Volume 1 | Number 1 | Jan 2013 | Pages 1–100

PCCP

Physical Chemistry Chemical Physics
www.rsc.org/pccp

ROYAL SOCIETY OF CHEMISTRY

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ARTICLE TYPE

The self assembly of proteins; Probing patchy protein interactions

Susan James, Michelle K. Quinn and Jennifer J. McManus*

5 The ability to control the self-assembly of biological molecules to form defined structures, with a high degree of predictability is a central aim for soft matter science and synthetic biology. Several examples of this are known for synthetic systems, such as anisotropic colloids. However, for biomacromolecules, such as proteins, success has been more limited, since aeolotopic (or anisotropic) interactions between protein molecules are not easily predicted. We have created three double mutants of human γD-crystallin for
10 which the phase diagrams for singly mutated proteins can be used to predict the behavior of the double mutants. These proteins provide a robust mechanism to examine the kinetic and thermodynamic properties of proteins in which competing interactions exist due to the anisotropic or patchy nature of the protein surface.

Introduction

15 While there is increasing awareness about how best to screen solution conditions to increase the probability of protein crystallization, mostly, this is unsuccessful, since protein-protein interactions are strongly determined by the specific chemistry of the surface amino acids. Using a phase diagram approach, three
20 double mutant proteins of human gamma D-crystallin have been created. In this way, the competitive nature of these anisotropic protein interactions can be examined. For these mutants, the behavior of the double mutants can be qualitatively predicted by the single mutant variants, and in one case, two crystals for the
25 same protein were formed, each with opposite temperature dependence of the solubility lines; one melts when it's heated and one melts when it's cooled.

Anisotropic interactions are ubiquitous in nature, responsible for
30 the formation of lipid membranes,[1] viruses[2] and protein assemblies.[3] Recently, this biologically inspired anisotropy has been exploited for the controlled self-assembly of DNA-functionalized colloids.[4,5] The controlled self-assembly of soft materials including nanoparticles,[6] colloids[4] and
35 biomacromolecules[7] is a fundamental theme in soft matter science and has generated new structures with important applications in photonics, as catalytic supports and as biomaterials.[8,9] The degree to which self assembly can be controlled is greatly improved by using anisotropic particles.

40
These anisotropic, or "patchy" particles, may be anisotropic in shape,[10,11,12,13] surface charge density,[14] size ratio[15] or have surface functionality.[4,8] Anisotropy is central to protein self assembly.[16] The self-assembly of proteins is important for both
45 protein folding[17] and in the assembly of higher order structures.[18] The formation of higher order protein assemblies includes virus formation,[19] assemblies required for normal protein function and

condensed phases that occur during the pathogenesis of several protein condensation diseases.[20] However, despite many decades
50 of research, we are not yet at the point where predictable control of protein self-assembly occurs and this is most apparent in the limited success of protein crystallization which, for many classes of protein remains elusive.[21]

55 Protein phase diagrams map the physical state of a protein for a range of solution conditions[21] and in many ways this behavior is similar to phase diagrams for colloidal particles.[22,23] A more complete analysis of protein phase behavior however must include anisotropic (or aeolotopic) interactions between protein
60 particles to accurately describe the experimentally observed behavior.[24,25] While the protein surface is inherently patchy or anisotropic due to the variation in its surface chemistry from different amino acids, in some cases, an averaged net inter-protein interaction potential is enough to explain protein
65 behaviour. However, in many cases, this is insufficient since, at particular positions on the protein surface, specific amino acids (or groups of amino acids) contribute more to the behavior of the protein than others, which is not captured in an isotropic interaction potential. Therefore, it is convenient to think of the
70 protein consisting of a number of patches (corresponding to either a single amino acid, or a group of amino acids) giving rise to an anisotropic interaction potential. Several aeolotopic, or "patchy" particle models have been developed to describe the directional nature of protein interactions[25,26,27,28] and the process of protein
75 self-assembly to form protein condensed states;[25,29] however, predictable protein phase behavior has yet to materialize experimentally.

Human γD-crystallin (HGD) is one of the most important
80 structural proteins found in the eye-lens. The phase behavior for HGD is defined by net attractive short-range interactions responsible for liquid-liquid phase separation and

crystallization.[30,31] HGD does not spontaneously crystallize. One of the speculated reasons is that the directional interactions of proteins are not compatible with simple crystal lattices.[27] However, several known single amino acid substitutions of HGD
5 result in proteins that crystallize readily with significantly lower solubility than for the native protein.[31] Such single mutants of HGD are associated with congenital cataracts.[30,31] A second type of single amino acid substitution in HGD at position 23 (P23V, P23T, P23S) inverts the temperature dependence of the solubility
10 line leading to unusual phase behavior.[32,33,34]

Given the large number of single point mutants for HGD, which are known to alter the phase behavior of the protein, we can use this as a model to further probe how changes to surface amino
15 acids in specific locations (and hence protein anisotropy) contributes to the behavior of the resulting protein. In the simplest case, a single amino acid substitution is used. For HGD, a number of these have been previously described.[31,32] To increase complexity, we have instead introduced two mutations to
20 the same protein. Both mutations have been demonstrated to give rise to distinctly different phase behaviours. In this case, by mapping the behavior of this new protein, we can clearly identify if characteristics of the original single mutant proteins are maintained and, furthermore, if the positions of the phase
25 boundaries associated with the single mutations are the same in the double mutant. Using the characteristics of these individual single mutants, we have created three double mutants of HGD incorporating two distinct properties; 1) Inverted solubility and 2) Propensity to crystallize.

