Title

The “Sticky Patch” Model of Crystallization and Modification of Proteins for Enhanced Crystallizability

Authors

Derewenda, Zygmunt S

Godzik, Adam

Publication Date

2017

DOI

10.1007/978-1-4939-7000-1_4

Peer reviewed

HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2017 August 24.

Published in final edited form as:

Methods Mol Biol. 2017 ; 1607: 77–115. doi:10.1007/978-1-4939-7000-1_4.

The 'Sticky Patch' Model of Crystallization and Modification of Proteins for Enhanced Crystallizability

Zygmunt S. Derewenda and

Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine

Adam Godzik

Bioinformatics and Systems Biology Program, Sanford Burnham Prebys Medical Discovery Institute

Abstract

Crystallization of macromolecules has long been perceived as a stochastic process, which cannot be predicted or controlled. This is consistent with another popular notion that the interactions of molecules within the crystal, i.e. crystal contacts, are essentially random and devoid of specific physicochemical features. In contrast, functionally relevant surfaces, such as oligomerization interfaces and specific protein-protein interaction sites, are under evolutionary pressures so their amino acid composition, structure and topology are distinct. However, current theoretical and experimental studies are significantly changing our understanding of the nature of crystallization. The increasingly popular 'sticky patch' model, derived from soft matter physics, describes crystallization as a process driven by interactions between select, specific surface patches, with properties thermodynamically favorable for cohesive interactions. Independent support for this model comes from various sources including structural studies and bioinformatics. Proteins that are recalcitrant to crystallization can be modified for enhanced crystallizability through chemical or mutational modification of their surface to effectively engineer 'sticky patches' which would drive crystallization. Here, we discuss the current state of knowledge of the relationship between the microscopic properties of the target macromolecule and its crystallizability, focusing on the 'sticky patch' model. We discuss state-of-art *in silico* methods that evaluate the propensity of a given target protein to form crystals based on these relationships, with the objective to design of variants with modified molecular surface properties and enhanced crystallization propensity. We illustrate this discussion with specific cases where these approaches allowed to generate crystals suitable for structural analysis.

Keywords

Protein crystallization; sticky patch model; surface entropy reduction; lysine methylation

1 Introduction

Conventional, single-crystal X-ray diffraction analysis is only feasible if the target sample (e.g. protein, protein-DNA complex, etc.) can be obtained in the crystal form capable of diffracting X-rays. Early on, pioneers of macromolecular crystallography relied on a

portfolio of proteins and viruses crystallized in the 1920s and 1930s, by such biochemists as James Sumner, John Northrop and Wendell Stanley, who received the 1946 Nobel Prize for Chemistry for this work [1]. Their efforts were never intended to generate crystals for structural analyses: this was an early era of protein biochemistry, and crystallization was the final step in the purification of proteins. Given the earlier discovery of X-ray diffraction in 1912, there was obvious interest if protein crystals, observed since mid-19th century, also diffract X-ray. It was John Desmond Bernal, who showed in 1934 that only if protein crystals are kept within their mother liquor (he used pepsin), they exhibit beautiful diffraction up to virtually atomic resolution [2], effectively demonstrating that these macromolecules have distinct chemical structure, in contrast to the then prevailing colloidal theory [3]. This set the stage for macromolecular crystallography, which revolutionized contemporary biology and medicine by providing tools to explore structure-function relationships in protein, nucleic acids, intact viruses, and other complex structures as ribosomes [4].

From the very beginning it was obvious that proteins exhibit very different propensities to form crystals: some crystallize under a range of conditions yielding distinct crystal forms, others precipitate in an amorphous way, form gels or oils. As early as 1909, the physiologist E.T Reichert and the mineralogist A.P. Brown (both from the University of Pennsylvania) published a remarkable volume, entitled '*The Crystallography of Haemoglobins*' in which they described hundreds of distinct morphologies of crystals of hemoglobin obtained from blood of various vertebrates [5]. Numerous micrographs (Fig 1) illustrate how species-dependent variations between proteins result in crystals with markedly different morphologies. At the time, the underlying chemical nature was unknown; today we understand that amino acid sequence differences are the cause. Half a century later, in 1953, John Kendrew showed that crystal forms of the same protein from different species may show very different diffraction properties; by screening myoglobin crystals from different sources he was able to select the best diffracting crystal form (sperm whale myoglobin) which eventually led to crystallographic characterization of the first protein molecule [6].

As the list of proteins known to crystallize upon purification was slowly getting exhausted, in the 1960s crystallographers faced a challenge of having to crystallize their target macromolecules first. Thus the science (or art) of protein crystallization was born. One of the first concepts to be introduced was that of screening of a spectrum of conditions, be it buffers or precipitants, in search of one where a protein would precipitate in a crystalline form as the solution passed the saturation point (see Chapter … for detailed discussion). Little attention was paid to recording data from failed experiments, and the whole process was thought to be stochastic: some proteins crystallized, others did not, and no correlation with solution conditions was apparent. The microscopic nature of interactions underlying crystallization was virtually ignored. When crystal contacts where finally recognized as a valid target of structural analysis, early studies concluded that they are essentially random patches of protein surfaces, attesting to the stochastic process of assembly of proteins into nuclei and crystals [7–9]. This was bad news, because stochastic phenomena cannot be easily controlled and directed, and so macromolecular crystallization appeared to be destined forever to Edisonian approach of trial and error, i.e. random screening.

The advent of molecular biology in the 1980s, and more recently of high-throughput (HT) methods that enabled the Structural Genomics initiative, brought substantial changes to the way we approach crystallization. We are no longer restricted to wild-type proteins from natural sources; in most cases the targets are recombinant proteins, and often they are custom-designed fragments of specific targets. This makes it possible, in principle, to manipulate the initial cDNA construct to enhance its propensity to crystallize. Further, the HT Structural Genomics laboratories introduced highly standardized crystallization pipelines, carefully recording all outcomes, both positive and negative. This uncovered hitherto unappreciated correlations between protein properties (as encoded by the amino acid sequences) and their crystallization propensities, clearly revealing that some are much more amenable to crystallization than others [10–14]. As the number of structures deposited in the Protein Data Bank grew at a rapid pace, new opportunities for data mining opened up. Further, studies from fields such as soft matter physics, bioinformatics, and molecular biology began to reshape our understanding of the microscopic mechanisms underlying crystallization, coming to conclusions that are in stark contrast to the ‘stochastic model’. In its place, a new general ‘sticky patch’ model has emerged, emphasizing the microscopic variations of the macromolecular surface and the physicochemical phenomena behind low-affinity, yet specific molecular interactions, including those governing the formation of contacts in nascent crystals. In this chapter, we review the current microscopic view of crystallization based on the premise of directional, specific molecular interactions, and discuss experimental methods that exploit those concepts for the design of chemically or mutationally modified protein targets for enhanced crystallizability (NB: Macromolecular crystallography encompasses broadly the study of proteins, nucleic acids and their complexes as well as a range of chemical entities; most of this Chapter focuses specifically on proteins, but the phenomena and methods described herein also apply to all protein complexes).

2 Theoretical and Experimental Evidence for the ‘Sticky Patch’ Model

2.2 Crystallization *in silico*: lessons from soft matter physics

Our current understanding of protein crystallization owes much to experimental and theoretical soft matter physics, and particularly to the study of colloids [15]. More than two decades ago, it has been observed that both colloidal particles and proteins tend to crystallize when the osmotic second virial coefficient, B_2, which depends only on the pair interaction between the particles, lies in the favorable, crystallization ‘slot’[16]. Studies of crystallization of isotropic spheres show that it proceeds through a slow process of nucleation, whose rate is enhanced close to the metastable liquid-vapor coexistence binodal, followed by growth [17]. Proteins may behave in this fashion, and they (like colloids) can also form amorphous aggregates that kinetically impair crystallization below the binodal, and they can be (meta)stable in the crystallization slot of the second virial coefficient without crystallization ever taking place [18–21]. The complexity of the phenomenon prompted theoretical and computational efforts to generate suitable models for phase transitions and crystallization of both these systems. Initial attempts focused on simple models with a relatively short-range interparticle attraction [22]. Subsequently, various pair potentials have been studied, allowing for variable (yet still small) range attraction and more

complex potentials that included a repulsive barrier [23]. In general terms, these models required the particles in the liquid phase to be very close to each other for the attraction force to become significant. Initially this phase behavior was thought to be reasonably well suited as a starting point for simulations of globular proteins, with their roughly spherical shape, isotropic electrostatic repulsion and short-range van der Waals and hydrophobic attraction.

One of the key assumptions underlying the early colloidal models was the isotropic nature of the interactions [22]. Most obviously, the microscopic nature of the colloid and protein solid phase is different, as illustrated by the fact that proteins do not form close packed crystals. More subtly, the overall shape of the bimodal is also incompatible with isotropic attraction. These problems suggest that additional features should be included in the minimal model. One key such features is interaction anisotropy [24–26]. In fact, a similar question has surfaced in colloid physics, particularly given the considerable effort to design complex colloidal particles with physically patterned surfaces, or ‘patchy’ particles [27]. It was therefore natural to extend this notion to proteins, in order to capture the orientation dependence of protein-protein interactions. Lomakin *et al.* [28] first developed a model taking account of the spatial variation of the protein surface, underlying varying short-range interactions. It was used, among others, by Gogelein *et al* [29] to describe the phase behavior of lysozyme dispersions. This early model involves repulsive screened Coulomb interactions, with incorporated attractive surface patches that mediate interactions between molecules.

More detailed computer simulations then revealed the impact of attractive surface patches on the crystal lattice, concluding that anisotropic interactions can lead to a variety of different crystal structures, depending on the geometry and strength of the patchy interactions [30]. A variant of the model, which contained competing sets of attractive patches, has been used to explain why nearly identical conditions sometimes yield different crystal forms of the same protein, specifically homodimeric an monomeric crystal forms [31]. The concept was further expanded by the introduction of a model based on spheres decorated randomly with a large number of attractive patches, to study the formation of structures with $P2_12_12_1$ symmetry, the most prevalent space group among proteins [32]. The conclusions of this study are particularly interesting. The unit cell with the lowest energy is not necessarily the one that grows fastest, because growth is favored when new particles attach through enough patches to the growth front and if particles can attach in crystallographically inequivalent positions with the same affinity. Importantly, when non-specific interactions that are not part of the set of crystal contacts are few and weaker than the actual crystal contacts, both nucleation and growth are successful [32]. Recently, a computational study of crystals of three proteins from the rubredoxin family characterized crystal contacts and used it to parametrize patchy particles models (Fig 2) [33]. This first explicit bridge between soft matter physics with structural biology not only obtained reasonable theoretical phase diagrams, but also microscopic-level insight into specific patterns of residues that make up crystal contacts.

To conclude, the ‘sticky patch’ model described crystallization as a non-stochastic process, made possible by few, attractive patches on the surface of a protein, which under specific crystallization conditions impact critically on the success of nucleation and growth type as

well as crystal lattice. We will now discuss how other, parallel advances in the understanding of crystallization thermodynamics, the chemistry and stereochemistry of crystal contacts, and the recent progress in the understanding of weak protein-protein interactions, all of which support and complement the 'sticky patch' model.

2.2 Thermodynamics of crystallization: a microscopic view

The canonical, macroscopic view of crystallization thermodynamics, including phase diagrams [34] (see Chapter…), has little predictive value and does not address the microscopic mechanisms of molecular interactions leading to three-dimensional order during crystal growth, or—conversely—does not explain the failure of molecules to form crystals under conditions of supersaturation, as opposed to amorphous precipitate or gel. However, recent interpretations of thermodynamic changes that accompany crystallization of macromolecules give us new insights into the microscopic aspects of the phenomenon, and taken together with the 'sticky patch model' allow to answer a number of questions [35–37].

Like any equilibrium process, crystallization is driven by the reduction in Gibbs free energy, $\mathbf{\Delta G^o_{cryst}}$, on transfer of molecules from solution to the crystalline phase. At constant temperature **T,** this is the net effect of changes in enthalpy ($\mathbf{\Delta H^o_{cryst}}$) and entropy ($\mathbf{\Delta S^o_{cryst}}$):

$$\Delta G^{o}{}_{cryst} = \Delta H^{o}{}_{cryst} - T\Delta G^{o}{}_{cryst}$$

Direct determination of $\mathbf{\Delta G^o_{cryst}}$ is difficult, but available data suggest that it is modestly negative, i.e. in the range of −10 to −100 kJ mol^{-1} [37]. This explains why crystallization is subject to 'butterfly effects', because even extremely subtle phenomena (e.g. minute change of temperature) that can occur at any point during the process can shift $\mathbf{\Delta G^o_{cryst}}$ into positive or negative range, with dramatic impact on the outcome of the process.

An interesting question is if either $\mathbf{\Delta H^o_{cryst}}$ (enthalpy) or $\mathbf{\Delta S^o_{cryst}}$ (entropy), preferentially drive the free energy change. In the case of macromolecular crystallization, enthalpy changes cannot be large, because no strong bonds are formed. In few cases where experimental measurements of $\mathbf{\Delta H^o_{cryst}}$ were made, the values were consistently small [38–40]. This suggests that entropic effects should be playing a dominant role, although the notion is counter-intuitive, because the formation and growth of the three-dimensionally ordered crystal is by definition associated with significant, unfavorable decrease in entropy. Indeed, a loss of three translational and three rotational degrees of freedom per molecule, is estimated to result in a change of $\mathbf{T\Delta S^o_{cryst}}$ of 30 to 100 kJ mol^{-1} [41,42]. However, it is when we take into account the microscopic effects associated with the formation of crystal contacts, that the situation gets much worse.

In general terms, a protein molecule can be described as having a solvent-inaccessible core with rigid secondary structure elements, and more flexible, solvent exposed loops that create the molecular surface. Much of this surface is populated by conformationally labile, large side chains, such as Lys, Arg, Glu and Gln. (NB: It has been suggested, in fact that the presence of high-entropy side chains on protein surfaces could be the result of early evolutionary pressures [43]; given the high protein concentrations in living cells, and the

associated overcrowding effects, it is reasonable to hypothesize that globular proteins have been under evolutionary pressures to avoid non-functional specific interactions, hence the presence of the 'entropy shield' on the surface [44,45]). When protein molecules assemble within nascent crystals, specific intermolecular contacts are formed. At these sites, many flexible side chains become sequestered and consequently ordered (Fig 3). Although the magnitude of side chain conformational entropies of Lys, Arg, Glu, and Gln are highly dependent on the rotamer and secondary structure context, it is generally agreed that it may range at room temperature from ~2 kJ mol^{-1} in regions of defined secondary structure to ~8 kJ mol^{-1} in coil regions[46,47]. Thus, formation of contacts that involve many such side chains is thermodynamically prohibitive. Also, the N- and C-termini of the polypeptide chain, often disordered in solution, may also become trapped at crystal contacts, leading to further decrease in entropy. The same applies to flexible loops, sequestered upon crystallization.

