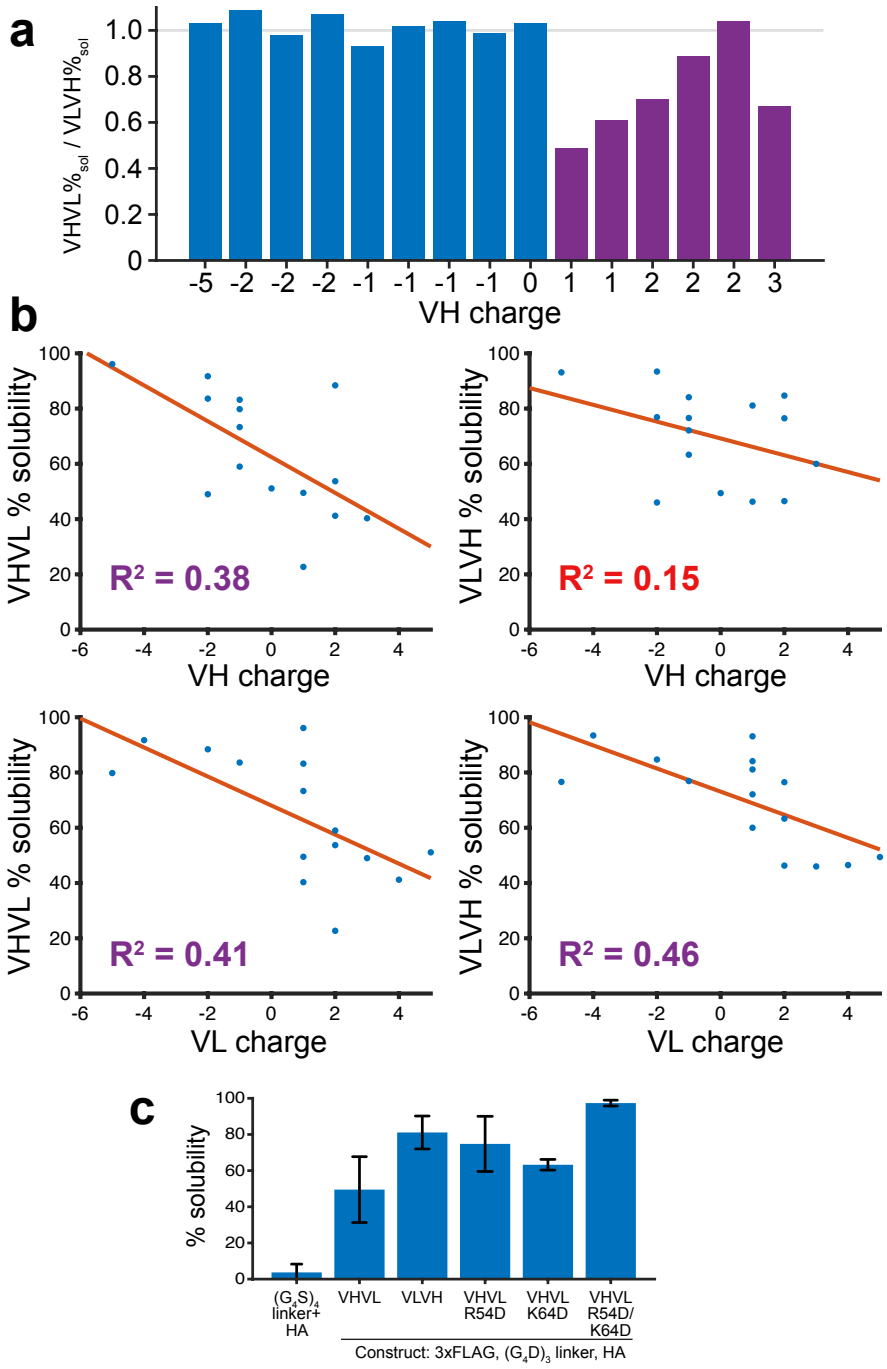

Figure S3. Impact of variable heavy domain charge on scFv intrabody solubility. a. Ratio of VHVL/VLVH solubility for fifteen intrabodies with varying VH charge. **b.** Relationship between scFv intrabody variable domain charge and solubility of VHVL and VLVH constructs with coefficient of determination. **c.** Solubility of AMF7-63-derived anti-SOD1 scFv intrabodies and mutants. VHVL has lower solubility than the VLVH orientation, but this can be overcome by inclusion of charge swap mutations in the VH domain. n≥2 $(G_4S)_4$/linker/HA vs VHVL R54D/K64D P=0.000056. Chothia numbering.