30

We find that the position of the liquidus phase boundary for each of the single mutants, predicts both equilibrium and kinetic properties of the double mutants formed. In all cases, the double mutant protein retains its secondary structure and crystallizes
35 after mutagenesis. Additionally, two different crystal types are observed for one of these new proteins, each with opposite temperature dependence of the solubility line. These fascinating observations are important in their own right, but this approach may be seen as a strategy for exploring the degree to which
40 aeolotopic interactions are responsible for protein self-assembly and crystallization experimentally.

Results and Discussion

Phase Diagrams of Double Mutants.

45 The R36S single mutant of HGD crystallizes spontaneously at significantly lower concentrations than the native protein.[31] While arginine 36 is not a crystal contact point in the native protein, it becomes one in the mutant, resulting in this behavior.[35,36] Replacing proline at position 23 inverts the temperature
50 dependence of the solubility line, i.e. the protein forms aggregates/crystals at higher temperatures, which melt as the temperature is lowered, but is not a crystal contact point in either the native or single mutant protein.[32,37] The first double mutant of HGD that we have created, P23VR36S, incorporates both types
55 of mutant. CD spectroscopy measurements confirm that no significant change in protein secondary structure (Figure S1) or

tertiary structure (Figure S2) occurs upon mutagenesis. Furthermore, the double mutant protein forms protein crystals, consistent with the behavior of the single mutant R36S (Figure
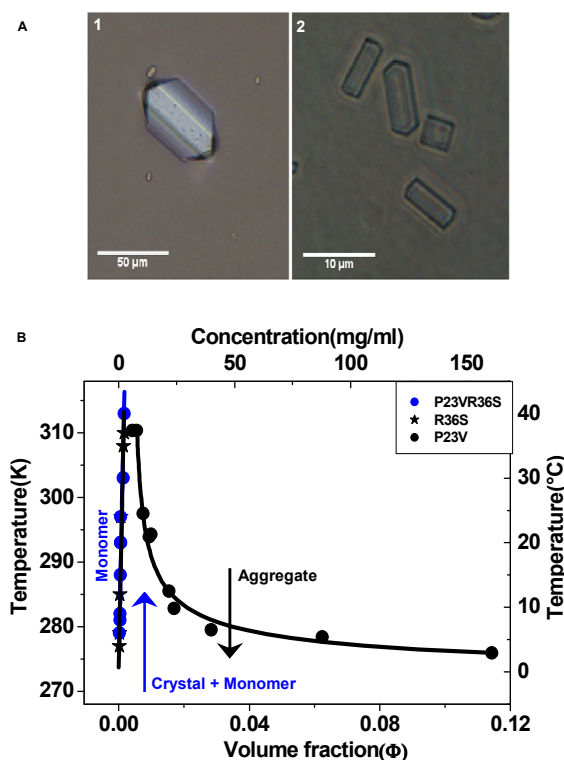60 1A). In fact, the equilibrium solubility lines for R36S and P23VR36S overlap (Figure 1B).

Figure 1. Crystal formation for mutants of human gamma D crystallin. When two mutants with distinct phase behavior are
90 incorporated in one double mutant, its phase behavior is predicted by that of the single mutants. Top: A1, microscopy image for the crystals formed by the single mutant HGDR36S. A2, crystals formed by the double mutant HGDP23VR36S. Both crystal types have the same morphology. B: Phase diagram for the two single mutants of HGD, R36S and P23V
95 (black lines) and the double mutant P23VR36S (blue). Liquidus lines for all three mutants are shown. Data for the previously described mutants P23V and R36S is taken from ref 32 and 30 respectively. R36S crystals have normal solubility, while P23V mutants exhibit an inverse temperature dependence of the solubility line. The double mutant crystal
100 shows only normal temperature dependence (arrows indicate the temperature change required for crystal melting).

The temperature dependence of the solubility line is normal and not inverted. Hence, the behaviours associated with the P23V mutation are suppressed in this double mutant. This is at first not
105 surprising and can be expected from the phase behavior of the single mutants. The solubility line associated with the P23V mutant occurs at significantly higher concentrations than it is possible to obtain for this double mutant, since crystallization occurs at low protein concentrations and on short timescales. This
110 initial mutant provides useful insights; that the double mutants are soluble, retain their native structure and that they crystallize in a similar way to one of the parent single mutants. However, it also raises a number of questions. The phase behavior of the P23V single mutant is unusual. The co-existence curves (liquid-

liquid phase separation) for both the native and mutant proteins overlap, indicating that there is little difference in the net-interaction potential between the two proteins. The inverted solubility observed was explained in terms of aeolotopic interactions, which were only engaged in the solid phase. Here however, this amino acid substitution is present in the double mutant and the solid phase formed has normal solubility. The behavior associated with the P23V substitution is not engaged at all. While the phase behavior of this initial double mutant is consistent with the phase diagrams for each of the parent single mutants, a number of questions remain. Hence, we created a second double mutant.