To identify the specific thermodynamic driving phenomenon in crystallization, we have to turn to solvent effects. Any high-resolution crystal structure of a protein reveals large numbers of ordered water molecules covering both hydrophobic and hydrophilic solvent-exposed surfaces [48,49]. While lacking the dynamic aspect, these crystal structures are largely representative of the hydration shell, that encapsulates macromolecules in solution and is two to three molecules deep, [50,51]. As the protein molecules become incorporated into the growing crystal, and direct contacts between them form, the structured solvent is released from the surfaces. Based on the entropy gain of transfer of one molecule of water from clathrate, crystal hydrate, or other ice-like structures into the liquid phase, it has been estimated that release of one water from a protein surface at ambient temperature into the bulk phase leads to an entropy gain of ~ 6 kJ mol^{-1} [52]. If sufficient number of water molecules are released into the bulk solvent, the overall entropy gain will compensate for the losses ascribed to other phenomena (see above) and provide the driving free energy for crystallization [36,37,53]. Indeed, the estimated values of $-T\Delta S^{o}_{solvent}$ (i.e. free energy decrease due to water release) during macromolecular crystallization at ambient temperature range from ~30 kJ mol^{-1} to ~180 kJ mol^{-1}, corresponding to the release of ~ 5 to 30 water or solvent molecules [37,54,36].

It is important to note, that the thermodynamic outcome of the crystallization process can only be probed experimentally on a macroscopic scale as the combined effect of all the molecular interactions and solvent effects. In reality, a single molecule (or independent structural entity) within a crystal, may form as few as three or as many as eighteen interfaces with satellite molecules [55]. Such interfaces in crystal structures are identifiable solely by distance criteria (e.g. with atoms in the two molecules separated by < 4.5 Å), and there is essentially no way to discriminate between cohesive and non-cohesive interactions. However, only three cohesive contacts are necessary for the integrity of three-dimensional crystal, with the exception of space group $P2_12_12_1$, where only two are required [56]. Remaining contacts may have neutral, cohesive, or distinctly repulsive character and may be forced on the ensemble by the intrinsic ruggedness of the molecular shape.

In conclusion, the microscopic aspects of thermodynamics of intermolecular interaction during crystallization are consistent with a model in which the assembly of protein

molecules in the nuclei and nascent crystals is orientation-dependent in order to minimize the unfavorable entropy gains stemming from loss of flexibility to fragments of protein structure (exposed side chins, loops and flexible termini), while maximizing favorable solvent effects. Only select surface 'sticky patches', with a tendency to form cohesive interactions, serve that purpose, enforcing specific orientations.

2.3 Structural support for the 'sticky patch' model

The crystal structures deposited in the Protein Data Bank (PDB) offer a wealth of structural data for the analysis of macromolecular packing and the nature of the protein-protein interactions (PPIs). As remarked earlier, the main effort was to identify biologically 'functional' interfaces against the background of what was believe to be random interactions. A number of methods were developed for automated *in silico* analysis of the interfaces and identification of functional interfaces, including those taking advantage of the evolutionary conservation as defined by Shannon entropy [57–60]. Currently, the most popular method for the analysis of protein-protein interfaces in crystals is the PISA algorithm available as a server (https://www.ebi.ac.uk/msd-srv/prot_int/) [61].

Unfortunately, a strictly binary classification of protein-protein interactions, i.e. functional *vs* crystal contacts (*vide* random), is overly simplistic. A more recent study utilizing a non-redundant PDB database of strictly monomeric proteins, and a regression analysis methodology, demonstrated that crystal contacts are not random, but are in fact enriched in small and hydrophobic amino acids (e.g. Ala, Val), and depleted in large and charged residues, such as Lys, Glu and Gln, in a manner similar to functional PPIs (Fig 4)[62]. This is an important observation, even though the dataset of interfaces that were subject to analysis by necessity had to include all contacts identified by distance criteria, regardless of whether they are thermodynamically cohesive or not. It is almost certain that if it were possible to computationally identify the attractive interactions only, their amino acid content would be even more distinct. These observations are in full agreement with the simulations mentioned earlier that assessed the impact of the weak, non-specific interactions on nucleation and crystal growth [32].

Another important observation comes from the analysis and comparisons of interfaces across different crystal forms of the same, or homologous, proteins. As was already noted in Chapter (…), macromolecules show considerable polymorphism and often the same protein crystallizes in various forms, sometimes from the same solution conditions. It has been recently reported that the portion of the PDB entries with at least two crystal forms available is 64% [63]. Although reproducibility of crystal contacts in different crystal forms of multimeric proteins is normally taken as evidence of physiological homo-oligomerization, such functionality is often not known *in vivo*. Further, even taking conservatively annotated monomeric proteins into consideration, a third shows shared interfaces in different crystal forms. A striking example is the homodimeric association of the globular domain of RhoGDI (discussed in more detail in 4.2) which is reproduced across multiple crystal forms obtained using dramatically different crystallization conditions (Fig 5). All these observations strongly support the 'sticky patch model', and show that protein surfaces are decorated with distinct patches mediating specific interactions which under favorable

conditions allow for the formation of crystal contacts. The interfaces that mediate such contacts are not explicitly distinct form so-called ‘functional interfaces’ but the two are simply examples of opposite ends of a continuum of interactions, all of which have potential functional significance, even though in most cases we have not yet linked a particular interaction to a physiological phenomenon.

2.4 The ‘Sticky Patch’ Model and Ultra-Weak Protein-Protein Interactions (UWPPIs)

The diversity of protein-protein interactions (PPIs) is well illustrated by the differences in their amino acid composition and size, ranging from surfaces burying in excess of 2,000 $Å^2$, with distinctive hydrophobic core, to small patches limited to few amino acids of diverse nature. As a consequence, PPIs span a huge range of affinities, from the tightest interactions with $\boldsymbol{K_D}$ in the < pM range, to weak and ultra-weak interactions (WPPIs and UWPPIs) with $\boldsymbol{K_D}$ > 1 μM and even > 100 μM, respectively. Historically, tight and obligate interactions have always been under intense scrutiny, but WPPIs and UWPPIS were only recently appreciated as biologically important [64]. This is in part because transient and weak complexes are often very difficult to identify, isolate and evaluate by methods such as TAP (tandem affinity purification), surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC). Nevertheless, (U)WPPIs are increasingly recognized as key factors in gene regulation [65], cell adhesion [66], virus assembly [67], and other phenomena. On the technical side, heteronuclear NMR emerged as a powerful method for probing (U)WPPIs [68,69].

It is important to realize that *in vivo*, (U)WPPIs are likely to show significantly higher *effective* affinities than those measured *in vitro* for isolated proteins [68]. *In vivo* proteins function under conditions of macromolecular crowding [70,71], with total concentrations ranging from 80 mg/mL in blood plasma, to ~200 mg/ml and ~400 mg/mL in eukaryotic and prokaryotic cells, respectively [72,44,71]. The thermodynamic activities are consequently significantly higher than actual concentrations. Under crowding conditions, the activity coefficient of a 30 kDa protein triples, and for a 50 kDa protein increases by two orders of magnitude [72]. This dramatically favors association of molecules, with $\boldsymbol{K_D}$ often lowered by two to three orders of magnitude, depending on the molecule’s size and shape. The equilibrium of the association of a monomeric protein of 40 kDa into a tetramer shifts by $10^3 – 10^5$ in the *E. coli* cell environment, compared to isolated protein [72]. It is now well established that macromolecular crowding but modulates (U)WPPIs in a biologically relevant manner. For example, when ubiquitin is monitored by NMR in the *E. coli* cell, it tumbles so slowly that no detectable HSQC spectra can be recorded [68]. This is due to transient, targeted (U)WPPIs interactions with large proteins or complexes, which under crowding conditions have much higher affinities than those determined *in vitro*. A number of computational studies aimed at characterization of (U)WPPIs revealed several differences between weak, transient interactions and tight, obligate associations. Importantly, it has been shown that total accessible surface area (ASA) and polarity of the relevant surface patches constitute critical parameters [73,74].

Crystal contacts constitute an unexplored wealth of information regarding protein surfaces that may engage in (U)WPPIs of functional importance. For example, the 1.8 Å resolution

structure of human erythrocyte ubiquitin [75] showed a crystal contact involving Leu8, Ile44 and Val70 (Fig 6). This contact buries a modest 386 $Å^2$ of surface, and is only the second largest. Its biological function was recognized only significantly later [76]. Similarly, the original crystal structure of the protein tyrosine phosphatase [77] revealed a crystal contact mediated by Tyr130 and Tyr131, to which no functional significance was initially attributed. Much later, NMR titration experiments showed it to be as an ultra-weak (K_D ~ 1.5 mM) interaction, and functional studies revealed its significance [78].

To conclude, many of the 'sticky patches' mediating contacts in protein crystals may have hitherto unknown functional significance. In a more general sense, all cohesive crystal contacts represent sites where the protein may potentially interact with other proteins, especially under molecular crowding conditions. It seems that rather than to attempt a binary classification of PPIs, it is safer to see these interactions as forming a continuum, from ultra-weak to high-affinity, all playing some role in protein's physiology.

2.5 Sequence-derived properties and crystallization

The physicochemical properties of protein surfaces are defined singularly by the solvent-exposed amino acids, and therefore by the amino acid composition and sequence. It is therefore quite reasonable to assume that sequence based properties of protein should be correlated with the presence and type of attractive patches, and therefore with the crystallization outcome. In other words, if 'sticky patches' constitute an integral feature of easily crystallizable protein, one should be able to detect their fingerprint using sequence analysis. Indeed, extensive datamining *in silico* studies have been recently made possible by the databases accumulated by HT Structural Genomics Centers. Unlike the worldwide Biomolecular Crystallization Database[79], and to some extent the PDB, these new databases contain information about both negative and positive outcomes of millions of crystallization experiments, making it possible to probe the issue of what sequence-dependent biophysical properties correlate with the binary outcome of crystallization using regression analysis and other mathematical methods. Here we briefly discuss three specific studies and their conclusions, with the emphasis on the relevance to the 'sticky patch' model.

The study focusing on the proteome of *Thermatoga maritima* was carried out by the Joint Center for Structural Genomics (JCSG) [80]. Detailed in silico analyses were carried out for the 1877 Open Reading Frames (ORFs) of *T. maritima* and the subsets of those proteins that were successfully overexpressed (539 cases) and crystallized (464). Among the properties that were negatively correlated with crystallization were: excessive polypeptide length, i.e. > ~600 amino acids; presence of transmembrane helices; high isoelectric point; low percentage of charged residues; and a high GRAVY hydropathy index, i.e. > 0.3. Although this information was helpful in the filtering of potential targets for structural studies, it did not reveal much about the microscopic aspects of crystallization. Another, somewhat more informative study probed a diverse group of nearly 700 proteins investigated by the Northeast Structural Genomic Consortium (NESGC) [81]. This study looked at an expanded set of molecular properties, such as thermal stability and oligomerization, but also at the frequency of each amino acid, mean hydrophobicity, mean side chain entropy, total and net

electrostatic charge, pI, the fraction of residues predicted to be disordered, and chain length. The sequence-derived parameters were analyzed using logistic regression to evaluate the impact of a continuous variable on the binary outcome of crystallization screens. Hydropathy (Gravy index) and side-chain entropy exhibited strong negative correlation with crystallization success rates. Interestingly, it was also discovered that high frequencies of Lys and Glu amino acids, correlated negatively with crystallization outcome (Fig 7). These conclusions are in agreement with the thermodynamic argument that preponderance of high-conformational entropy side chains on the surface reduces the chances for suitable 'sticky patches' that can mediate crystal contacts.

A third study looked in detail at the behavior of 182 proteins (also from the NESGC) which were each subject to extensive crystallization screen using 1,536 conditions and developed by the High-Throughput Crystallization Screening Laboratory of the Hauptman Woodward Medical Research Institute [82]. Statistical models were trained on this sample capturing trends driving crystallization. Once again, low level of side chain entropy was found to be correlated with positive crystallization outcomes. In addition, a new correlation was found between crystallization and complementary electrostatic interactions. The study concluded that crystal contacts have 'specific physico-chemical signature even if they are not biologically functional' [82].

Taken together, these analyses of sequence derived properties are consistent with the 'sticky patch' model, identifying both side chain entropy, hydrophobicity and electrostatics as surface properties which render specific surface patches as particularly conducive to mediation of specific intermolecular interactions.

3 Prediction of protein propensity for crystallization

Given the existence of detectable correlations between sequence-derived protein physiochemical properties and the propensity to crystallize, it should be possible to design predictive algorithms that evaluate in more rigorous ways the probability of a given protein yielding crystals in an extensive screen. Various "rules of thumb" based on anecdotal observations have been used by individual crystallographer to select and optimize constructs since be early days of crystallography. The first effort to automate this process [83] was based on data-mining on a positive protein set (i.e. proteins with solved structures). Comprehensive negative datasets (lists of proteins and constructs which failed to crystallize) were not available until Protein Structure Initiative (PSI) started producing and screening large sets of proteins and reporting both successes and failures. Availability of such data, collected in the databases of individual PSI centers and, with some caveats, in the TargetDB database [84], enabled development of a first generation of algorithms for evaluation of probability of crystallization of proteins [85–88]. The first publicly available server that provided such evaluation interactively, XtalPred [89] used statistical analysis of seven physicochemical features, sequence length, iso-electric point, gravy hydropathy index, number of residues in the longest disordered region, protein instability index and two different measures of the amount of coiled-coil structure, to develop a single "crystalizability score". Since then, more complex models were developed, often in conjunction with machine learning techniques. These models including ParCrys [90], CRYSTALP2 [91],

MetaPPCP [92], PXS[81], MCSG-Z score [93], PPCpred [94], and XtalPred-RF [95](Fig 8) have allowed users to assess the probability of successful structure determination prior to any experimental work and adjust their target selection strategies. Such algorithms were most useful in the context of high throughput structure determination such as done in structural genomics centers [96–99], where typically protein families, but not specific proteins were targeted for structure determination. In such application, several, typically 5–10 most promising candidates from a protein family were selected for structure determination and successful structure determination of any of them was considered a success. Even modest enrichment in the number of crystallizable proteins in the target pool as compared to random selection improved overall production in the structural genomics centers and allowed them to solve thousands of protein structures, including hundreds of first representatives of novel protein families.

Individual structural biology groups, which often target specific, high value targets still used such approaches [100–102] but often found them lacking resolution needed to distinguish changes to crystallization propensities made by small changes in construct boundaries or individual mutations. New generation of algorithms, now in development, is aiming at the first task [83], while the second clearly remains out of reach of statistics based methods.

Failure of methods based on average physicochemical features of the protein to provide more decisive improvements in selecting or designing optimal constructs for structure determination is easy to explain in the context of the ‘sticky patch model’ of crystallization. While average values of hydrophobicity or instability can effectively predict protein solubility and recognize some features detrimental to crystallization (such as long disorder segments) they do not see individual crystal interfaces and methods targeting individual “sticky patches” are needed to improve on statistical models.

4 Target protein modification for enhanced crystallizability

The ‘sticky patch’ model of crystallization opens a new, exciting possibility for direct enhancement of success rates in crystallization screens by modifying the surfaces of the target protein or complex. Briefly, if crystallization is facilitated by the existence of specific ‘sticky patches’ on the surface of the target molecule, then it should be possible to engineer such patches by chemical modification or site-directed mutagenesis. The key question is what modifications can be effectively used, and what should they target.