Supplementary tables

Region	Charge at pH 7.5 r^2 with solubility	Charge at pH 5.5 r^2 with solubility	Isoelectric point r^2 with solubility	Hydropathicity r^2 with solubility	Aliphatic index r^2 with solubility
Fv (+ tags & linker)	0.74	0.67	0.38	0.14	0.00
VH	0.09	0.07	0.08	0.00	0.01
VL	0.13	0.08	0.11	0.03	0.06
All CDR	0.14	0.10	0.09	0.03	0.01
VH CDR	0.04	0.03	0.07	0.03	0.00
VH CDR1	0.12	0.12	0.08	0.08	0.12
VH CDR2	0.05	0.05	0.08	0.01	0.05
VH CDR3	0.01	0.02	0.03	0.00	0.00
VL CDR	0.13	0.08	0.13	0.00	0.00
VL CDR1	0.09	0.08	0.15	0.00	0.00
VL CDR2	0.01	0.01	0.00	0.00	0.06
VL CDR3	0.13	0.03	0.05	0.01	0.05
Framework Fv	0.05	0.02	0.03	0.03	0.13
Framework VH	0.02	0.01	0.03	0.00	0.00
Framework VL	0.02	0.01	0.01	0.07	0.00

Property	r^2 with solubility
Mean pLDDT	0.00
Mean pLDDT*	0.07
Mean pLDDT#	0.05
DeltaG IF*	0.15
DeltaG IF#	0.07
DeltaG*	0.15
DeltaG#	0.07

*Whole molecule with aspartate linkers.
#Whole molecule with serine linkers.

Supplementary table S1. Coefficient of determination (r^2) for predicted physicochemical and biophysical properties of scFv molecules broken down into regions of the molecule with experimentally determined solubility.

					Variable heavy				Variable light				Framework		
Numbering scheme	**Fv charge**	**VH charge**	**VL charge**	**All CDR**	**CDR 1 2 3**	**CDR 1**	**CDR 2**	**CDR 3**	**CDR 1 2 3**	**CDR 1**	**CDR 2**	**CDR 3**	**Fv**	**VH**	**VL**
IMGT	1.49	1.24	0.25	-1.57	-0.92	-0.16	-0.23	-0.52	-0.65	-0.03	-0.35	-0.27	3.05	2.15	0.90
Chothia	1.49	1.22	0.26	-1.41	-1.88	-0.14	-0.40	-1.34	0.46	0.45	0.29	-0.27	2.90	3.10	-0.20
Kabat	1.49	1.17	0.32	-0.74	-1.20	-0.14	0.28	-1.34	0.46	0.45	0.29	-0.27	2.22	2.37	-0.14

Supplementary table S2. Comparison of variable domain CDR and framework mean charge of 8993 antibody-coding memory B cells[63] using IMGT, Chothia and Kabat amino acid numbering schemes.

References

1. Kovaltsuk, A. *et al.* Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. *J. Immunol.* **201**, 2502–2509 (2018).
2. Abanades, B. *et al.* The Patent and Literature Antibody Database (PLAbDab): an evolving reference set of functionally diverse, literature-annotated antibody sequences and structures. *Nucleic Acids Res.* **52**, D545–D551 (2024).
3. Kim, D.-S. *et al.* Engineering IgG antibodies for intracellular targeting and drug delivery. *J. Controlled Release* **382**, 113727 (2025).
4. Herrmann, A. *et al.* An effective cell-penetrating antibody delivery platform. *JCI Insight* **4**, (2019).
5. Bird, R. E. *et al.* Single-chain antigen-binding proteins. *Science* **242**, 423–426 (1988).
6. Huston, J. S. *et al.* Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proc. Natl. Acad. Sci.* **85**, 5879–5883 (1988).
7. Auf der Maur, A., Tissot, K. & Barberis, A. Antigen-independent selection of intracellular stable antibody frameworks. *Methods* **34**, 215–224 (2004).
8. Jung, S. & Plückthun, A. Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng. Des. Sel.* **10**, 959–966 (1997).
9. Knappik, A. & Plückthun, A. Engineered turns of a recombinant antibody improve its in vivo folding. *Protein Eng.* **8**, 81–89 (1995).
10. Proba, K., Wörn, A., Honegger, A. & Plückthun, A. Antibody scFv fragments without disulfide bonds, made by molecular evolution. *J. Mol. Biol.* **275**, 245–253 (1998).
11. Rego, M., Houston, D. W., Fan, M., Murray, K. D. & Trimmer, J. S. Open-source antibodies as a path to enhanced research reproducibility and transparency. *New Biotechnol.* **87**, 121–129 (2025).
12. Oreste, U., Ametrano, A. & Coscia, M. R. On Origin and Evolution of the Antibody Molecule. *Biology* **10**, 140 (2021).
13. Otey, C. A., Dixon, R., Stack, C. & Goicoechea, S. M. Cytoplasmic Ig-domain proteins: Cytoskeletal regulators with a role in human disease. *Cell Motil.* **66**, 618–634 (2009).
14. Richardson, J. S., Richardson, D. C., Thomas, K. A., Silverton, E. W. & Davies, D. R. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. *J. Mol. Biol.* **102**, 221–235 (1976).
15. Wright, G. S. A. Bacterial Evolutionary Precursors of Eukaryotic Copper–Zinc Superoxide Dismutases. *Mol. Biol. Evol.* **38**, 3789–3803 (2021).
16. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098–3100 (2017).
17. Cagiada, M., Ovchinnikov, S. & Lindorff-Larsen, K. Predicting absolute protein folding stability using generative models. *Protein Sci.* **34**, e5233 (2025).
18. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
19. Akdel, M. *et al.* A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* **29**, 1056–1067 (2022).