The P23T mutant of HGD was used for the second double mutant, since it exhibits the same inverse temperature dependence of the solubility line, but at significantly lower concentrations than for P23V. The double mutant incorporating this and the R36S propensity to crystallize (P23TR36S) also maintains its secondary structure after mutagenesis (Figure S1). In this case, two different protein crystal morphologies occur for this one double mutant protein; one, a rod shaped crystal, which has normal temperature dependence and a second, with rhombic shaped crystals that have inverted solubility (Figure 2A). Again, in this case, the equilibrium solubility lines for the single mutants are similar to those for each of the crystal types formed in the double mutant, and in many ways are qualitatively predictable (Figure 2B), with the liquidus line for the rhombic shaped crystals consistent with that for the P23T single mutant and the rod shaped crystals consistent with the R36S solubility line.
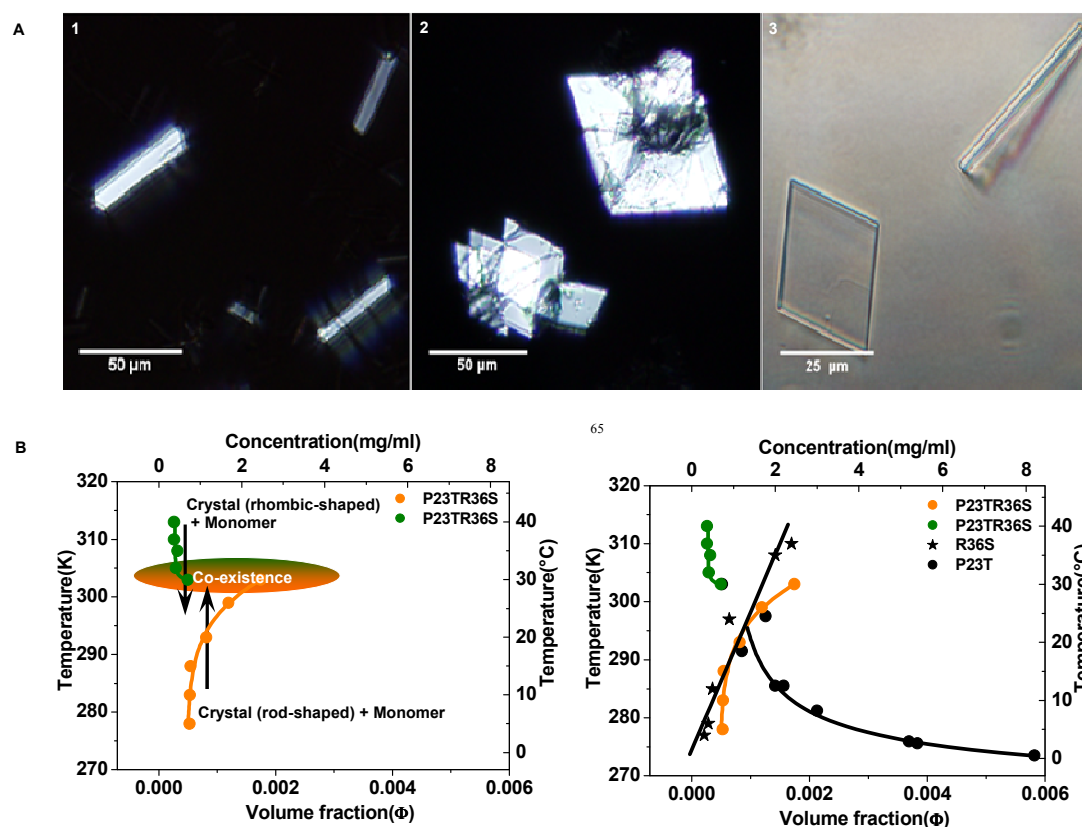

Figure 2. One protein, two crystals. The phase behaviour for the double mutant P23TR36S is fully described by that for the parent single mutants. In this case, two distinct crystal types with well-defined morphologies are formed; Rod shaped crystals, with normal temperature dependence of the solubility line (A1) and rhombic shaped crystals with inverse temperature dependence of the solubility line (A2). Polarization microscopy images are shown in each case, confirming crystal formation. The phase diagram, showing liquidus lines for both single mutants and newly formed double mutant, again indicate that the parent single mutants are predictive of the equilibrium solubility lines for the double mutant (B). Furthermore, co-existence of both crystal types is observed at $\approx$ 303K (30°C) (A3).

Interestingly, neither crystal type is morphologically similar to the P23VR36S double mutant described earlier. Indeed, both crystal types form spontaneously. This is noteworthy, since the P23T single mutant forms aggregates far more easily than crystals. In fact, the solubility line for the single mutant of P23T actually represents a monomer-aggregate equilibrium.[32] However, for the double mutant, only crystals were observed. There is a temperature region $\approx$ 303K (30°C) on the phase diagram where the two different crystal morphologies co-exist (Figure 2B). This observation is unprecedented in that the protein

exhibiting this behavior was essentially predetermined by the behavior of the single mutants.