Recombinant methods and protein chemistry offer a plethora of avenues for protein modification, and comprehensive discussion of all these methods is beyond the scope of this chapter (NB: some of these methods may be designed to overcome other potential bottlenecks, such as intermolecular disulfide bridges, low solubility, etc. [103,104]). Here we are primarily interested in methods that modify very specific patches to achieve the potential for cohesive interactions, driving the formation of crystal contacts. In the most general terms there are four such strategies: (1) optimization of the recombinant construct to remove high-entropy N- and C-termini and loops; (2) the use of proteases to remove unstructured regions; (3) mutational modification of protein surface to reduce excess conformational entropy (the Surface Entropy Reduction protocol); (4) chemical modification of the protein surface by

targeting specific amino acids; (5) the use of small molecule additives that specifically bind in surface crevices and modify local surface properties; (6) the use of protein chaperones which may stabilize the target protein or complex and provide additional surfaces with 'sticky patches' assisting in the crystallization of the target. We will briefly address each of these strategies, and refer the readers to a number of extensive reviews.

4.1 Construct optimization and proteolytic digestion

As is evident from our earlier discussion, the presence of disordered regions in crystallization targets, i.e. N- and C-termini and large flexible loops, is very unfavorable. This is a very important point because most target proteins under study are fragments, e.g. signaling or catalytic domains, and the correct choice of N- and C-terminal boundaries is of paramount importance. Historically, isolated stable domains have been obtained using limited proteolysis and subsequent purification of the smallest functional domain. The contemporary approach is *in situ* proteolysis, i.e. addition of small amounts of select proteases to the crystallization mixture, so that the enzyme acts on the protein under crystallization conditions, and the proteolysed fragment is allowed to form crystals in the same drop [105–107]. Another strategy is *in silico* analysis, using tools such a XtalPred [95] or DisMeta [108], to identify the boundaries of the folded stable fragment, and to clone the target fragment accordingly [109–112]. The functional core units can also be identified experimentally, following limited proteolysis, by mass spectrometry [113]. Finally, experimental methods can be used to identify the disordered regions directly, such as deuterium-hydrogen exchange coupled to mass spectrometry (DXMS) [114–116], or NMR [117]. Unfortunately, many target variants may have to be screened to identify one amenable to crystallization, because in some cases short disordered fragments may even help[118]. For example, in the case of the MAPKAP kinase 2, sixteen truncation variants were assayed, all containing the catalytic domain, and shown to have dramatically different solubility and propensity for crystallization [109].

A complementary approach is to remove disordered loops, is such can be identified by other means and are believed to interfere with crystallization. For example, the variant used in the successful crystallization of the HIV gp120 envelope glycoprotein had two flexible loops which were replaced with Gly-Ala-Gly linkages to obtain a crystallizable variant [119,120]. In the case of 8R-lipoxygenase, the replacement of a flexible Ca^{2+}-dependent membrane insertion loop, consisting of five amino acids, with a Gly-Ser dipeptide resulted in crystals that diffracted to a resolution 1 Å higher than the wild-type protein [121]. An interesting variation of this approach was introduced for the preparation of crystals of the β-subunit of the signal recognition particle receptor. A twenty-six residue-long flexible loop was removed, but instead of replacing it with a shorter sequence, the authors connected the native N- and C-termini of the protein using a heptapeptide GGGSGGG, thus creating circular permutation of the polypeptide chain [122].

In summary, removal of flexible fragments in the crystallization targets reduces the possibility of a prohibitive loss of entropy during crystal formation as the unstructured regions become sequestered in the crystal contacts, and exposes the cohesive patches which can mediate thermodynamically favorable crystal contacts.

4.2 Surface Entropy Reduction – engineering ‘sticky patches’

The Surface Entropy Reduction (SER) strategy, uses site-directed mutagenesis to generate protein variants with surface cohesive patches (sticky patches) designed to increase the propensity for crystallization. The concept was initially based on a broadly formulated hypothesis, consistent with the microscopic interpretation of entropy contribution to crystallization, that solvent exposed amino acids with long, flexible side chains (e.g. Lys or Glu) impede crystallization because high conformational entropy would is lost as the amino acid is sequestered in crystal contacts [123]. It was therefore suggested that surface patches enriched in these amino acids are very unlikely to mediate protein interactions, and consequently crystal contacts. Conversely, variants in which Lys and/or Glu within such patches were mutated to small residues such as Ala, should have increased probability of being involved in interactions that could consequently mediate crystal contacts. This is essentially as direct way of engineering ‘sticky patches’ to enhance the protein’s crystallizability. The hypothesis was first tested using a model system of the globular domain of the human signaling protein RhoGDI, which is unusually rich in Lys and Glu, and is recalcitrant to crystallization in its wild-type form [123–125]. The mutated variants containing single or multiple Lys→Ala or Glu→Ala mutations within a single patch (identified by close sequence proximity) have indeed shown much higher success rate in routine crystallization screens [123–125]. Importantly, the crystal structures of these variant revealed that the mutated patches directly mediate select crystal contacts, corroborating the underlying hypothesis and the ‘sticky patch’ model. One of the interesting of the RhoGDI variants was that multiple mutations within a single patch were more noticeably effective than single mutants.

The SER strategy was successfully applied to a number of new proteins that were recalcitrant to crystallization in the wild-type form. The first new structure to be solved was the RGSL domain from the signaling protein PDZRhoGEF [126,127]: a triple mutant in which two Lys and one Glu were mutated to alanine yielded good quality crystals (Fig 9). Other successful applications quickly followed, and a number of high-profile structures were solved. Among them were: EscJ, a component of the type III secretion system, the structure of which helps to understand key aspects of virulence in Gram-negative pathogens [128]; ALIX/AIP, programmed cell death 6-interacting protein, key to the understanding of mechanisms involved retrovirus budding and endosomal protein sorting [129]; an HIV-capsid component, which helps understand the maturation of HIV [130]; a complex of the K^+ gated channel, KChIP1 with the Kv channel interacting protein (Kv4.3 T1), [131]; and the crystal structure of the BetP Na^+/betaine symporter [132]. In virtually all cases, the strategy was to target clusters of 2 to 3 Lys or Glu (or both) residues that were consecutive in sequence, and change them to Ala.

The basic premise of the SER strategy is strongly corroborated by the aforementioned data-mining studies showing that preponderance of high-entropy amino acids in protein sequence is negatively correlated with crystallization success [81,82]. In an effort to better understand what mutational strategy is optimal, a more comprehensive study was carried out using the same model system of RhoGDI, expanding the target amino acids to Lys, Glu and Gln (which has the same conformational entropy as Glu, but has no charge) and replacing them

with Ser, Thr and His [133]. Further, an additional approach was tested in which Tyr was used as another alternative amino acid to replace the high-entropy residues. The rationale there was that tyrosines occur with relatively high frequency at protein-protein interfaces [134], and are known to play a crucial role at antibody-antigen recognition sites [135]. Tyrosine has only two degrees of conformational freedom, compared to four in Lys and three in Glu, so that the entropy loss upon sequestration at an interface is lower, but also has a bulky hydrophobic moiety as well as a hydroxyl group capable of forming directional H-bonds. Interestingly, the latter was most successful substitution, in terms of the success rate of crystallization in a standard screen [133]. As was the case with other SER variants, those containing Tyr also crystallized with the engineered patch mediating crystal contact. However, it has also been observed that Tyr variants, especially those with two or three of these residues adjacent in sequence, display significantly lower expression yields, limiting the applicability of the method. Tyrosine remains a rare choice for the replacement of Lys, Gln and Glu in the SER approach.

Currently, the choice of which residues in the target protein should be mutated is made easier by a dedicated server (SERp, http://services.mbi.ucla.edu/SER/) with a predictive algorithm to identify suitable surface sites for mutagenesis, based on amino acid sequence information only [136,137] (Fig 10). The server seeks to identify solvent-exposed loops with patches populated by high-entropy residues (Lys, Glu and Gln). The investigator then makes the choice what type of amino acid is to replace the wild-type residues (typically this is Ala).

The SERp server has been used in a multitude of studies to design crystallizable variants of many proteins and macromolecular complexes. There are currently over 150 non-redundant entries, with a total of over 450 depositions in the PDB based on SER strategy. This database allows for an interesting insight how the engineered SER patches affect molecular packing and consequently crystallization at the microscopic level. A preliminary survey (to be submitted) reveals that in over 90% of cases the mutated patches are directly involved in crystal contact. This is an irrefutable validation of the original notion that mutations of high-entropy residues create ‘sticky patches’ with enhanced propensity to mediate protein-protein interactions and crystal contacts. It is also interesting to note that the SER ‘sticky patches’ generate crystal contacts with unique topological patters. Most of the mediate homotypic contacts, i.e. interactions between two identical engineered patches in neighboring molecules (see Fig 9). This specific interaction gives rise to two-fold symmetry, either crystallographic or non-crystallographic, in which case a dimer occupied the asymmetric unit. A minority of SER patches forms heterotypic contacts, in which the mutated patch interacts with a completely different, wild-type patch on the surface of the neighboring molecule. These contacts are associated with crystallographic screw axes typically in such space groups as ($P2_1$, $P2_12_12_1$ or C2) but are also responsible in many cases for translational contacts, especially in the rather rare P1 space group (Fig 11).

Although precise calculation of the **ΔG** free energy change based on structure is notoriously difficult, it is interesting to note that crystal contacts generated by SER appear to be generally thermodynamically cohesive, based on calculations by PISA [61]. This suggest that the SER contacts constitute in fact one of the minimum three, cohesive interactions actually responsible for the integrity of the lattice. It is also possible that SER-mediated

interactions, particularly the homotypic ones that lead to dimerization, exist in solution prior to nucleation and crystallization.

The vast majority of the successful applications of the SER strategy was limited to engineering a single patch. In a typical case, the SER crystal contact either generates a crystallographic (or non-crystallographic) oligomer (i.e. primary contact), or mediates interactions between oligomers (secondary contacts). We note, that often the primary contacts have significantly larger buried interfaces than the secondary contacts.

SER strategy offers also another advantage; it can be used to generate novel crystal forms with superior diffraction qualities, in those cases where wild type protein yields a poorly diffracting crystal forms. In a majority of crystallographic studies, one is typically satisfied if screens yield crystals that allow for the determination of structure to ~2.1–2.5 Å resolution. Effort is typically invested in the optimization of the crystals obtained in a screen, rather than in searching for other crystal forms. However, as has been often demonstrated, the quality of diffraction is dependent on a particular crystal form, rather than being correlated to a specific protein. Thus, if the wild type protein crystallizes, the variants generated by the SER approach are very likely to yield novel crystal forms, often with better diffraction quality and higher resolution. This has been demonstrated early on during the studies of RhoGDI, which in its wild-type form never yields crystals diffracting to better than ~2.8 Å resolution. In contrast, the double mutant E154A, E155A resulted in crystals which allowed for collection of data and refinement of the structure to 1.3 Å resolution [124]. One of the additional advantages of having multiple crystal forms is that the packing of molecules nay be quite different, with certain specific surfaces, such as active sites, open to solvent in some forms, while occluded in other forms. The availability of different forms allows to choose one best suited for particular functional experiments.

The combination of the potential advantages associated with using SER as a method to produce alternative crystal forms of target proteins, has made this technique very useful and popular in drug discovery, and was incorporated into the arsenal of tools in the pharmaceutical industry. One of the first published successful examples of this approach was the improvement in the quality of crystal of the intracellular kinase domain of the insulin-like growth factor, a potential drug target [138]; the resolution was increased from 2.7 Å to 1.5 Å. The HIV-1 reverse transcriptase was successfully engineered to yield crystals diffracting to 1.8 Å; there are 39 PDB entries for this protein [139–143]. A SER variant of the β-site amyloid precursor protein cleaving enzyme (BACE-1), a target in Alzheimer's, was extensively used in drug-discovery research [144–153], with 29 PDB depositions. A drug discovery effort targeting *Trypanosoma brucei*, the causal agent of sleeping sickness, was made possible by SER variants of the methionyl-tRNA synthetase [154,155], generating 29 coordinate sets of complexes. An effort to design inhibitors of the nitric oxide synthase oxygenase, serving as antibiotics against Gram-positive pathogens, utilized a variant of the enzyme predicted by SERp [156–159]: 46 respective PDB entries are available. Finally, the epidermal growth factor receptor kinase domain, a target for non-small cell lung cancer, has been successfully engineered to yield 21 PDB depositions of complexes with drug leads [160–162].

4.3 Reductive methylation

Although there is a whole range of protocols for chemical modification of proteins [163], only one, i.e. reductive methylation, has become a routine method used to enhance crystallization success rates. This is because it is the only one that is technically facile, quick, and produces homogeneously modified samples. Also, reductive methylation selectively modifies lysines, which – as discussed above –disfavor specific interactions and formation of crystal contacts.

The method was initially introduced in the challenging study of myosin subfragment-1 [164]. Detailed protocols have been published [165,166]; only the free amino groups (ε-amino groups of lysines, and the N-terminal amino group) are modified. Formaldehyde is the methylating agent with dimethylamine-trifluoroborane acting as the reducing agent. The most common outcome is dimethylation (n,N,-dimethyl lysine; dmLys), as the monomethylated derivative is more susceptible to modification that the non-methylated amine. As is invariably the case in crystallization screens, assessment of success rates is not trivial. Nevertheless, HT Structural Genomics centers reported 10 – 30% success rates with selections of proteins recalcitrant to crystallization in unmodified form [167,168,166]. Why does this strategy enhance crystallization? Although methylation results in a slight increase in conformational entropy of Lys side chain, it also increases the size of the hydration shell of the side chain, ordering a number of water molecules [169]. Upon packing at a crystal contact, the site containing dmLys is therefore likely to release more solvent molecules with favorable change in entropy. Further, the methyl groups bound to the Nε are polarized, and therefore capable of participating as donors in C-H…O hydrogen bonds [169]. A careful recent study evaluated 40 protein structures solved using crystals made from methylated samples, and compared them with a non-redundant database of 18,972 non-methylated protein structures; for 10 proteins both wild-type and methylated structures were scrutinized [170]. The results revealed that dmLys is more likely to form interactions with Glu across crystal contacts then unmodified Lys, and that this is correlated with the C-H…O directional bonds mediated by the methyl groups (Fig 12). In the case of a ParB-like nuclease, it has been shown that the methylation protocol resulted in a crystal form stabilized by intermolecular contacts that involve 44 C-H…O interactions mediated by nine dmLys residues [171]. Other effects associated with methylation, which may impact crystallization, include a slight reduction in the isoelectric point (pI) [172].

It should be noted that methylation constitutes just one variation of reductive alkylation, which may involve introduction of larger groups, such as ethyl and isopropyl [166]. However, examples of the latter are rare and have not been reported to be very successful.

To conclude, reductive methylation targets lysines on the protein's surface, and modifies them in a manner that increases the probability of these residues forming cohesive intermolecular interactions as part of 'sticky patches'.