20. Kabayama, H. *et al.* An ultra-stable cytoplasmic antibody engineered for in vivo applications. *Nat. Commun.* **11**, 336 (2020).
21. Dudgeon, K. *et al.* General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc. Natl. Acad. Sci.* **109**, 10879–10884 (2012).
22. Peters-Libeu, C. *et al.* Crystallization and diffraction properties of the Fab fragment of 3B5H10, an antibody specific for disease-causing polyglutamine stretches. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **61**, 1065–1068 (2005).
23. Glockshuber, R., Malia, M., Pfitzinger, I. & Plückthun, A. A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry* **29**, 1362–1367 (1990).
24. Goverde, C. A. *et al.* Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
25. Dahiyat, B. I. & Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **278**, 82–87 (1997).
26. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
27. Sumida, K. H. *et al.* Improving Protein Expression, Stability, and Function with ProteinMPNN. *J. Am. Chem. Soc.* **146**, 2054–2061 (2024).
28. Dauparas, J. *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **22**, 717–723 (2025).
29. Hsu, H.-J. *et al.* Antibody Variable Domain Interface and Framework Sequence Requirements for Stability and Function by High-Throughput Experiments. *Structure* **22**, 22–34 (2014).
30. Akbar, R. *et al.* A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell Rep.* **34**, (2021).
31. Fujisawa, T. *et al.* A novel monoclonal antibody reveals a conformational alteration shared by amyotrophic lateral sclerosis-linked SOD1 mutants. *Ann. Neurol.* **72**, 739–749 (2012).
32. Caron de Fromentel, C. *et al.* Restoration of transcriptional activity of p53 mutants in human tumour cells by intracellular expression of anti-p53 single chain Fv fragments. *Oncogene* **18**, 551–557 (1999).
33. Urushitani, M., Ezzi, S. A. & Julien, J.-P. Therapeutic effects of immunization with mutant superoxide dismutase in mice models of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2495–2500 (2007).
34. Ayers, J. I. *et al.* Conformational specificity of the C4F6 SOD1 antibody; low frequency of reactivity in sporadic ALS cases. *Acta Neuropathol. Commun.* **2**, 55 (2014).
35. Pickles, S. *et al.* ALS-linked misfolded SOD1 species have divergent impacts on mitochondria. *Acta Neuropathol. Commun.* **4**, 43 (2016).
36. Maier, M. *et al.* A human-derived antibody targets misfolded SOD1 and ameliorates motor symptoms in mouse models of amyotrophic lateral sclerosis. *Sci. Transl. Med.* **10**, eaah3924 (2018).
37. Bakavayev, S. *et al.* Exposure of β6/β7-Loop in Zn/Cu Superoxide Dismutase (SOD1) Is Coupled to Metal Loss and Is Transiently Reversible During Misfolding. *ACS Chem. Neurosci.* **12**, 49–62 (2021).