As a final step, a third mutant type was created, P23VR58H. Again, this mutant incorporates a single amino acid substitution associated with inverted solubility and a second associated with crystallization. The R58H single mutant crystallizes with a liquidus line almost identical to R36S, however the protein has significantly higher solubility (up to 100mg/ml) and crystallization occurs more slowly.[31] Unlike the R36S mutant, position 58 is a crystal contact in the native protein but the strength of the contact at this position actually decreases upon mutagenesis.[36] Like the other double mutants, P23VR58H

maintains its secondary structure after mutagenesis. All three known single mutations at site 23 of HGD form large aggregates which co-exist with protein monomer with an inverted solubility line.[32] P23T and P23V have been shown to crystallize[33,37] and the crystal-monomer solubility line appears to overlap with the aggregate-monomer solubility line.[33] The P23VR58H double mutant also forms large aggregates as the protein is concentrated at temperatures above the expected solubility line for a P23V mutant. However, over time (6 hrs – 1 day depending on the temperature), crystallization of protein occurs on the surface of the aggregates which form immediately after the solubility limit is reached (Figure 3A).
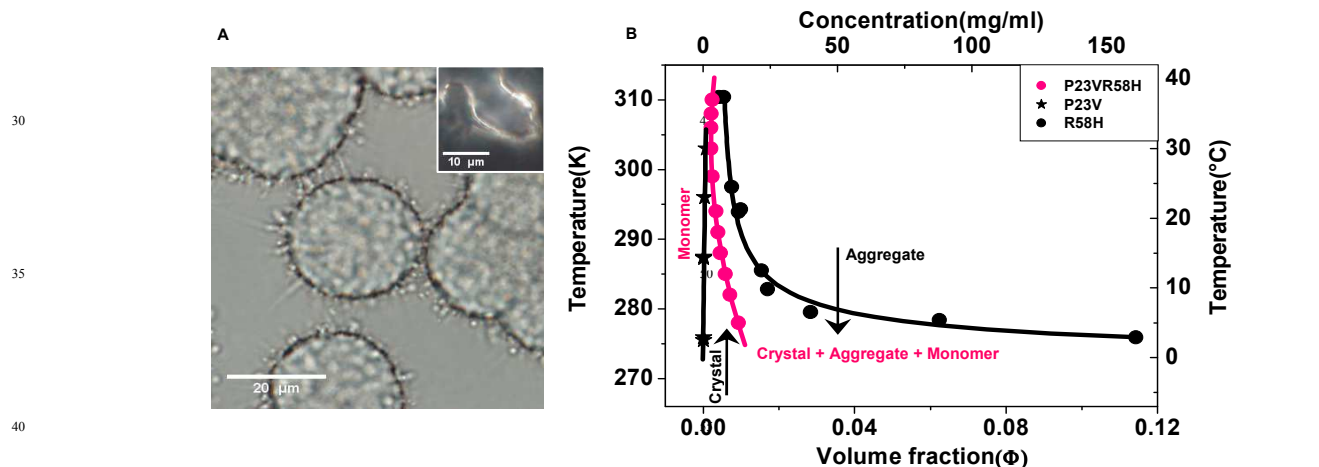

Figure 3. The battle between aggregates and crystals. A, the third double mutant, P23VR58H exhibits both distinctive behaviours of the parent single mutants, however both condensed phases form under the same solution conditions, but with different kinetics, leading to a hybrid structure where protein aggregates (due to the P23V mutation) become encased in protein crystals (formed after aggregate formation, due to the R58H mutation). The crystalline nature of the surface structures was confirmed by polarization microscopy (inset). B, The equilibrium solubility line for the aggregate/crystal structure lies between that for the parent single mutants P23V[32] and R58H[30]. Arrows indicate the solid (crystal, aggregate) to liquid (monomer) transition.

After the formation of protein aggregates (which have inverted solubility), the concentration of monomer in solution decreases but is still sufficiently high to allow protein crystal formation (due to the R58H amino acid substitution). If we attempt to measure the liquidus line for this double mutant we find that both aggregates and crystals co-exist and the solid (crystal + aggregates) - liquid (monomer) equilibrium occurs at concentrations between those expected for the individual single mutants (Figure 3B). The position of this "equilibrium" line moves depending on the temperature at which initial aggregate formation occurs (i.e. how much material is sequestered as aggregates before the onset of crystallization). When less aggregation occurs at lower temperatures, increasing numbers of crystals are formed and the "liquidus" line moves to lower concentrations, closer to that of the single R58H mutant. Indeed, it is possible to suppress aggregate formation by keeping the concentration of protein below the solubility line for the P23V mutant and allowing protein crystallization to occur (Figure 4).