4.4 The use of non-covalently bound small molecule – 'sticky bridges'

The alternative to the use of covalent modification is the use of small molecular weight organic or inorganic compounds which bind specifically – although typically with low

affinity – in crevices of the target protein's surface, and provide modified surface patches mediating crystal contacts, i.e. 'packing bridges'. Sequestration of various small molecules within crystal contacts has been observed quite often in the past, including metal cations, organic and inorganic anions, glycerol, and much larger moieties such as organic inhibitors or DNA oligomers. These molecules could be part of the crystallization mix, they might be carried over accidentally from protein purification protocol, or be purposefully added to the crystallization screen [173]. The small molecules may be multifunctional (i.e. these may be inhibitors stabilizing the enzyme or components of buffers) but here we are only concerned with the manner in which they can sterically mediate crystal contacts.

A recent datamining analysis explored a subset of the PDB database for the presence of small molecules and ions serving as packing bridges, and discovered that about 11.5% of interactions between symmetry-related macromolecules are mediated by a heteroatom (i.e. an atom that does not belong to the macromolecule) [174]. This represented nearly half (45%) of the structures. The small molecules most frequently found within the bridges were sulfate ions, glycerol, 1,2-ethanediol as well as acetate, phosphate and chloride ions and calcium ions [174]. A systematic study of the impact of diverse small molecules (other than the usual buffers and additives) on crystallization, screened 200 compounds with respect to their potential to serve as crystallization 'catalysts' for 81 diverse proteins, using only two fundamental crystallization conditions [173]. Nearly 85% of the proteins crystallized, often in new crystal forms, although they were not subjected to systematic structural investigations that might reveal specific interactions mediated by the additive. However, subsequent applications of this strategy revealed explicit examples of additives promoting crystallization by acting as bridges across crystal contacts. For example, cobalamine added to the crystallization mix was found to mediate contacts between oligomers of Δ1-pyrroline-5-carboxylate dehydrogenease [175]. while a tellurium (VI)-centred polyoxotungstate was found to mediate contacts in crystals of hen egg-white lysozyme [176].

The key problem in this field is the unpredictability of what compounds might be helpful in crystallization, or how they might form the packing bridges. The other important question is if these interactions are indeed cohesive and contribute to the integrity of the crystal, or if they might represent serendipitous 'trapping' of small molecules between macromolecules, that contributes little to the overall thermodynamics balance. It is very likely that many of the examples uncovered in the datamining study of the PDB[8] are indeed cases of fortuitously bound ions in contact with two molecules. Control experiments (i.e. crystallization without these small molecules) were never conducted or reported. However, in those cases where crystallization appears to be contingent on the presence of a small molecule, and if the latter is found to form a packing bridge, it is almost certain that the bridge is thermodynamically cohesive (i.e. 'sticky bridge'). Similarly, if the bridge constitutes one of the primary contacts, it has to be cohesive.

There are also interesting examples which show that specific residues or motifs may be 'coupled' to certain ions or compounds, and consequently they can be introduced into proteins by mutagenesis. For example, crystal of the *E. coli* apo acyl carrier protein, rich in carboxylic acids were obtained in the presence of Zn^{2+} ions which provided bridging interactions [177]. It has been suggested that mutational introduction of aspartates on the

surface in protein with high intrinsic pI could provide a useful strategy for crystallization with metal ions. In a related example, His-Cys pairs were introduced on the surface (using T4 lysozyme as a template), allowing for coordination of Zn^{2+} ions that effectively induced dimerization, and engineering a key crystal contact [178]. In yet another case, Ca^{2+} ions were shown to form a 'sticky bridge' between two molecules of the YkoF, engineered by the SER strategy [179]. Here, the removal of high entropy sidechains exposed main chain carbonyls, creating a metal binding site. Interestingly, a recent theoretical study presented a general model of multivalent cation bridges as a method to activate attractive positive patches on the protein surface, bringing small molecules and ions directly into the realm of the 'sticky patch' model [180].

An example that shows potential for more general application is that of combining the use of sulfates as precipitants with surface engineering. A mutant of RhoGDI with two Arg replacing adjacent Lys residues was crystallized in the presence of amonium sulfate, and the surface ions were found to bridge the Arg-rich surface patches [125]. In this particular case, the bridging sulfates may neutralize potential electrostatic repulsion, allowing this secondary contact to form, although it may not *per se* serve as a cohesive interaction.

To summarize, engineering of crystal contact bridges using small molecules or ions, either into wild-type or mutated protein, offers the possibility of creating a 'sticky bridge', thermodynamically cohesive contact, or allows for creation of an interaction eliminating potential electrostatic repulsion. It is conceivable that more general recipes can be designed to couple this approach with surface mutagenesis.

4.5 Crystallization chaperones – using surrogate surfaces

Perhaps the most challenging and complex strategy of altering the surface of the target protein is to use a partner protein (chaperone) that lends its surface to mediate crystal contacts, enabling the crystallization of the complex. (NB: chaperone proteins also serve other purposes, e.g. they may stabilize a particular conformation or enhance solubility of the target; here we focus exclusively on crystallization.) There are two options: either the chaperone is expressed in fusion with the target protein, or the chaperone is generated separately, and the complex is purified and crystallized. Below we briefly discuss the first approach, and expand more on the second, which is more popular and much more successful.

Given that most proteins are overexpressed for purposes of crystallization as fusion proteins with globular affinity tags (e.g. GST, MBP, thioredoxin, T4 lysozyme) the use of these fusion proteins is an obvious and easy option. A number of such fusion proteins have been crystallized: e.g. the DNA-binding domain of DNA replication-related element-binding factor, DREF, in fusion with GST [181]; and the U2AF homology motif domain of splicing factor Puf60 in fusion with thioredoxin [182]. The drawback is the intrinsic flexibility of a two-domain architecture, which impedes crystallization. A remedy is to shorten the linker between the two proteins, to achieve rigidity owing to steric restraint [183–187].

An alternative to N- or C-terminal fusions is an insertion-fusion, in which the chaperone is inserted into a loop of the target. This approach has been used exclusively in membrane

protein crystallization, and was initially pioneered for the *E. coli* lactose permease, in which cytochrome b562, flavodoxin and T4 lysozyme were tested as chaperones [188,189]. A similar insertion fusion with T4 lysozyme, replacing the third intracellular loop of the β2-adrenergic receptor was a key to successful crystallization allowing for structure determination at 2.4 Å resolution [190,191]. This strategy has since been used in a number of crystallographic studies of the G-protein coupled receptors and other membrane proteins [192].

A more universal alternative to fusion proteins are non-covalent crystallization chaperones, i.e. binding proteins engineered to produce a high-affinity complex with the target macromolecule. The most commonly used engineered chaperones are Fab fragments of antibodies [193–199]. In its canonical version, animals were immunized with target antigen, followed by purification of hybridoma-derived antibodies and their proteolytic digestion to obtain antigen binding fragments [200,193]. Alternatively, the Fab fragment was directly sequenced and expressed in heterologous cells for subsequent use [201]. This strategy is costly, inefficient and prone with challenges. A far more powerful and efficient approach is *in vitro* selection of Fab fragments using phage display [202,203] or, less often, ribosome display [204,205]. Multiple templates have been used, but the most common is that based on the herceptin scaffold. Although initially such synthetic antibodies were weaker binders than wild-type ones [206,207], the problem was overcome by using ‘reduced genetic code’, which uses only select types of amino acids which produce high affinity binders [208,202]. Synthetic Fab fragments can be generated against a broad variety of targets, unique conformations of proteins, complexes, or weak antigens such as RNA. Automated platforms are available for high-throughput production [203]. Many targets have been successfully crystallized using synthetic Fabs based on the herceptin scaffold as chaperones. Recent examples include the Nsp1-Nup49-Nup57 channel nucleoporin heterotrimer bound to Nic96 nuclear pore complex attachment site [209]; human paxillin LD2 and LD4 motifs [210]; structure of the Get3 targeting factor with its membrane protein cargo[211]; and the prolactin receptor [212].

The *in vitro* display methods also allow for engineering of non-Fab scaffolds [207]. Examples include nanobodies, i.e. single chain fragments derived from camelid antibodies [213–215]; fibronectin type III domain (FN3) scaffold [135,216]; and DARPins, i.e. designed ankyrin repeat proteins [217,218], used in the crystallization of several proteins, including the polo-like kinase-1 [219], the integral membrane multidrug transporter AcrB [220], and the receptor-binding protein (RBP, the BppL trimer) of the baseplate complex of the lactococcal phage TP901-1 [221].

The success of the chaperones in crystallization is, of course, dependent on its ability to mediate crystal contacts in a more effective way than the target protein alone. The various chaperones described above are all well studied and all show high propensity for crystallization in isolation. However, they may still be not-optimal, and could be subject to surface engineering or other modifications. The wild-type T4 lysozyme, for example, is not ideally suited because of intrinsic flexibility and recently it has been engineered for the use as an internal fusion in GPCRs by addition of stabilizing disulfides, or by reduction in size of the N-lobe (miniT4); these modified molecules proved to be superior as crystallization

chaperones when fused into the third loop of the M3 muscarinic receptor[222]. In another study, T4 lysozyme has also been modified including the mutation of the three C-terminal residues to Ala to reduce conformational entropy[223]. However, perhaps the most relevant for this discussion is the example of the mutants of MBP specifically engineered by SER mutations for enhanced crystallizability [224]. They were was used as an N-terminal fusion chaperone to crystallize the signal transduction regulator RACK1 from *Arabidopsis thaliana* [224] and the SER patches on MBP indeed served as the 'sticky patches' for the fusion protein.

6 Conclusions and perspectives

The challenge in the field is to obtain sufficiently detailed insight into the mechanism of the PPIs underlying crystallization, to enable us to rationally modify the crystallization experiment and its outcome. Although not long ago this may have seemed like a phantasy, we are not far from realizing this goal, even if only in some specific cases. The progress is vividly illustrated by a recent success in computational design of a protein that not only self-assembles in to yield macroscopic crystals, but does so yielding the expected P6 space group symmetry [225] (Fig 13).

The 'sticky patch' model provides not only a unifying theory for a wide spectrum of PPIs, but also rationalizes many of the thermodynamic macroscopic observations, and paves the way for strategies to rationally modify the macromolecular targets to dramatically enhance their crystallizability, either by covalent or not covalent chemical modification, or by protein engineering. Nevertheless, much remains to be learned about PPIs, and ways in which we can modify them through site directed mutagenesis, and ways in which we can control crystallization. The SER methodology for enhancement of protein crystallizability, has gained considerable support in recent years from various experimental, theoretical and data mining studies which collectively generate the comprehensive patchy model. The underlying concept - the reduction in 'excess surface entropy'- is, of course, an oversimplification, because the mutations of polar, charged residues to Ala or similar smaller amino acids alter many physical properties of the protein, including electrostatic potential and solubility. Nevertheless, the distinct propensity of the SER patches to form crystal contacts, most of which are homotypic and result in transient homodimers, shows that the mutations generate the very 'sticky patches' that the model invokes. A change in pI or solubility could not rationalize these effects, but it is important to take all of these properties into account as well as the role of the solvent. Although many crystal structures have been obtained using proteins modified by SER or reductive methylation, there are also numerous examples of failures of such protocols. One possibility to overcome this problem is a design of multi-patch SER strategy, which could overcome problems with particularly recalcitrant proteins with distinct paucity of attractive patches on their surfaces. Several successful examples have already been reported. The *Arabidopsis* medium/long chain prenyl pyrophosphate synthase has been crystallized using a two-patch variant [226]. The structure revealed that the two SER patches assist in forming an octamer (the wild-type protein is a homodimer in solution) within the asymmetric unit, and generate secondary contacts between the octamers to allow for 3D packing. In another study, a triple-patch SER strategy was necessary to

overcome extreme difficulties in the crystallization of the human vaccinia related kinase 1 (PDB code: 3OP5). Again, all patches were involved in crystal contacts.

Finally, perhaps the most intriguing question is when exactly the crystal contacts are formed, and how do they drive the crystallization process. The current theory of nucleation and crystal growth strongly suggests that protein nuclei form within clusters of protein-dense liquid, metastable with respect to protein solution and hundreds of nanometers in size [53,227]. Within these clusters, overcrowding effects will contribute to significant enhancement of attractive interactions between proteins molecules. Whether formation of specific oligomers, as defined by the 'sticky patches' underlies nucleation, and defines the symmetry of the nascent crystal, will hopefully be elucidated by ongoing research.

Acknowledgments

We thank a number of colleagues whose comments and help with figures have greatly helped in the preparation of this Chapter, including Patrick Charbonneau and Diane Fusco (Duke University), Peter Vekilov (UH), Urszula Derewenda (UVA), Heping Zheng (UVA), Jagoda Mika (UVA); this work was supported by the NIH grant GM095847

LITERATURE

1. Manchester KL. The crystallization of enzymes and virus proteins: laying to rest the colloidal concept of living systems. Endeavour. 2004; 28(1):25–29. [PubMed: 15036925]
2. Bernal JD, Crowfoot D. X-ray photographs of crystalline pepsin. Nature. 1934; 133(3369):794–795.
3. Deichmann U. Collective phenomena and the neglect of molecules: A historical outlook on biology. Hist Phil Life Sci. 2007; 29(1):83–86.
4. Jaskolski M, Dauter Z, Wlodawer A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. Febs J. 2014; 281(18):3985–4009. DOI: 10.1111/febs.12796 [PubMed: 24698025]
5. Reichert, ET., Brown, AP. The differentiation and specificity of corresponding proteins and other vital substances in relation to biological classification and organic evolution: the crystallography of haemoglobin. Carnegie Instituion of Washington; Washington, D.C: 1909.
6. Kendrew JC, Parrish RG, Marrack JR, Orlans ES. The species specificity of myoglobin. Nature. 1954; 174(4438):946–949. [PubMed: 13214049]
7. Janin J, Rodier F. Protein-protein interaction at crystal contacts. Proteins. 1995; 23(4):580–587. [PubMed: 8749854]
8. Carugo O, Argos P. Protein-protein crystal-packing contacts. Protein Sci. 1997; 6(10):2261–2263. [PubMed: 9336849]
9. Janin J. Specific versus non-specific contacts in protein crystals. Nat Struct Biol. 1997; 4(12):973–974. [PubMed: 9406542]
10. Bonanno J. Structural genomics [In Process Citation]. Curr Biol. 1999; 9(23):R871–872. [PubMed: 10607575]
11. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S. Structural genomics: beyond the human genome project. Nat Genet. 1999; 23(2):151–157. [PubMed: 10508510]
12. Chandonia JM, Brenner SE. The impact of structural genomics: expectations and outcomes. Science. 2006; 311(5759):347–351. 311/5759/347 [pii]. DOI: 10.1126/science.1121018 [PubMed: 16424331]
13. Gaasterland T. Structural genomics taking shape. Trends Genet. 1998; 14(4):135. S0168-9525(98)01430-9 [pii]. [PubMed: 9594659]