38. Weihofen, A. *et al.* Development of an aggregate-selective, human-derived α-synuclein antibody BIIB054 that ameliorates disease phenotypes in Parkinson's disease models. *Neurobiol. Dis.* **124**, 276–288 (2019).
39. Games, D. *et al.* Reducing C-Terminal-Truncated Alpha-Synuclein by Immunotherapy Attenuates Neurodegeneration and Propagation in Parkinson's Disease-Like Models. *J. Neurosci.* **34**, 9441–9454 (2014).
40. Emadi, S., Kasturirangan, S., Wang, M. S., Schulz, P. & Sierks, M. R. Detecting Morphologically Distinct Oligomeric Forms of α-Synuclein. *J. Biol. Chem.* **284**, 11048–11058 (2009).
41. Kumar, S. T. *et al.* How specific are the conformation-specific α-synuclein antibodies? Characterization and validation of 16 α-synuclein conformation-specific antibodies using well-characterized preparations of α-synuclein monomers, fibrils and oligomers with distinct structures and morphology. *Neurobiol. Dis.* **146**, 105086 (2020).
42. Emadi, S., Barkhordarian, H., Wang, M. S., Schulz, P. & Sierks, M. R. Isolation of a Human Single Chain Antibody Fragment Against Oligomeric α-Synuclein that Inhibits Aggregation and Prevents α-Synuclein-induced Toxicity. *J. Mol. Biol.* **368**, 1132–1144 (2007).
43. Lang, A. E. *et al.* Trial of Cinpanemab in Early Parkinson's Disease. *N. Engl. J. Med.* **387**, 408–420 (2022).
44. Pagano, G. *et al.* Trial of Prasinezumab in Early-Stage Parkinson's Disease. *N. Engl. J. Med.* **387**, 421–432 (2022).
45. Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N. & Scholtz, J. M. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **102**, 1907–1915 (2012).
46. Kvam, E., Sierks, M. R., Shoemaker, C. B. & Messer, A. Physico-chemical determinants of soluble intrabody expression in mammalian cell cytoplasm. *Protein Eng. Des. Sel.* **23**, 489–498 (2010).
47. Joshi, S. N., Butler, D. C. & Messer, A. Fusion to a highly charged proteasomal retargeting sequence increases soluble cytoplasmic expression and efficacy of diverse anti-synuclein intrabodies. *mAbs* **4**, 686–693 (2012).
48. Michaelis, L. & Pechstein, H. Der isoelektrische Punkt des Caseins. *Biochem. Z.* **47**, 260–268 (1912).
49. Shanehsazzadeh, A. *et al.* Unlocking de novo antibody design with generative artificial intelligence. 2023.01.08.523187 Preprint at https://doi.org/10.1101/2023.01.08.523187 (2023).
50. Bennett, N. R. *et al.* Atomically accurate de novo design of antibodies with RFdiffusion. Preprint at https://doi.org/10.1101/2024.03.14.585103 (2024).
51. Shanker, V. R., Bruun, T. U. J., Hie, B. L. & Kim, P. S. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* **385**, 46–53 (2024).
52. Lum, J. *et al.* Development of a targeted BioPROTAC degrader selective for misfolded SOD1. Preprint at https://doi.org/10.21203/rs.3.rs-6070925/v1 (2025).
53. Benn, J. *et al.* Aggregate-selective removal of pathological tau by clustering-activated degraders. *Science* **385**, 1009–1016 (2024).

54. Gasser, H.-C., Oyarzún, D. A., Alfaro, J. A. & Rajan, A. Tuning ProteinMPNN to reduce protein visibility via MHC Class I through direct preference optimization. *Protein Eng. Des. Sel.* **38**, gzaf003 (2025).
55. Rotunno, M. S. *et al.* Identification of a Misfolded Region in Superoxide Dismutase 1 that is Exposed in Amyotrophic Lateral Sclerosis. *J. Biol. Chem.* **289**, 28527–28538 (2014).
56. Olsen, T. H., Boyles, F. & Deane, C. M. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci.* **31**, 141–146 (2022).
57. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
58. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
59. Kozlowski, L. P. IPC – Isoelectric Point Calculator. *Biol. Direct* **11**, 55 (2016).
60. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
61. Ikai, A. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem. (Tokyo)* **88**, 1895–1898 (1980).
62. Li, L. *et al.* AbRSA: A robust tool for antibody numbering. *Protein Sci.* **28**, 1524–1531 (2019).
63. King, H. W. *et al.* Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci. Immunol.* **6**, eabe6291 (2021).
64. Jaffe, D. B. *et al.* Functional antibodies exhibit light chain coherence. *Nature* **611**, 352–357 (2022).
65. Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917 (1987).
66. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
67. Agostini, F., Cirillo, D., Livi, C. M., Delli Ponti, R. & Tartaglia, G. G. cc SOL omics : a webserver for solubility prediction of endogenous and heterologous expression in Escherichia coli. *Bioinformatics* **30**, 2975–2977 (2014).
68. Thumuluri, V. *et al.* NetSolP: predicting protein solubility in Escherichia coli using language models. *Bioinformatics* **38**, 941–946 (2022).
69. Oeller, M. *et al.* Sequence-based prediction of pH-dependent protein solubility using CamSol. *Brief. Bioinform.* **24**, bbad004 (2023).
70. Magnan, C. N., Randall, A. & Baldi, P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**, 2200–2207 (2009).
71. Bhandari, B. K., Gardner, P. P. & Lim, C. S. Solubility-Weighted Index: fast and accurate prediction of protein solubility. *Bioinformatics* **36**, 4691–4698 (2020).
72. Hon, J. *et al.* SoluProt: prediction of soluble protein expression in Escherichia coli. *Bioinformatics* **37**, 23–28 (2021).