Analysis of the Solubility Data. If we analyse the data further, some interesting observations emerge. In previous work, the transfer chemical potential for the closely related single mutants, P23T, P23S and P23V was calculated from the solubility data (based on the monomer-aggregate solubility line).[33]
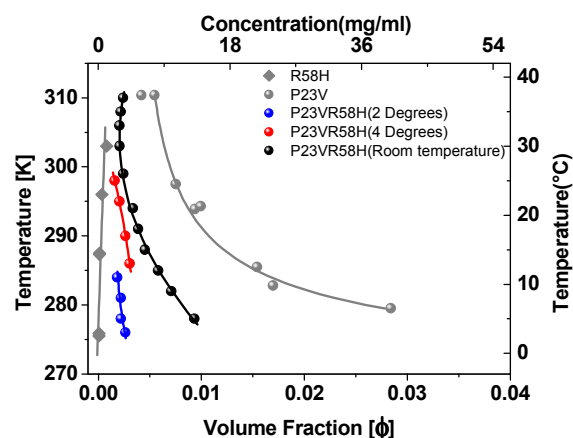

Figure 4. Phase diagram for P23VR58H double mutant displaying shifts in the solubility line depending on the temperature at which initial aggregates/crystals were formed in solution.

We use the same method here to probe if indeed the phase boundaries for the double mutants are quantitatively the same, i.e. is the binding energy for the crystals in the double mutants the

same as for the single mutant variant which gives rise to a specific behavior. The transfer chemical potential, $\Delta\mu_{trans}$ for the double mutant proteins have been estimated from their liquidus/solubility lines previously.[33] In brief, the van't Hoff law relates the volume fraction, ϕ, along the solubility line to the transfer chemical potential, $\Delta\mu_{trans}$ by:

$$\ln\phi = -\frac{\Delta\mu_{trans}}{kT} \qquad (1)$$

By defining a value $X = T_0/T$ and $T_0 = 303K$ (the temperature at which the solubility lines for the R36S and P23T single mutants intersect), and including the second virial coefficient $B_2 = -4[(e^\varepsilon - 1)(\lambda^3 - 1) - 1]$ defined by Lomakin *et. al.*,[25] where $\varepsilon = \varepsilon_{eff}/kT$ from the LLPS for HGD, $\lambda = 1.25$ and $\varepsilon_{eff}/kT_c = 1.27$, the solubility line at low volume fractions ($\phi < \sim 0.02$), becomes:

$$\phi \cong \phi_0 \exp[(V - S)\Delta\tau] \qquad (2)$$

where

$$\phi_0 = \exp-\left(\frac{\Delta\mu_{trans}(T_0)}{kT_0}\right) \qquad (3)$$

and $V = \Delta\mu_{trans}(T_0)/kT_0$; $S = [\partial(\Delta\mu_{trans}/kT_0)\partial\tau]_{\tau=1}$; and $\Delta\tau = (T-T_0)/T_0$. Using the solubility data for the native and double mutant proteins, the transfer chemical potential $\Delta\mu_{trans}$ for each protein was determined. Plotting $\Delta\mu_{trans}/kT_0$ vs. T/T_0, we find that the slope is negative for proteins displaying normal solubility and positive for inverted solubility and is determined by the sign of the value $K = ([\Delta\mu_{trans}(T_0)/kT_0] - S)$.

If we examine the solubility data, we see that the liquidus line for the R36S and the P23VR36S mutants both qualitatively and quantitatively overlap (Figure 1B and Table 1). There are two competing contributions to the protein behaviour and in this case the R36S mutation dominates this behavior. The values for V for both the single and double mutants are almost the same (V = 6.9 and 6.8 respectively). No evidence of inverted solubility is present.

For the P23TR36S mutant, several interesting observations are made. The solubility line for the R36S-type crystal indicates that there is a slight shift in the position of the liquidus line to higher concentrations in the double mutant protein, due to a change in binding energy ≈ 0.4 kT (at 303K). Therefore the P23T mutation is influencing the position of the phase boundary for the R36S-type crystal in the double mutant. For the crystal with inverted solubility, there is a small change in the value of V between the single (P23T) mutant and the double mutant of ≈ 0.2 kT (at 303K), but in this case only crystals are observed whereas for the single mutant, aggregation occurs more frequently. This small change in binding energy is unlikely to fully explain this difference in behavior and suggests some influence from the R36S substitution.

Table 1. Values of the parameters S, the slope for $\Delta\mu_{trans}/kT_0$ vs. T/T_0 and V the intercept at $T = T_0 = 303K$ for each protein.

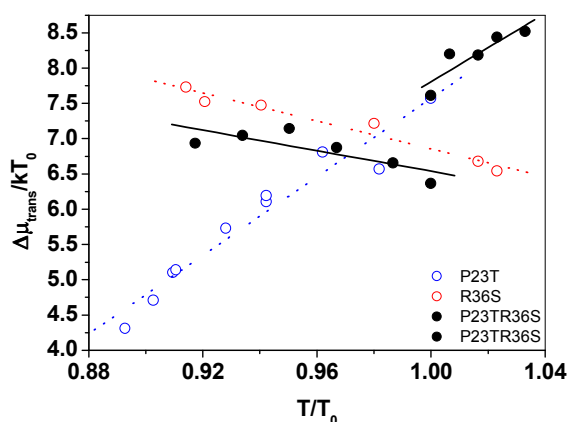| Protein	S	V	K = V - S
HGD	-19	4.4	23.4
P23V	20	5.1	-14.9
P23T	28	7.6	-20.4
R36S	-10	6.9	16.9
P23VR36S	-7.5	6.8	14.3
P23TR36S (Normal)	-7.3	6.5	13.8
P23TR36S (Inverted)	24	7.8	-16.2

Figure 5. Temperature dependence on solubility for single and double mutants of HGD.