14. Skarina T, Xu X, Evdokimova E, Savchenko A. High-throughput crystallization screening. Methods Mol Biol. 2014; 1140:159–168. DOI: 10.1007/978-1-4939-0354-2_12 [PubMed: 24590716]
15. Fusco D, Charbonneau P. Soft matter perspective on protein crystal assembly. Colloids Surf B Biointerfaces. 2016; 137:22–31. DOI: 10.1016/j.colsurfb.2015.07.023 [PubMed: 26236019]
16. George A, Wilson WW. Predicting protein crystallization from a dilute solution property. Acta Cryst D. 1994; 50(Pt 4):361–365. DOI: 10.1107/S0907444994001216 [PubMed: 15299385]
17. ten Wolde PR, Frenkel D. Enhancement of protein crystal nucleation by critical density fluctuations. Science. 1997; 277(5334):1975–1978. [PubMed: 9302288]
18. Wilson WW. Light scattering as a diagnostic for protein crystal growth--a practical approach. J Struct Biol. 2003; 142(1):56–65. [PubMed: 12718919]
19. Muschol M, Rosenberger F. Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. J Chem Phys. 1997; 107:1953–1958.
20. Liu Y, Wang X, Ching CB. Toward Further Understanding of Lysozyme Crystallization: Phase Diagram, Protein–Protein Interaction, Nucleation Kinetics, and Growth Kinetics. Cryst Growth Des. 2010; 10:548–558.
21. Lu PJ, Zaccarelli E, Ciulla F, Schofield AB, Sciortino F, Weitz DA. Gelation of particles with short-range attraction. Nature. 2008; 453(7194):499–503. DOI: 10.1038/nature06931 [PubMed: 18497820]
22. Rosenbaum D, Zamora PC, Zukoski CF. Phase behavior of small attractive colloidal particles. Phys Rev Lett. 1996; 76(1):150–153. DOI: 10.1103/PhysRevLett.76.150 [PubMed: 10060456]
23. Noro MG, Frenkel D. Extended corresponding-states behavior for particles with variable range attractions. J Chem Phys. 2000; 113(8):2941–2944. Pii [S0021-9606(00)52432-5]. DOI: 10.1063/1.1288684
24. Doye JP, Louis AA, Lin IC, Allen LR, Noya EG, Wilber AW, Kok HC, Lyus R. Controlling crystallization and its absence: proteins, colloids and patchy models. Phys Chem Chem Phys. 2007; 9(18):2197–2205. DOI: 10.1039/b614955c [PubMed: 17487316]
25. Liu H, Kumar SK, Douglas JF. Self-assembly-induced protein crystallization. Phys Rev Lett. 2009; 103(1):018101.doi: 10.1103/PhysRevLett.103.018101 [PubMed: 19659179]
26. Kern N, Frenkel D. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction. J Chem Phys. 2003; 118:9882–9893.
27. Bianchi E, Blaak R, Likos CN. Patchy colloids: state of the art and perspectives. Phys Chem Chem Phys. 2011; 13(14):6397–6410. DOI: 10.1039/c0cp02296a [PubMed: 21331432]
28. Lomakin A, Asherie N, Benedek GB. Aeolotopic interactions of globular proteins. Proc Natl Acad Sci U S A. 1999; 96(17):9465–9468. [PubMed: 10449715]
29. Gogelein C, Nagele G, Tuinier R, Gibaud T, Stradner A, Schurtenberger P. A simple patchy colloid model for the phase behavior of lysozyme dispersions. J Chem Phys. 2008; 129(8):085102.doi: 10.1063/1.2951987 [PubMed: 19044852]
30. Chang J, Lenhoff AM, Sandler SI. Determination of fluid--solid transitions in model protein solutions using the histogram reweighting method and expanded ensemble simulations. J Chem Phys. 2004; 120(6):3003–3014. DOI: 10.1063/1.1638377 [PubMed: 15268448]
31. Fusco D, Charbonneau P. Competition between monomeric and dimeric crystals in schematic models for globular proteins. J Phys Chem B. 2014; 118(28):8034–8041. DOI: 10.1021/jp5011428 [PubMed: 24684539]
32. Staneva I, Frenkel D. The role of non-specific interactions in a patchy model of protein crystallization. J Chem Phys. 2015; 143(19):194511.doi: 10.1063/1.4935369 [PubMed: 26590546]
33. Fusco D, Headd JJ, De Simone A, Wang J, Charbonneau P. Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study. Soft Matter. 2014; 10(2):290–302. DOI: 10.1039/C3sm52175c [PubMed: 24489597]
34. Asherie N. Protein crystallization and phase diagrams. Methods. 2004; 34(3):266–272. [PubMed: 15325646]
35. Derewenda ZS, Vekilov PG. Entropy and surface engineering in protein crystallization. Acta Cryst D. 2006; 62(Pt 1):116–124. [PubMed: 16369101]

36. Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D. Solvent entropy contribution to the free energy of protein crystallization. Acta Cryst D. 2002; 58(Pt 10 Pt 1):1611–1616. [PubMed: 12351872]
37. Vekilov PG. Solvent entropy effects in the formation of protein solid phases. Methods Enzymol. 2003; 368:84–105. [PubMed: 14674270]
38. Yau ST, Petsev DN, Thomas BR, Vekilov PG. Molecular-level thermodynamic and kinetic parameters for the self-assembly of apoferritin molecules into crystals. J Mol Biol. 2000; 303(5): 667–678. [PubMed: 11061967]
39. Paunov VN, Kaler EW, Sandler SI, Petsev DN. A Model for Hydration Interactions between Apoferritin Molecules in Solution. J Colloid Interface Sci. 2001; 240(2):640–643. DOI: 10.1006/jcis.2001.7671 [PubMed: 11482976]
40. Gliko O, Neumaier N, Pan W, Haase I, Fischer M, Bacher A, Weinkauf S, Vekilov PG. A metastable prerequisite for the growth of lumazine synthase crystals. Journal of the American Chemical Society. 2005; 127(10):3433–3438. [PubMed: 15755162]
41. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng. 1989; 3(1):1–3. [PubMed: 2813338]
42. Tidor B, Karplus M. The contribution of vibrational entropy to molecular association. The dimerization of insulin. J Mol Biol. 1994; 238(3):405–414. [PubMed: 8176732]
43. Doye JPK. Inhibition of protein crystallization by evolutionary negative design. Physical Biology. 2004; 1:P9–P13. [PubMed: 16204814]
44. Ellis RJ. Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol. 2001; 11(1):114–119. S0959-440X(00)00172-X [pii]. [PubMed: 11179900]
45. Zorrilla S, Rivas G, Acuna AU, Lillo MP. Protein self-association in crowded protein solutions: a time-resolved fluorescence polarization study. Protein Sci. 2004; 13(11):2960–2969. [PubMed: 15459331]
46. Pal D, Chakrabarti P. Estimates of the loss of main-chain conformational entropy of different residues on protein folding. Proteins. 1999; 36(3):332–339. [PubMed: 10409826]
47. Chellgren BW, Creamer TP. Side-chain entropy effects on protein secondary structure formation. Proteins. 2006; 62(2):411–420. DOI: 10.1002/prot.20766 [PubMed: 16315271]
48. Lee J, Kim SH. Water polygons in high-resolution protein crystal structures. Protein Sci. 2009; 18(7):1370–1376. DOI: 10.1002/pro.162 [PubMed: 19551896]
49. Nakasako M. Water-protein interactions from high-resolution protein crystallography. Philos Trans R Soc Lond B Biol Sci. 2004; 359(1448):1191–1204. discussion 1204–1196. DOI: 10.1098/rstb.2004.1498 [PubMed: 15306376]
50. Ball P. Chemical physics: How to keep dry in water. Nature. 2003; 423(6935):25–26. DOI: 10.1038/423025a [PubMed: 12721610]
51. Pal SK, Zewail AH. Dynamics of water in biological recognition. Chem Rev. 2004; 104(4):2099–2123. DOI: 10.1021/cr020689l [PubMed: 15080722]
52. Dunitz JD. The entropic cost of bound water in crystals and biomolecules. Science. 1994; 264(5159):670.doi: 10.1126/science.264.5159.670 [PubMed: 17737951]
53. Vekilov PG, Vorontsova MA. Nucleation precursors in protein crystallization. Acta Crystallogr F Struct Biol Commun. 2014; 70(Pt 3):271–282. DOI: 10.1107/S2053230X14002386 [PubMed: 24598910]
54. Vekilov PG. Dense liquid precursor for the nucleation of ordered solid phases from solution. Crystal Growth & Design. 2004; 4(4):671–685.
55. Carugo O, Djinovic-Carugo K. How many packing contacts are observed in protein crystals? J Struct Biol. 2012; 180(1):96–100. DOI: 10.1016/j.jsb.2012.05.009 [PubMed: 22634724]
56. Wukovitz SW, Yeates TO. Why protein crystals favour some space-groups over others. Nat Struct Biol. 1995; 2(12):1062–1067. [PubMed: 8846217]
57. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci. 1998; 23(9):358–361. [PubMed: 9787643]
58. Ponstingl H, Kabir T, Thornton JM. Automatic inference of protein quaternary structure from crystals. J Appl Cryst. 2003; 36:1116–1122.

59. Elcock AH, McCammon JA. Calculation of weak protein-protein interactions: The pH dependence of the second virial coefficient. Biophysical Journal. 2001; 80(2):613–625. [PubMed: 11159430]

60. Elcock AH, McCammon JA. Identification of protein oligomerization states by analysis of interface conservation. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98(6):2990–2994. [PubMed: 11248019]

61. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372(3):774–797. [PubMed: 17681537]

62. Cieslik M, Derewenda ZS. The role of entropy and polarity in intermolecular contacts in protein crystals. Acta Cryst D. 2009; 65(Pt 5):500–509. DOI: 10.1107/S0907444909009500 [PubMed: 19390155]

63. Xu Q, Dunbrack RL Jr. The protein common interface database (ProtCID)--a comprehensive database of interactions of homologous proteins in multiple crystal forms. Nucleic Acids Res. 2011; 39(Database issue):D761–770. gkq1059 [pii]. DOI: 10.1093/nar/gkq1059 [PubMed: 21036862]

64. Rowe AJ. Ultra-weak reversible protein-protein interactions. Methods. 2011; 54(1):157–166. S1046-2023(11)00042-9 [pii]. DOI: 10.1016/j.ymeth.2011.02.006 [PubMed: 21338686]

65. Huxford T, Mishler D, Phelps CB, Huang DB, Sengchanthalangsy LL, Reeves R, Hughes CA, Komives EA, Ghosh G. Solvent exposed non-contacting amino acids play a critical role in NF-kappaB/IkappaBalpha complex formation. J Mol Biol. 2002; 324(4):587–597. S002228360201149X [pii]. [PubMed: 12460563]

66. van der Merwe PA, Davis SJ. Molecular interactions mediating T cell antigen recognition. Annu Rev Immunol. 2003; 21:659–684. [pii]. DOI: 10.1146/annurev.immunol.21.120601.141036120601.141036 [PubMed: 12615890]

67. Ceres P, Zlotnick A. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. Biochemistry. 2002; 41(39):11525–11531. bi0261645 [pii]. [PubMed: 12269796]

68. Wang Q, Zhuravleva A, Gierasch LM. Exploring weak, transient protein--protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. Biochemistry. 2011; 50(43):9225–9236. DOI: 10.1021/bi201287e [PubMed: 21942871]

69. Vaynberg J, Qin J. Weak protein-protein interactions as probed by NMR spectroscopy. Trends Biotechnol. 2006; 24(1):22–27. S0167-7799(05)00253-2 [pii]. DOI: 10.1016/j.tibtech.2005.09.006 [PubMed: 16216358]

70. Zhou HX, Rivas G, Minton AP. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys. 2008; 37:375–397. [PubMed: 18573087]

71. McGuffee SR, Elcock AH. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol. 2010; 6(3):e1000694.doi: 10.1371/journal.pcbi.1000694 [PubMed: 20221255]

72. Ellis RJ. Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci. 2001; 26(10):597–604. S0968-0004(01)01938-7 [pii]. [PubMed: 11590012]

73. Nooren IM, Thornton JM. Diversity of protein-protein interactions. Embo J. 2003; 22(14):3486–3492. [PubMed: 12853464]

74. Nooren IM, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. J Mol Biol. 2003; 325(5):991–1018. [PubMed: 12527304]

75. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol. 1987; 194(3):531–544. 0022-2836(87)90679-6 [pii]. [PubMed: 3041007]

76. Dikic I, Wakatsuki S, Walters KJ. Ubiquitin-binding domains - from structures to functions. Nat Rev Mol Cell Biol. 2009; 10(10):659–671. nrm2767 [pii]. DOI: 10.1038/nrm2767 [PubMed: 19773779]

77. Tabernero L, Evans BN, Tishmack PA, Van Etten RL, Stauffacher CV. The structure of the bovine protein tyrosine phosphatase dimer reveals a potential self-regulation mechanism. Biochemistry. 1999; 38(36):11651–11658. bi990381x [pii]. [PubMed: 10512620]

78. Akerud T, Thulin E, Van Etten RL, Akke M. Intramolecular dynamics of low molecular weight protein tyrosine phosphatase in monomer-dimer equilibrium studied by NMR: a model for changes

in dynamics upon target binding. J Mol Biol. 2002; 322(1):137–152. S0022283602007143 [pii]. [PubMed: 12215420]

79. Tung M, Gallagher DT. The Biomolecular Crystallization Database Version 4: expanded content and new features. Acta Cryst D. 2009; 65(Pt 1):18–23. DOI: 10.1107/S0907444908035440 [PubMed: 19153462]
80. Canaves JM, Page R, Wilson IA, Stevens RC. Protein biophysical properties that correlate with crystallization success in Thermotoga maritima: maximum clustering strategy for structural genomics. J Mol Biol. 2004; 344(4):977–991. [PubMed: 15544807]
81. Price WN 2nd, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J, Tong SN, Forouhar F, Swaminathan SS, Acton T, Xiao R, Luft JR, Lauricella A, DeTitta GT, Rost B, Montelione GT, Hunt JF. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol. 2009; 27(1):51–57. [PubMed: 19079241]
82. Fusco D, Barnum TJ, Bruno AE, Luft JR, Snell EH, Mukherjee S, Charbonneau P. Statistical analysis of crystallization database links protein physico-chemical features with crystallization mechanisms. PLoS One. 2014; 9(7):e101123.doi: 10.1371/journal.pone.0101123 [PubMed: 24988076]
83. Carugo O, Argos P. Protein-protein crystal-packing contacts. Protein Science. 1997; 6(10):2261–2263. [PubMed: 9336849]
84. Chen L, Oughtred R, Berman HM, Westbrook J. TargetDB: a target registration database for structural genomics projects. Bioinformatics. 2004; 20(16):2860–2862. bth300 [pii]. DOI: 10.1093/bioinformatics/bth300 [PubMed: 15130928]
85. Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, Kozlov G, Maxwell KL, Wu N, McIntosh LP, Gehring K, Kennedy MA, Davidson AR, Pai EF, Gerstein M, Edwards AM, Arrowsmith CH. Structural proteomics of an archaeon. Nature Structural Biology. 2000; 7(10):903–909. [PubMed: 11017201]
86. Canaves JM, Page R, Wilson IA, Stevens RC. Protein biophysical properties that correlate with crystallization success in Thermotoga maritima: Maximum clustering strategy for structural genomics. Journal of Molecular Biology. 2004; 344(4):977–991. DOI: 10.1016/J.Jmb.2004.09.076 [PubMed: 15544807]
87. Goh C-S, Lan N, Douglas SM, Wu B, Echols N, Smith A, Milburn D, Montelione GT, Zhao H, Gerstein M. Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. Journal of molecular biology. 2004; 336(1):115–130. [PubMed: 14741208]
88. Smialowski P, Schmidt T, Cox J, Kirschner A, Frishman D. Will my protein crystallize? A sequence-based predictor. Proteins-Structure Function and Bioinformatics. 2006; 62(2):343–355. DOI: 10.1002/Prot.20789
89. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. Bioinformatics. 2007; 23(24):3403–3405. DOI: 10.1093/bioinformatics/btm477 [PubMed: 17921170]
90. Overton IM, Padovani G, Girolami MA, Barton GJ. ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction. Bioinformatics. 2008; 24(7):901–907. btn055 [pii]. DOI: 10.1093/bioinformatics/btn055 [PubMed: 18285371]
91. Kurgan L, Razib AA, Aghakhani S, Dick S, Mizianty M, Jahandideh S. CRYSTALP2: sequence-based protein crystallization propensity prediction. Bmc Structural Biology. 2009; 9 Artn 50. doi: 10.1186/1472-6807-9-50
92. Mizianty MJ, Kurgan L. Meta prediction of protein crystallization propensity. Biochemical and Biophysical Research Communications. 2009; 390(1):10–15. DOI: 10.1016/J.Bbrc.2009.09.036 [PubMed: 19755114]
93. Babnigg G, Joachimiak A. Predicting protein crystallization propensity from protein sequence. Journal of structural and functional genomics. 2010; 11(1):71–80. [PubMed: 20177794]
94. Mizianty MJ, Kurgan L. Sequence-based prediction of protein crystallization, purification and production propensity. Bioinformatics. 2011; 27(13):I24–I33. DOI: 10.1093/Bioinformatics/Btr229 [PubMed: 21685077]

95. Jahandideh S, Jaroszewski L, Godzik A. Improving the chances of successful protein structure determination with a random forest classifier. Acta Cryst D. 2014; 70(Pt 3):627–635. S1399004713032070 [pii]. DOI: 10.1107/S1399004713032070 [PubMed: 24598732]

96. Jaroszewski L, Slabinski L, Wooley J, Deacon AM, Lesley SA, Wilson IA, Godzik A. Genome Pool Strategy for Structural Coverage of Protein Families. Structure. 2008; 16(11):1659–1667. DOI: 10.1016/J.Str.2008.08.018 [PubMed: 19000818]

97. Gabanyi MJ, Adams PD, Arnold K, Bordoli L, Carter LG, Flippen-Andersen J, Gifford L, Haas J, Kouranov A, McLaughlin WA, Micallef DI, Minor W, Shah R, Schwede T, Tao YP, Westbrook JD, Zimmerman M, Berman HM. The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods. J Struct Funct Genomics. 2011; 12(2):45–54. DOI: 10.1007/s10969-011-9106-2 [PubMed: 21472436]

98. Savitsky P, Bray J, Cooper CDO, Marsden BD, Mahajan P, Burgess-Brown NA, Gileadi O. High-throughput production of human proteins for crystallization: The SGC experience. Journal of Structural Biology. 2010; 172(1):3–13. DOI: 10.1016/J.Jsb.2010.06.008 [PubMed: 20541610]

99. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GV, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. J Struct Biol. 2010; 172(1):21–33. DOI: 10.1016/j.jsb.2010.07.011 [PubMed: 20688167]

100. Lee CK, Cheong C, Jeon YH. The N-terminal domain of human holocarboxylase synthetase facilitates biotinylation via direct interaction with the substrate protein. Febs Letters. 2010; 584(4):675–680. DOI: 10.1016/J.Febslet.2009.12.059 [PubMed: 20085763]

101. Oyenarte I, Lucas M, Gomez Garcia I, Martinez-Cruz LA. Purification, crystallization and preliminary crystallographic analysis of the CBS-domain protein MJ1004 from Methanocaldococcus jannaschii. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2011; 67(Pt 3):318–324. DOI: 10.1107/S1744309110053479

102. Gomez-Garcia I, Stuiver M, Ereno J, Oyenarte I, Corral-Rodriguez MA, Muller D, Martinez-Cruz LA. Purification, crystallization and preliminary crystallographic analysis of the CBS-domain pair of cyclin M2 (CNNM2). Acta Crystallographica Section F-Structural Biology and Crystallization Communications. 2012; 68:1198–1203. DOI: 10.1107/S1744309112035348

103. Derewenda ZS. Application of protein engineering to enhance crystallizability and improve crystal properties. Acta Crystallographica Section D-Biological Crystallography. 2010; 66:604–615. DOI: 10.1107/S090744491000644x

104. McPherson A, Nguyen C, Cudney R, Larson SB. The Role of Small Molecule Additives and Chemical Modification in Protein Crystallization. Crystal Growth & Design. 2011; 11(5):1469–1474.

105. Giedroc DP, Keating KM, Williams KR, Konigsberg WH, Coleman JE. Gene 32 protein, the single-stranded DNA binding protein from bacteriophage T4, is a zinc metalloprotein. Proc Natl Acad Sci U S A. 1986; 83(22):8452–8456. [PubMed: 3490667]

106. Dong A, Xu X, Edwards AM, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, Egorova O, Evdokimova E, Filippova E, Gu J, Guthrie J, Ignatchenko A, Joachimiak A, Klostermann N, Kim Y, Korniyenko Y, Minor W, Que Q, Savchenko A, Skarina T, Tan K, Yakunin A, Yee A, Yim V, Zhang R, Zheng H, Akutsu M, Arrowsmith C, Avvakumov GV, Bochkarev A, Dahlgren LG, Dhe-Paganon S, Dimov S, Dombrovski L, Finerty P Jr, Flodin S, Flores A, Graslund S, Hammerstrom M, Herman MD, Hong BS, Hui R, Johansson I, Liu Y, Nilsson M, Nedyalkova L, Nordlund P, Nyman T, Min J, Ouyang H, Park HW, Qi C, Rabeh W, Shen L, Shen Y, Sukumard D, Tempel W, Tong Y, Tresagues L, Vedadi M, Walker JR, Weigelt J, Welin M, Wu H, Xiao T, Zeng H, Zhu H. In situ proteolysis for protein crystallization and structure determination. Nat Methods. 2007; 4(12):1019–1021. [PubMed: 17982461]

107. Wernimont A, Edwards A. In situ proteolysis to generate crystals for structure determination: an update. PLoS One. 2009; 4(4):e5094. [PubMed: 19352432]

108. Huang YJ, Acton TB, Montelione GT. DisMeta: a meta server for construct design and optimization. Methods Mol Biol. 2014; 1091:3–16. DOI: 10.1007/978-1-62703-691-7_1 [PubMed: 24203321]

109. Malawski GA, Hillig RC, Monteclaro F, Eberspaecher U, Schmitz AA, Crusius K, Huber M, Egner U, Donner P, Muller-Tiemann B. Identifying protein construct variants with increased crystallization propensity--a case study. Protein Sci. 2006; 15(12):2718–2728. 15/12/2718 [pii]. DOI: 10.1110/ps.062491906 [PubMed: 17132859]
110. Ding HT, Ren H, Chen Q, Fang G, Li LF, Li R, Wang Z, Jia XY, Liang YH, Hu MH, Li Y, Luo JC, Gu XC, Su XD, Luo M, Lu SY. Parallel cloning, expression, purification and crystallization of human proteins for structural genomics. Acta Cryst D. 2002; 58(Pt 12):2102–2108. [PubMed: 12454471]
111. Quevillon-Cheruel S, Leulliot N, Gentils L, van Tilbeurgh H, Poupon A. Production and crystallization of protein domains: how useful are disorder predictions ? Curr Protein Pept Sci. 2007; 8(2):151–160. [PubMed: 17430196]
112. Page R. Strategies for improving crystallization success rates. Methods Mol Biol. 2008; 426:345–362. DOI: 10.1007/978-1-60327-058-8_22 [PubMed: 18542875]
113. Cohen SL, Ferredamare AR, Burley SK, Chait BT. Probing the Solution Structure of the DNA-Binding Protein Max by a Combination of Proteolysis and Mass-Spectrometry. Protein Science. 1995; 4(6):1088–1099. [PubMed: 7549873]
114. Hamuro Y, Coales SJ, Southern MR, Nemeth-Cawley JF, Stranz DD, Griffin PR. Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. J Biomol Tech. 2003; 14(3):171–182. [PubMed: 13678147]
115. Pantazatos D, Kim JS, Klock HE, Stevens RC, Wilson IA, Lesley SA, Woods VL Jr. Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. Proc Natl Acad Sci U S A. 2004; 101(3):751–756. [PubMed: 14715906]
116. Sharma S, Zheng H, Huang YPJ, Ertekin A, Hamuro Y, Rossi P, Tejero R, Acton TB, Xiao R, Jiang M, Zhao L, Ma LC, Swapna GVT, Aramini JM, Montelione GT. Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry. Proteins-Structure Function and Bioinformatics. 2009; 76(4):882–894. DOI: 10.1002/Prot.22394
117. Gray FL, Murai MJ, Grembecka J, Cierpicki T. Detection of disordered regions in globular proteins using (1)(3)C-detected NMR. Protein Sci. 2012; 21(12):1954–1960. DOI: 10.1002/pro.2174 [PubMed: 23047544]
118. Carugo O. Participation of protein sequence termini in crystal contacts. Protein Sci. 2011; 20(12):2121–2124. DOI: 10.1002/pro.690 [PubMed: 21739502]
119. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998; 393(6686):648–659. [PubMed: 9641677]
120. Kwong PD, Wyatt R, Desjardins E, Robinson J, Culp JS, Hellmig BD, Sweet RW, Sodroski J, Hendrickson WA. Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). J Biol Chem. 1999; 274(7):4115–4123. [PubMed: 9933605]
121. Neau DB, Gilbert NC, Bartlett SG, Dassey A, Newcomer ME. Improving protein crystal quality by selective removal of a Ca2+-dependent membrane-insertion loop. Acta Crystallogr F. 2007; 63:972–975. DOI: 10.1107/S1744309107050993
122. Schwartz TU, Walczak R, Blobel G. Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor beta subunit. Protein Sci. 2004; 13(10):2814–2818. [PubMed: 15340174]
123. Longenecker KL, Garrard SM, Sheffield PJ, Derewenda ZS. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Cryst D. 2001; 57(Pt 5):679–688. [PubMed: 11320308]
124. Mateja A, Devedjiev Y, Krowarsch D, Longenecker K, Dauter Z, Otlewski J, Derewenda ZS. The impact of Glu-->Ala and Glu-->Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 A resolution. Acta Cryst D. 2002; 58(Pt 12):1983–1991. [PubMed: 12454455]

125. Czepas J, Devedjiev Y, Krowarsch D, Derewenda U, Otlewski J, Derewenda ZS. The impact of Lys-->Arg surface mutations on the crystallization of the globular domain of RhoGDI. Acta Cryst D. 2004; 60(Pt 2):275–280. [PubMed: 14747703]

126. Garrard SM, Longenecker KL, Lewis ME, Sheffield PJ, Derewenda ZS. Expression, Purification, and Crystallization of the RGS-like Domain from the Rho Nucleotide Exchange Factor, PDZ-RhoGEF, Using the Surface Entropy Reduction Approach. Protein Expr Purif. 2001; 21(3):412–416. [PubMed: 11281715]

127. Longenecker KL, Lewis ME, Chikumi H, Gutkind JS, Derewenda ZS. Structure of the RGS-like domain from PDZ-RhoGEF: linking heterotrimeric g protein-coupled signaling to Rho GTPases. Structure (Camb). 2001; 9(7):559–569. [PubMed: 11470431]

128. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, Pfuetzner RA, Frey EA, Finlay BB, Miller SI, Strynadka NC. Structural characterization of the molecular platform for type III secretion system assembly. Nature. 2005; 435(7042):702–707. [PubMed: 15931226]

129. Fisher RD, Chung HY, Zhai Q, Robinson H, Sundquist WI, Hill CP. Structural and biochemical studies of ALIX/AIP1 and its role in retrovirus budding. Cell. 2007; 128(5):841–852. [PubMed: 17350572]

130. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, Sundquist WI, Hill CP, Yeager M. X-ray structures of the hexameric building block of the HIV capsid. Cell. 2009; 137(7):1282–1292. [PubMed: 19523676]

131. Pioletti M, Findeisen F, Hura GL, Minor DL Jr. Three-dimensional structure of the KChIP1-Kv4.3 T1 complex reveals a cross-shaped octamer. Nat Struct Mol Biol. 2006; 13(11):987–995. [PubMed: 17057713]

132. Ressl S, Terwisscha van Scheltinga AC, Vonrhein C, Ott V, Ziegler C. Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. Nature. 2009; 458(7234):47–52. nature07819 [pii]. DOI: 10.1038/nature07819 [PubMed: 19262666]

133. Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z. Protein crystallization by surface entropy reduction: optimization of the SER strategy. Acta Cryst D. 2007; 63(Pt 5):636–645. [PubMed: 17452789]

134. Conte LL, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol. 1999; 285(5):2177–2198. [PubMed: 9925793]

135. Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary-code interface. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(16):6632–6637. DOI: 10.1073/pnas.0700149104 [PubMed: 17420456]

136. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci. 2007; 16(8):1569–1576. [PubMed: 17656576]

137. Goldschmidt L, Eisenberg D, Derewenda ZS. Salvage or recovery of failed targets by mutagenesis to reduce surface entropy. Methods Mol Biol. 2014; 1140:201–209. DOI: 10.1007/978-1-4939-0354-2_16 [PubMed: 24590720]

138. Munshi S, Hall DL, Kornienko M, Darke PL, Kuo LC. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 A resolution. Acta Cryst D. 2003; 59(Pt 10):1725–1730. [PubMed: 14501110]

139. Bauman JD, Das K, Ho WC, Baweja M, Himmel DM, Clark AD Jr, Oren DA, Boyer PL, Hughes SH, Shatkin AJ, Arnold E. Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. Nucleic Acids Res. 2008; 36(15):5083–5092. [PubMed: 18676450]

140. Das K, Bauman JD, Clark AD Jr, Frenkel YV, Lewi PJ, Shatkin AJ, Hughes SH, Arnold E. High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: strategic flexibility explains potency against resistance mutations. Proc Natl Acad Sci U S A. 2008; 105(5):1466–1471. [PubMed: 18230722]

141. Frey KM, Puleo DE, Spasov KA, Bollini M, Jorgensen WL, Anderson KS. Structure-based evaluation of non-nucleoside inhibitors with improved potency and solubility that target HIV reverse transcriptase variants. J Med Chem. 2015; 58(6):2737–2745. DOI: 10.1021/jm501908a [PubMed: 25700160]

142. Gray WT, Frey KM, Laskey SB, Mislak AC, Spasov KA, Lee WG, Bollini M, Siliciano RF, Jorgensen WL, Anderson KS. Potent Inhibitors Active against HIV Reverse Transcriptase with K101P, a Mutation Conferring Rilpivirine Resistance. ACS Med Chem Lett. 2015; 6(10):1075–1079. DOI: 10.1021/acsmedchemlett.5b00254 [PubMed: 26487915]

143. Lee WG, Frey KM, Gallardo-Macias R, Spasov KA, Chan AH, Anderson KS, Jorgensen WL. Discovery and crystallography of bicyclic arylaminoazines as potent inhibitors of HIV-1 reverse transcriptase. Bioorg Med Chem Lett. 2015; 25(21):4824–4827. DOI: 10.1016/j.bmcl.2015.06.074 [PubMed: 26166629]