Crystallization vs Aggregation.

In the third double mutant the solubility line (or lines) represents three co-existing protein forms; monomer, aggregate and crystal, hence a solubility line analysis cannot provide any useful insights and only a qualitative analysis is possible. It is significant however that both crystals and aggregates co-exist. Thermodynamically, the R58H crystal should be the lowest energy phase for the protein (lowest solubility and crystalline form). However, there is no change in position of the solubility boundary formed once the initial formation of aggregates (with inverted solubility) at a particular temperature has been established over several days. The solubility boundary can be shifted to lower concentrations by decreasing the temperature at which the initial aggregates are formed, which decreases the concentration of aggregated particles formed before crystallization proceeds on the surface (Figure 4). No crystallization of protein in the bulk is observed and nucleation appears to proceed heterogeneously on the surface of the pre-formed aggregates only. Hence the behavior of the double mutant is controlled by the kinetics of the crystallization process (which would also hold true for the P23VR36S mutant). The relative concentrations of either aggregate or crystal are defined both by the phase boundary of the initial single mutant driving that

behaviour, but also by the rate at which that particular condensed phase forms. While this would explain why the solubility line lies between those of the single mutants, the data does imply that the aggregated material is as thermodynamically stable as the crystal. Conventional thought suggests that the entropy loss due to condensation is balanced by the decrease in free energy due to the formation of contacts in the crystal. The number of contact points in the crystal should also be higher than in the amorphous aggregate, making the crystal the most thermodynamically stable form. It is of course possible that the aggregated material is kinetically trapped and would eventually crystallise resulting in liquidus line similar to the R58H single mutant if given sufficient time. However, the original work on the P23V single mutant did demonstrate that aggregated and crystalline material had the same solubility line. The phase diagram for this double mutant seems consistent with that view and we believe that in fact the monomer-aggregate-crystal co-existence is an equilibrium arrangement.

Double mutants as a means to examine protein anisotropy.

This work offers important insights into the factors driving the self-assembly of proteins mediated by anisotropic interactions. The formation of new crystal contacts is considered a driving force in promoting protein crystallization.[38] For some proteins this is likely to be true, but for the four individual mutant proteins described here, only the R36S mutant forms a new crystal contact. It is possible that new crystal contacts are formed in the double mutants that do not exist in the single mutants. However there is a remarkable preservation in the position of the phase boundaries for each condensed phase formed in the double mutants when compared to the single mutant form, which suggests that the existing contact points are maintained. Based on the P23VR36S mutant alone, it would appear that the crystal contact formed by the R36S substitution is dominating the behaviour of the double mutant. The phase diagram for the P23VR58H mutant however would suggest that this might just be a kinetic effect. The R36S-type crystal forms most quickly and it is never possible to obtain protein concentrations high enough to form the aggregates associated with the P23V mutant, but these are observed in the P23VR58H mutant since crystallization is sufficiently slow to allow the faster aggregation process to occur. Of course, x-ray analysis of the double mutants will be required to confirm this. X-ray structures for the native protein and the R36S, R58H and P23T single mutants of the protein are known and have been discussed in some detail [30-37]. For each single and double mutant described here, no major change in protein structure is observed upon mutagenesis. Symmetry is a further matter to consider.[18] Therefore, the factors controlling the crystallization and, more widely, the self-assembly of proteins are complex and will depend on a combination of influences including solution conditions, the formation of crystal contacts and symmetry which are of course influenced strongly by the protein anisotropy. These directional interactions, as a consequence of the microscopic surface features of the protein, complicate efforts to manipulate protein-protein interactions to facilitate crystallization and have been difficult to explore systematically until now. Screening a range of solution conditions to find those suitable for crystallization are often unsuccessful, even for closely homologous proteins.[16] Other strategies to improve the success rate of crystallization by surface engineering of proteins using surface-energy reduction (SER) have had success.[16] For the double mutants created here, no screening of solutions conditions was required and the amino acids selected for mutagenesis were chosen on the basis of the impact a single mutation had on the phase behavior of the protein. Hence, this is a new approach, which has produced from three attempts, three proteins for which the equilibrium solubility lines are preserved in the double mutant and are qualitatively (and semi-quantitatively) predicted.

Experimental Section

Materials and Methods.

Analytical grade sodium hydroxide, sodium chloride, glacial acetic acid, sodium phosphate dibasic heptahydrate, sodium dihydrogen phosphate dihydrate and sodium azide used for the preparation of buffers were purchased from Fisher Scientific (Dublin, Ireland), and used without further purification. All buffers were prepared in Milli-Q water and adjusted to the correct pH using sodium hydroxide or HCl as appropriate.

Molecular cloning, Site-directed mutagenesis, Expression and Purification of Recombinant proteins.