144. Yang W, Fucini RV, Fahr BT, Randal M, Lind KE, Lam MB, Lu W, Lu Y, Cary DR, Romanowski MJ, Colussi D, Pietrak B, Allison TJ, Munshi SK, Penny DM, Pham P, Sun J, Thomas AE, Wilkinson JM, Jacobs JW, McDowell RS, Ballinger MD. Fragment-based discovery of nonpeptidic BACE-1 inhibitors using tethering. Biochemistry. 2009; 48(21):4488–4496. [PubMed: 19284778]

145. Barrow JC, Stauffer SR, Rittle KE, Ngo PL, Yang Z, Selnick HG, Graham SL, Munshi S, McGaughey GB, Holloway MK, Simon AJ, Price EA, Sankaranarayanan S, Colussi D, Tugusheva K, Lai MT, Espeseth AS, Xu M, Huang Q, Wolfe A, Pietrak B, Zuck P, Levorse DA, Hazuda D, Vacca JP. Discovery and X-ray crystallographic analysis of a spiropiperidine iminohydantoin inhibitor of beta-secretase. J Med Chem. 2008; 51(20):6259–6262. [PubMed: 18811140]

146. McGaughey GB, Colussi D, Graham SL, Lai MT, Munshi SK, Nantermet PG, Pietrak B, Rajapakse HA, Selnick HG, Stauffer SR, Holloway MK. Beta-secretase (BACE-1) inhibitors: accounting for 10s loop flexibility using rigid active sites. Bioorg Med Chem Lett. 2007; 17(4):1117–1121. [PubMed: 17112725]

147. Stauffer SR, Stanton MG, Gregro AR, Steinbeiser MA, Shaffer JR, Nantermet PG, Barrow JC, Rittle KE, Collusi D, Espeseth AS, Lai MT, Pietrak BL, Holloway MK, McGaughey GB, Munshi SK, Hochman JH, Simon AJ, Selnick HG, Graham SL, Vacca JP. Discovery and SAR of isonicotinamide BACE-1 inhibitors that bind beta-secretase in a N-terminal 10s-loop down conformation. Bioorg Med Chem Lett. 2007; 17(6):1788–1792. [PubMed: 17257835]

148. Lindsley SR, Moore KP, Rajapakse HA, Selnick HG, Young MB, Zhu H, Munshi S, Kuo L, McGaughey GB, Colussi D, Crouthamel MC, Lai MT, Pietrak B, Price EA, Sankaranarayanan S, Simon AJ, Seabrook GR, Hazuda DJ, Pudvah NT, Hochman JH, Graham SL, Vacca JP, Nantermet PG. Design, synthesis, and SAR of macrocyclic tertiary carbinamine BACE-1 inhibitors. Bioorg Med Chem Lett. 2007; 17(14):4057–4061. [PubMed: 17482814]

149. Moore KP, Zhu H, Rajapakse HA, McGaughey GB, Colussi D, Price EA, Sankaranarayanan S, Simon AJ, Pudvah NT, Hochman JH, Allison T, Munshi SK, Graham SL, Vacca JP, Nantermet PG. Strategies toward improving the brain penetration of macrocyclic tertiary carbinamine BACE-1 inhibitors. Bioorg Med Chem Lett. 2007; 17(21):5831–5835. [PubMed: 17827011]

150. Stachel SJ, Coburn CA, Steele TG, Crouthamel MC, Pietrak BL, Lai MT, Holloway MK, Munshi SK, Graham SL, Vacca JP. Conformationally biased P3 amide replacements of beta-secretase inhibitors. Bioorg Med Chem Lett. 2006; 16(3):641–644. [PubMed: 16263281]

151. Rajapakse HA, Nantermet PG, Selnick HG, Munshi S, McGaughey GB, Lindsley SR, Young MB, Lai MT, Espeseth AS, Shi XP, Colussi D, Pietrak B, Crouthamel MC, Tugusheva K, Huang Q, Xu M, Simon AJ, Kuo L, Hazuda DJ, Graham S, Vacca JP. Discovery of oxadiazoyl tertiary carbinamine inhibitors of beta-secretase (BACE-1). J Med Chem. 2006; 49(25):7270–7273. [PubMed: 17149856]

152. Coburn CA, Stachel SJ, Jones KG, Steele TG, Rush DM, DiMuzio J, Pietrak BL, Lai MT, Huang Q, Lineberger J, Jin L, Munshi S, Katharine Holloway M, Espeseth A, Simon A, Hazuda D, Graham SL, Vacca JP. BACE-1 inhibition by a series of psi[CH2NH] reduced amide isosteres. Bioorg Med Chem Lett. 2006; 16(14):3635–3638. [PubMed: 16690314]

153. Coburn CA, Stachel SJ, Li YM, Rush DM, Steele TG, Chen-Dodson E, Holloway MK, Xu M, Huang Q, Lai MT, DiMuzio J, Crouthamel MC, Shi XP, Sardana V, Chen Z, Munshi S, Kuo L, Makara GM, Annis DA, Tadikonda PK, Nash HM, Vacca JP, Wang T. Identification of a small molecule nonpeptide active site beta-secretase inhibitor that displays a nontraditional binding mode for aspartyl proteases. J Med Chem. 2004; 47(25):6117–6119. [PubMed: 15566281]

154. Pedro-Rosa L, Buckner FS, Ranade RM, Eberhart C, Madoux F, Gillespie JR, Koh CY, Brown S, Lohse J, Verlinde CL, Fan E, Bannister T, Scampavia L, Hol WG, Spicer T, Hodder P. Identification of potent inhibitors of the Trypanosoma brucei methionyl-tRNA synthetase via high-throughput orthogonal screening. J Biomol Screen. 2015; 20(1):122–130. DOI: 10.1177/1087057114548832 [PubMed: 25163684]

155. Koh CY, Kim JE, Wetzel AB, de van der Schueren WJ, Shibata S, Ranade RM, Liu J, Zhang Z, Gillespie JR, Buckner FS, Verlinde CL, Fan E, Hol WG. Structures of Trypanosoma brucei methionyl-tRNA synthetase with urea-based inhibitors provide guidance for drug design against sleeping sickness. PLoS neglected tropical diseases. 2014; 8(4):e2775.doi: 10.1371/journal.pntd.0002775 [PubMed: 24743796]

156. Holden JK, Li H, Jing Q, Kang S, Richo J, Silverman RB, Poulos TL. Structural and biological studies on bacterial nitric oxide synthase inhibitors. Proc Natl Acad Sci U S A. 2013; 110(45): 18127–18131. DOI: 10.1073/pnas.1314080110 [PubMed: 24145412]

157. Jing Q, Li H, Fang J, Roman LJ, Martasek P, Poulos TL, Silverman RB. In search of potent and selective inhibitors of neuronal nitric oxide synthase with more simple structures. Bioorg Med Chem. 2013; 21(17):5323–5331. DOI: 10.1016/j.bmc.2013.06.014 [PubMed: 23867386]

158. Huang H, Li H, Martasek P, Roman LJ, Poulos TL, Silverman RB. Structure-guided design of selective inhibitors of neuronal nitric oxide synthase. J Med Chem. 2013; 56(7):3024–3032. DOI: 10.1021/jm4000984 [PubMed: 23451760]

159. Yang Z, Misner B, Ji H, Poulos TL, Silverman RB, Meyskens FL, Yang S. Targeting nitric oxide signaling with nNOS inhibitors as a novel strategy for the therapy and prevention of human melanoma. Antioxid Redox Signal. 2013; 19(5):433–447. DOI: 10.1089/ars.2012.4563 [PubMed: 23199242]

160. Hanan EJ, Baumgardner M, Bryan MC, Chen Y, Eigenbrot C, Fan P, Gu XH, La H, Malek S, Purkey HE, Schaefer G, Schmidt S, Sideris S, Yen I, Yu C, Heffron TP. 4-Aminoindazolyl-dihydrofuro[3,4-d]pyrimidines as non-covalent inhibitors of mutant epidermal growth factor receptor tyrosine kinase. Bioorg Med Chem Lett. 2016; 26(2):534–539. DOI: 10.1016/j.bmcl.2015.11.078 [PubMed: 26639762]

161. Heald R, Bowman KK, Bryan MC, Burdick D, Chan B, Chan E, Chen Y, Clausen S, Dominguez-Fernandez B, Eigenbrot C, Elliott R, Hanan EJ, Jackson P, Knight J, La H, Lainchbury M, Malek S, Mann S, Merchant M, Mortara K, Purkey H, Schaefer G, Schmidt S, Seward E, Sideris S, Shao L, Wang S, Yeap K, Yen I, Yu C, Heffron TP. Noncovalent Mutant Selective Epidermal Growth Factor Receptor Inhibitors: A Lead Optimization Case Study. J Med Chem. 2015; 58(22):8877–8895. DOI: 10.1021/acs.jmedchem.5b01412 [PubMed: 26455919]

162. Hanan EJ, Eigenbrot C, Bryan MC, Burdick DJ, Chan BK, Chen Y, Dotson J, Heald RA, Jackson PS, La H, Lainchbury MD, Malek S, Purkey HE, Schaefer G, Schmidt S, Seward EM, Sideris S, Tam C, Wang S, Yeap SK, Yen I, Yin J, Yu C, Zilberleyb I, Heffron TP. Discovery of selective and noncovalent diaminopyrimidine-based inhibitors of epidermal growth factor receptor containing the T790M resistance mutation. J Med Chem. 2014; 57(23):10176–10191. DOI: 10.1021/jm501578n [PubMed: 25383627]

163. Means GE, Feeney RE. Chemical modifications of proteins: history and applications. Bioconjug Chem. 1990; 1(1):2–12. [PubMed: 2095202]

164. Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM. Three-dimensional structure of myosin subfragment-1: a molecular motor. Science. 1993; 261(5117):50–58. [PubMed: 8316857]

165. Rayment I. Reductive alkylation of lysine residues to alter crystallization properties of proteins. Methods Enzymol. 1997; 276:171–179.

166. Tan K, Kim Y, Hatzos-Skintges C, Chang C, Cuff M, Chhor G, Osipiuk J, Michalska K, Nocek B, An H, Babnigg G, Bigelow L, Joachimiak G, Li H, Mack J, Makowska-Grzyska M, Maltseva N, Mulligan R, Tesar C, Zhou M, Joachimiak A. Salvage of failed protein targets by reductive alkylation. Methods Mol Biol. 2014; 1140:189–200. DOI: 10.1007/978-1-4939-0354-2_15 [PubMed: 24590719]

167. Walter TS, Meier C, Assenberg R, Au KF, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM. Lysine methylation as a routine rescue strategy for protein crystallization. Structure. 2006; 14(11):1617–1622. [PubMed: 17098187]

168. Kim Y, Quartey P, Li H, Volkart L, Hatzos C, Chang C, Nocek B, Cuff M, Osipiuk J, Tan K, Fan Y, Bigelow L, Maltseva N, Wu R, Borovilos M, Duggan E, Zhou M, Binkowski TA, Zhang RG, Joachimiak A. Large-scale evaluation of protein reductive methylation for improving protein crystallization. Nat Methods. 2008; 5(10):853–854. DOI: 10.1038/nmeth1008-853 [PubMed: 18825126]

169. Fan Y, Joachimiak A. Enhanced crystal packing due to solvent reorganization through reductive methylation of lysine residues in oxidoreductase from Streptococcus pneumoniae. J Struct Funct Genomics. 2010; 11(2):101–111. DOI: 10.1007/s10969-010-9079-6 [PubMed: 20127187]

170. Sledz P, Zheng H, Murzyn K, Chruszcz M, Zimmerman MD, Chordia MD, Joachimiak A, Minor W. New surface contacts formed upon reductive lysine methylation: improving the probability of protein crystallization. Protein Sci. 2010; 19(7):1395–1404. DOI: 10.1002/pro.420 [PubMed: 20506323]

171. Shaw N, Cheng C, Tempel W, Chang J, Ng J, Wang XY, Perrett S, Rose J, Rao Z, Wang BC, Liu ZJ. (NZ)CH…O contacts assist crystallization of a ParB-like nuclease. BMC Struct Biol. 2007; 7:46.doi: 10.1186/1472-6807-7-46 [PubMed: 17617922]

172. Means GE. Reductive alkylation of amino groups. Methods Enzymol. 1977; 47:469–478. [PubMed: 927199]

173. McPherson A, Cudney B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. J Struct Biol. 2006; 156(3):387–406. [PubMed: 17101277]

174. Carugo O, Djinovic-Carugo K. Packing bridges in protein crystal structures. J Appl Cryst. 2014; 47:458–461.

175. Lagautriere T, Bashiri G, Baker EN. Use of a “silver bullet” to resolve crystal lattice dislocation disorder: a cobalamin complex of Delta1-pyrroline-5-carboxylate dehydrogenase from Mycobacterium tuberculosis. J Struct Biol. 2015; 189(2):153–157. DOI: 10.1016/j.jsb.2014.12.007 [PubMed: 25557497]

176. Bijelic A, Molitor C, Mauracher SG, Al-Oweini R, Kortz U, Rompel A. Hen egg-white lysozyme crystallisation: protein stacking and structure stability enhanced by a Tellurium(VI)-centred polyoxotungstate. Chembiochem. 2015; 16(2):233–241. DOI: 10.1002/cbic.201402597 [PubMed: 25521080]

177. Qiu X, Janson CA. Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions. Acta Cryst D. 2004; 60(Pt 9):1545–1554. [PubMed: 15333924]

178. Laganowsky A, Zhao M, Soriaga AB, Sawaya MR, Cascio D, Yeates TO. An approach to crystallizing proteins by metal-mediated synthetic symmetrization. Protein Sci. 2011; 20(11):1876–1890. DOI: 10.1002/pro.727 [PubMed: 21898649]

179. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang RG, Lezondra L, Joachimiak A, Derewenda ZS. The Structure and Ligand Binding Properties of the B. subtilis YkoF Gene Product, a Member of a Novel Family of Thiamin/HMP-binding Proteins. J Mol Biol. 2004; 343(2):395–406. [PubMed: 15451668]

180. Roosen-Runge F, Zhang F, Schreiber F, Roth R. Ion-activated attractive patches as a mechanism for controlled protein interactions. Sci Rep. 2014; 4:7016.doi: 10.1038/srep07016 [PubMed: 25388788]

181. Kuge M, Fujii Y, Shimizu T, Hirose F, Matsukage A, Hakoshima T. Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. Protein Sci. 1997; 6(8):1783–1786. [PubMed: 9260294]

182. Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci. 2008; 17(12):2070–2079. [PubMed: 18780816]

183. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. Protein Sci. 2003; 12(7):1313–1322. [PubMed: 12824478]

184. Center RJ, Kobe B, Wilson KA, Teh T, Howlett GJ, Kemp BE, Poumbourios P. Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. Protein Sci. 1998; 7(7):1612–1619. [PubMed: 9684894]

185. Monne M, Han L, Schwend T, Burendahl S, Jovine L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature. 2008; 456(7222):653–657. [PubMed: 19052627]