Recombinant HGD was prepared by amplification of the coding sequence from a human fetal lens cDNA library, over-expressed in E.coli (BL21-Gold (DE3) competent cells, Stratagene, U.S.A), isolated and purified as reported previously.[30] Oligonucleotide primers for the desired mutations were synthesized by Life Technologies (Dublin, Ireland). Mutagenesis was performed with QuikChange II site-directed mutagenesis kit from Stratagene (USA) using DNA primers and HGD wild type plasmid DNA isolated and purified by using QIAGEN plasmid purification midi kit (Qiagen GmbH, Hilden, Germany). Each of the three double mutant plasmids obtained after mutagenesis were sequenced with the T7 promoter primer by using an automated capillary DNA sequencer (MRCPPU, College of Life Sciences, University of Dundee, Scotland) and was found to contain the desired mutations. Three mutant proteins (P23VR36S, P23TR36S and P23VR58H), each containing two amino acid substitutions were expressed and purified by the same methodology as for HGD.[30]

Characterization of mutant proteins.

For each mutant protein, SDS-PAGE and size exclusion HPLC were used to confirm protein purity at >98%. The intact molecular weight for the mutant proteins were analysed by electrospray ionization mass spectrometry (Finger Prints Proteomics Facility, College of Life Sciences, University of Dundee), which confirmed molecular mass of 20,607 Da for HGD, 20,539 $\pm$1 Da for P23VR36S, 20,541 $\pm$1 Da for P23TR36S and 20,589 $\pm$1 Da for P23VR58H. These mass values are consistent with those expected for these amino acid substitutions. Circular dichroism (CD) spectra were recorded on a JASCO J-810 spectropolarimeter (University of Glasgow, Scotland). Near-UV CD spectra were measured and normalized with respect to protein concentration; far-UV CD spectra were measured and normalized with respect to the concentration of the backbone

peptide bonds.

Solubility measurements (Liquidus lines).

Protein solutions were prepared initially by diafiltration against 100mM sodium phosphate buffer, pH 7.0 using Ultracel 10KDa ultrafiltration discs (Merck Millipore, Co.Cork, Ireland). Protein concentrations for the double mutants were determined by UV absorbance using the extinction coefficient value of 2.09 $mg^{-1}.ml.cm^{-1}$ after filtration through 0.22µm Millex-GV Millipore (Merck Millipore, Co.Cork, Ireland) syringe driven filter units. When required, protein solutions were further concentrated by ultrafiltration using Amicon Ultra-4 centrifugal filter units (Merck Millipore, Co.Cork, Ireland) and the protein concentration re-established by UV absorbance.

Crystallization/aggregation was allowed to proceed under the relevant conditions for each mutant protein and the liquidus lines were measured. The liquidus lines for P23VR36S, P23TR36S and P23VR58H were measured as described previously.[31] Protein crystals of the double mutants P23VR36S, P23TR36S and P23VR58H formed over a time period of 12 hours were allowed to settle and then separated from the supernatant. The crystals were dispersed in fresh buffer and stirred at a constant temperature to ensure uniform mixing. Every point on the solubility line is a measure of the protein monomer concentration monitored over a time period of 24-48 hours that has attained equilibrium with the protein crystals in solution (i.e. until there are no further changes in protein concentration). Phase diagrams were drawn using Origin v6.1 software.

Imaging.

Protein crystals were imaged by phase contrast and polarization microscopy. An Olympus BX61 microscope with either 60x or 100x magnification oil-immersion lens were utilized for visual observation of the protein crystals formed. The images were then recorded using CellF software and image analysis was performed using ImageJ software.

Conclusions

This work suggests a mechanism by which protein anisotropic (or aeolotopic) interactions can be probed in a systematic manner. This type of data is critical if good molecular models to predict protein behavior are to be developed. For the P23TR36S mutant, we created a protein, which forms two different crystal types, one that melts when the solution is heated and one that melts when the solution is cooled, with co-existence of the two crystal forms at 303K, the point at which the individual liquidus lines for the single mutant variants overlap. This observation is unprecedented. On a broader level, this work is a starting point which will require a combination of further experiments and complementary simulations to more clearly understand the interplay between the complex, competing forces controlling protein self-assembly and crystallization. However, it is clear that the surface characteristics of the protein, defined by the surface amino acids, can lead to a variety of condensed phases for the same protein. A change in the external environment, e.g. temperature, results in some amino acids contributing more to the protein self-assembly behaviour than others, leading to the variety of structures that we observe. For these three double mutants, we can conclude the following. The positions of the phase boundaries in the double mutants are predicted on the basis of the relative thermodynamic stabilities of the individual single mutants (i.e the mutant with lowest solubility at particular point on the phase diagram will dominate), with one exception. If it is possible to form both phases under the same solutions conditions, the kinetics of condensed phase that forms first will determine the position of the equilibrium solubility line of the double mutant. The kinetics for the formation of each phase in the double mutant can be predicted if the kinetics for each single mutant is known.

Notes and references

Department of Chemistry, Maynooth University, Maynooth, Co.Kildare, Ireland.; E-mail: jennifer.mcmanus@nuim.ie

† Electronic Supplementary Information (ESI) available. See DOI: 10.1039/b000000x/

1   S. Kawato, K. Jr. Kinosita and A. Ikegami, *Biochemistry*, 1977, **16**, 2319-2324.

2   S. N. Fejer, D. Chakrabarti and D. J. Wales, *ACS Nano*, 2010, **4**, 219-228.

3   A. Kurut, B. A. Persson, T. Akesson, J. Forsman and M. Lund, *J. Phys. Chem. Lett.*, 2012, **3**, 731-734.

4   Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, *Nature*, 2012, **491**, 51-55.

5   N. Geerts and E. Eiser, *Soft Matter*, 2010, **6**, 4647-4660.

6   M. Grzelczak, J. Vermant, E. M. Furst and L. M. Liz-Marzán, *ACS Nano*, 2010, **4**, 3591-3605.

7   G. M. Whitesides and M. Boncheva, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 4769-4774.

8   Q. Chen, S. C. Bae and S. Granick, *Nature*, 2011, **469**, 381-384.

9   S. C. Glotzer and M. J. Solomon, *Nat. Mater.*, 2007, **6**, 557-562.

10 M. R. Jones, R. J. Macfarlane, B. Lee, J. Zhang, K. L. Young, A. J. Senesi and C. A. Mirkin, *Nat. Mater.*, 2010, **9**, 913-917.

11 M. R. Jones, R. J. Macfarlane, A. E. Prigodich, P. C. Patel and C. A. Mirkin, *J. Am. Chem. Soc.*, 2011, **133**, 18865-18869.

12 S. Sacanna and D. J. Pine, *Curr. Opin. Colloid Interface Sci.*, 2011, **16**, 96-105.

13 K. L. Young, M. L. Personick, M. Engel, P. F. Damasceno, S. N. Barnaby, R. Bleher, T. Li, S. C. Glotzer, B. Lee and C. A. Mirkin, *Angew. Chem. Int. Ed.*, 2013, **52**, 13980-13984.

14 J. I. Park, T. D. Nguyen, G. Q. Silveira, J. H. Bahng, S. Srivastava, G. Zhao, K. Sun, P. Zhang, S. C. Glotzer and N. A. Kotov, *Nat. Commun.*, 2014, **5**, 3593.

15 R. L. Marson, C. L. Phillips, J. A. Anderson and S. C. Glotzer, *Nano Lett.*, 2014, **14**, 2071-2078.

16 Z. S. Derewenda, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **D67**, 243-248.

17 C. M. Dobson, *Nature*, 2003, **426**, 884-890.

18 Y. T. Lai, N. P. King and T. O. Yeates, *Trends Cell Biol.*, 2012, **22**, 653-661.

19 O. M. Elrad and M. F. Hagan, *Nano Lett.*, 2008, **8**, 3850-3857.

20 G. B. Benedek, *Invest. Ophthalmol. Vis. Sci.*, 1997, **38**, 1911-1921.

21 N. Asherie, *Methods*, 2004, **34**, 266-272.

22 N. Asherie, A. Lomakin and G. B. Benedek, *Phys. Rev.Lett.*, 1996, **77**, 4832-4835.

23   D. Rosenbaum, P. C. Zamora, C. F. Zukoski, *Phys. Rev. Lett.*, 1996, **76**, 150-153.

24   N. Dorsaz, L. Filion, F. Smallenburg and D. Frenkel, *Farad. Discuss.*, 2012, **159**, 9-21.

25   A. Lomakin, N. Asherie and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 9465-9468.

26   C. D. Michele, S. Gabrielli, P. Tartaglia and F. Sciortino, *J. Phys. Chem. B*, 2006, **110**, 8064-8079.

27   R. P. Sear, *J. Chem. Phys.*, 1999, **111**, 4800-4806.

28   E. Bianchi, P. Tartaglia, E. Zaccarelli, F. Sciortino, *J. Chem. Phys.*, 2008, **128**, 144504-144510.

29   J. P. K. Doye, A. A. Louis, I-Chun. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok and R. Lyus, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2197-2205.

30   A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. A. King, N. H. Lubsen, D. Walton and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 1993-1998.

31   A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. A. King and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 6116-6120.

32   A. Pande, O. Annunziata, N. Asherie, O. Ogun, G. B. Benedek and J. Pande, *Biochemistry*, 2005, **44**, 2491-2500.

33   J. J. McManus, A. Lomakin, O. Ogun, A. Pande, M. Basan, J. Pande and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 16856-16861.

34   A. Pande, J. Zhang, P. R. Banerjee, S. S. Puttamadappa, A. Shekhtman and J. Pande, *Biochem. Biophys. Res. Commun.*, 2009, **382**, 196-199.

35   S. Kmoch, J. Brynda, B. Asfaw, K. Bezouška, P. Novák, P. Rezácová, L. Ondrová, M. Filipec, J. Sedlácek and M. Elleder, *Hum. Mol. Genet.*, 2000, **9**, 1779-1786.

36   A. Basak, O. Bateman, C. Slingsby, A. Pande, N. Asherie, O. Ogun, G. B. Benedek and J. Pande, *J. Mol. Biol.*, 2003, **328**, 1137-1147.

37   F. Ji, L. M. I. Koharudin, J. Jung and A. M. Gronenborn, *Proteins: Struct., Funct., Bioinf.*, 2013, **81**, 1493-1498.

38   D. Fusco, J. J. Headd, A. De Simone, J. Wang and P. Charbonneau, *Soft Matter*, 2014, **10**, 290-32.