186. Wiltzius JJ, Sievers SA, Sawaya MR, Eisenberg D. Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process. Protein Sci. 2009; 18(7):1521–1530. [PubMed: 19475663]

187. Ke A, Wolberger C. Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. Protein Sci. 2003; 12(2): 306–312. [PubMed: 12538894]

188. Prive GG, Verner GE, Weitzman C, Zen KH, Eisenberg D, Kaback HR. Fusion proteins as tools for crystallization: the lactose permease from Escherichia coli. Acta Cryst D. 1994; 50(Pt 4):375–379. [PubMed: 15299388]

189. Engel CK, Chen L, Prive GG. Insertion of carrier proteins into hydrophilic loops of the Escherichia coli lactose permease. Biochim Biophys Acta. 2002; 1564(1):38–46. [PubMed: 12100994]

190. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. Science. 2007; 318(5854):1258–1265. [PubMed: 17962520]

191. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. GPCR engineering yields high-resolution structural insights into beta(2)-adrenergic receptor function. Science. 2007; 318(5854):1266–1273. [PubMed: 17962519]

192. Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, Kunken J, Xu F, Cherezov V, Hanson MA, Stevens RC. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. Structure. 2012; 20(6):967–976. DOI: 10.1016/j.str.2012.04.010 [PubMed: 22681902]

193. Kovari LC, Momany C, Rossmann MG. The use of antibody fragments for crystallization and structure determinations. Structure. 1995; 3(12):1291–1293. [PubMed: 8747455]

194. Hunte C, Michel H. Crystallisation of membrane proteins mediated by antibody fragments. Curr Opin Struct Biol. 2002; 12(4):503–508. [PubMed: 12163074]

195. Prongay AJ, Smith TJ, Rossmann MG, Ehrlich LS, Carter CA, McClure J. Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. Proc Natl Acad Sci U S A. 1990; 87(24):9980–9984. [PubMed: 2124709]

196. Ostermeier C, Iwata S, Ludwig B, Michel H. F-V Fragment Mediated Crystallization of the Membrane-Protein Bacterial Cytochrome-C-Oxidase. Nature Structural Biology. 1995; 2(10): 842–846. [PubMed: 7552705]

197. Jiang Y, Lee A, Chen J, Ruta V, Cadene M, Chait BT, MacKinnon R. X-ray structure of a voltage-dependent K+ channel. Nature. 2003; 423(6935):33–41. [PubMed: 12721618]

198. Dutzler R, Campbell EB, MacKinnon R. Gating the selectivity filter in ClC chloride channels. Science. 2003; 300(5616):108–112. [PubMed: 12649487]

199. Lee SY, Lee A, Chen J, MacKinnon R. Structure of the KvAP voltage-dependent K+ channel and its dependence on the lipid membrane. Proc Natl Acad Sci U S A. 2005; 102(43):15441–15446. [PubMed: 16223877]

200. Karpusas M, Lucci J, Ferrant J, Benjamin C, Taylor FR, Strauch K, Garber E, Hsu YM. Structure of CD40 ligand in complex with the Fab fragment of a neutralizing humanized antibody. Structure. 2001; 9(4):321–329. [PubMed: 11525169]

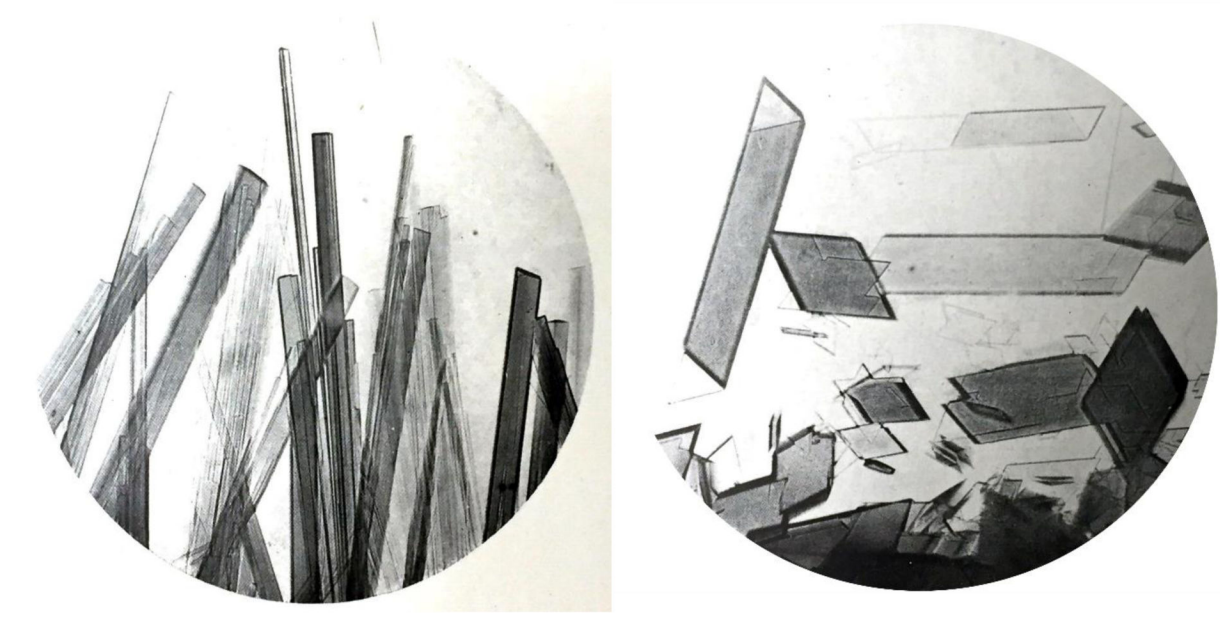
201. Nettleship JE, Ren J, Rahman N, Berrow NS, Hatherley D, Barclay AN, Owens RJ. A pipeline for the production of antibody fragments for structural studies using transient expression in HEK 293T cells. Protein Expr Purif. 2008; 62(1):83–89. [PubMed: 18662785]

202. Lee CV, Liang WC, Dennis MS, Eigenbrot C, Sidhu SS, Fuh G. High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. J Mol Biol. 2004; 340(5):1073–1093. [PubMed: 15236968]

203. Hornsby M, Paduch M, Miersch S, Saaf A, Matsuguchi T, Lee B, Wypisniak K, Doak A, King D, Usatyuk S, Perry K, Lu V, Thomas W, Luke J, Goodman J, Hoey RJ, Lai D, Griffin C, Li Z, Vizeacoumar FJ, Dong D, Campbell E, Anderson S, Zhong N, Graslund S, Koide S, Moffat J, Sidhu S, Kossiakoff A, Wells J. A High Through-put Platform for Recombinant Antibodies to Folded Proteins. Mol Cell Proteomics. 2015; 14(10):2833–2847. DOI: 10.1074/mcp.O115.052209 [PubMed: 26290498]
204. Lipovsek D, Pluckthun A. In-vitro protein evolution by ribosome display and mRNA display. J Immunol Methods. 2004; 290(1–2):51–67. [PubMed: 15261571]
205. Stafford RL, Matsumoto ML, Yin G, Cai Q, Fung JJ, Stephenson H, Gill A, You M, Lin SH, Wang WD, Masikat MR, Li X, Penta K, Steiner AR, Baliga R, Murray CJ, Thanos CD, Hallam TJ, Sato AK. In vitro Fab display: a cell-free system for IgG discovery. Protein Eng Des Sel. 2014; 27(4):97–109. DOI: 10.1093/protein/gzu002 [PubMed: 24586053]
206. Hawkins RE, Russell SJ, Winter G. Selection of Phage Antibodies by Binding-Affinity - Mimicking Affinity Maturation. Journal of Molecular Biology. 1992; 226(3):889–896. [PubMed: 1507232]
207. Koide S. Engineering of recombinant crystallization chaperones. Curr Opin Struct Biol. 2009; 19(4):449–457. [PubMed: 19477632]
208. Fellouse FA, Wiesmann C, Sidhu SS. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. Proc Natl Acad Sci U S A. 2004; 101(34): 12467–12472. [PubMed: 15306681]
209. Stuwe T, Bley CJ, Thierbach K, Petrovic S, Schilbach S, Mayo DJ, Perriches T, Rundlet EJ, Jeon YE, Collins LN, Huber FM, Lin DH, Paduch M, Koide A, Lu V, Fischer J, Hurt E, Koide S, Kossiakoff AA, Hoelz A. Architecture of the fungal nuclear pore inner ring complex. Science. 2015; 350(6256):56–64. DOI: 10.1126/science.aac9176 [PubMed: 26316600]
210. Nocula-Lugowska M, Lugowski M, Salgia R, Kossiakoff AA. Engineering Synthetic Antibody Inhibitors Specific for LD2 or LD4 Motifs of Paxillin. J Mol Biol. 2015; 427(15):2532–2547. DOI: 10.1016/j.jmb.2015.06.004 [PubMed: 26087144]
211. Mateja A, Paduch M, Chang HY, Szydlowska A, Kossiakoff AA, Hegde RS, Keenan RJ. Protein targeting. Structure of the Get3 targeting factor in complex with its membrane protein cargo. Science. 2015; 347(6226):1152–1155. DOI: 10.1126/science.1261671 [PubMed: 25745174]
212. Rizk SS, Kouadio JL, Szymborska A, Duguid EM, Mukherjee S, Zheng J, Clevenger CV, Kossiakoff AA. Engineering synthetic antibody binders for allosteric inhibition of prolactin receptor signaling. Cell Commun Signal. 2015; 13:1.doi: 10.1186/s12964-014-0080-8 [PubMed: 25589173]
213. Koide A, Tereshko V, Uysal S, Margalef K, Kossiakoff AA, Koide S. Exploring the capacity of minimalist protein interfaces: interface energetics and affinity maturation to picomolar K-D of a single-domain antibody with a flat paratope. Journal of Molecular Biology. 2007; 373(4):941–953. DOI: 10.1016/j.jmb.2007.08.027 [PubMed: 17888451]
214. Lam AY, Pardon E, Korotkov KV, Hol WG, Steyaert J. Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus. J Struct Biol. 2009; 166(1):8–15. [PubMed: 19118632]
215. Korotkov KV, Pardon E, Steyaert J, Hol WG. Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. Structure. 2009; 17(2): 255–265. [PubMed: 19217396]
216. Gilbreth RN, Esaki K, Koide A, Sidhu SS, Koide S. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. Journal of Molecular Biology. 2008; 381(2):407–418. DOI: 10.1016/j.jmb.2008.06.014 [PubMed: 18602117]
217. Sennhauser G, Grutter MG. Chaperone-assisted crystallography with DARPins. Structure. 2008; 16(10):1443–1453. [PubMed: 18940601]
218. Batyuk A, Wu Y, Honegger A, Heberling MM, Pluckthun A. DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. J Mol Biol. 2016; 428(8):1574–1588. DOI: 10.1016/j.jmb.2016.03.002 [PubMed: 26975886]
219. Bandeiras TM, Hillig RC, Matias PM, Eberspaecher U, Fanghanel J, Thomaz M, Miranda S, Crusius K, Puetter V, Amstutz P, Gulotti-Georgieva M, Binz HK, Holz C, Schmitz AAP, Lang C,

Donner P, Egner U, Carrondo MA, Muller-Tiemann B. Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin. Acta Crystallographica Section D-Biological Crystallography. 2008; 64:339–353.

220. Sennhauser G, Amstutz P, Briand C, Storchenegger O, Grutter MG. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. Plos Biol. 2007; 5(1):106–113.
221. Veesler D, Dreier B, Blangy S, Lichiere J, Tremblay D, Moineau S, Spinelli S, Tegoni M, Pluckthun A, Campanacci V, Cambillau C. Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARPin and camelid VHH binding mode. J Biol Chem. 2009; 284(44):30718–30726. [PubMed: 19740746]
222. Thorsen TS, Matt R, Weis WI, Kobilka BK. Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallogenesis. Structure. 2014; 22(11):1657–1664. DOI: 10.1016/j.str.2014.08.022 [PubMed: 25450769]
223. Notti RQ, Bhattacharya S, Lilic M, Stebbins CE. A common assembly module in injectisome and flagellar type III secretion sorting platforms. Nat Commun. 2015; 6:7125.doi: 10.1038/ncomms8125 [PubMed: 25994170]
224. Ullah H, Scappini EL, Moon AF, Williams LV, Armstrong DL, Pedersen LC. Structure of a signal transduction regulator, RACK1, from Arabidopsis thaliana. Protein Sci. 2008; 17(10):1771–1780. [PubMed: 18715992]
225. Lanci CJ, MacDermaid CM, Kang SG, Acharya R, North B, Yang X, Qiu XJ, DeGrado WF, Saven JG. Computational design of a protein crystal. Proc Natl Acad Sci U S A. 2012; 109(19):7304–7309. DOI: 10.1073/pnas.1112595109 [PubMed: 22538812]
226. Hsieh FL, Chang TH, Ko TP, Wang AH. Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol. 2011; 155(3):1079–1090. pp. 110.168799 [pii]. DOI: 10.1104/pp.110.168799 [PubMed: 21220764]
227. Vorontsova MA, Maes D, Vekilov PG. Recent advances in the understanding of two-step nucleation of protein crystals. Faraday Discuss. 2015; 179:27–40. DOI: 10.1039/c4fd00217b [PubMed: 25859918]



A B

Fig. 1. Two different forms of hemoglobin crystals

(A) Crystals obtained from hemoglobin of the mule, and (B) those from the hemoglobin of Indian antelope (*antelope cervicapra*). Both figures from Reichert ET, Brown AP (1909) The differentiation and specificity of corresponding proteins and other vital substances in relation to biological classification and organic evolution: the crystallography of haemoglobin. Carnegie Instituion of Washington, Washington, D.C.

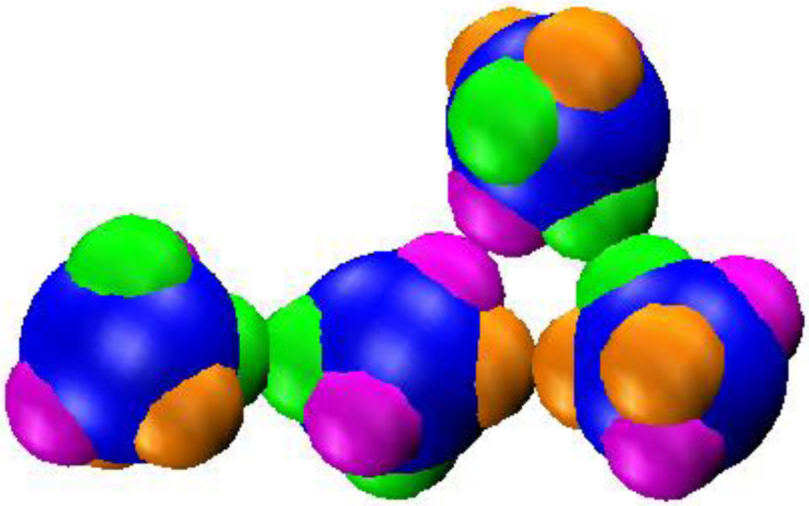


Fig. 2. The 'patchy model' of proteins and their interactions

The blue spheres are proteins on which each pair of patches corresponds to the crystal interface of the same color. From: Fusco *et al.* (2014) Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study. Soft Matter 10 (2):290–302.

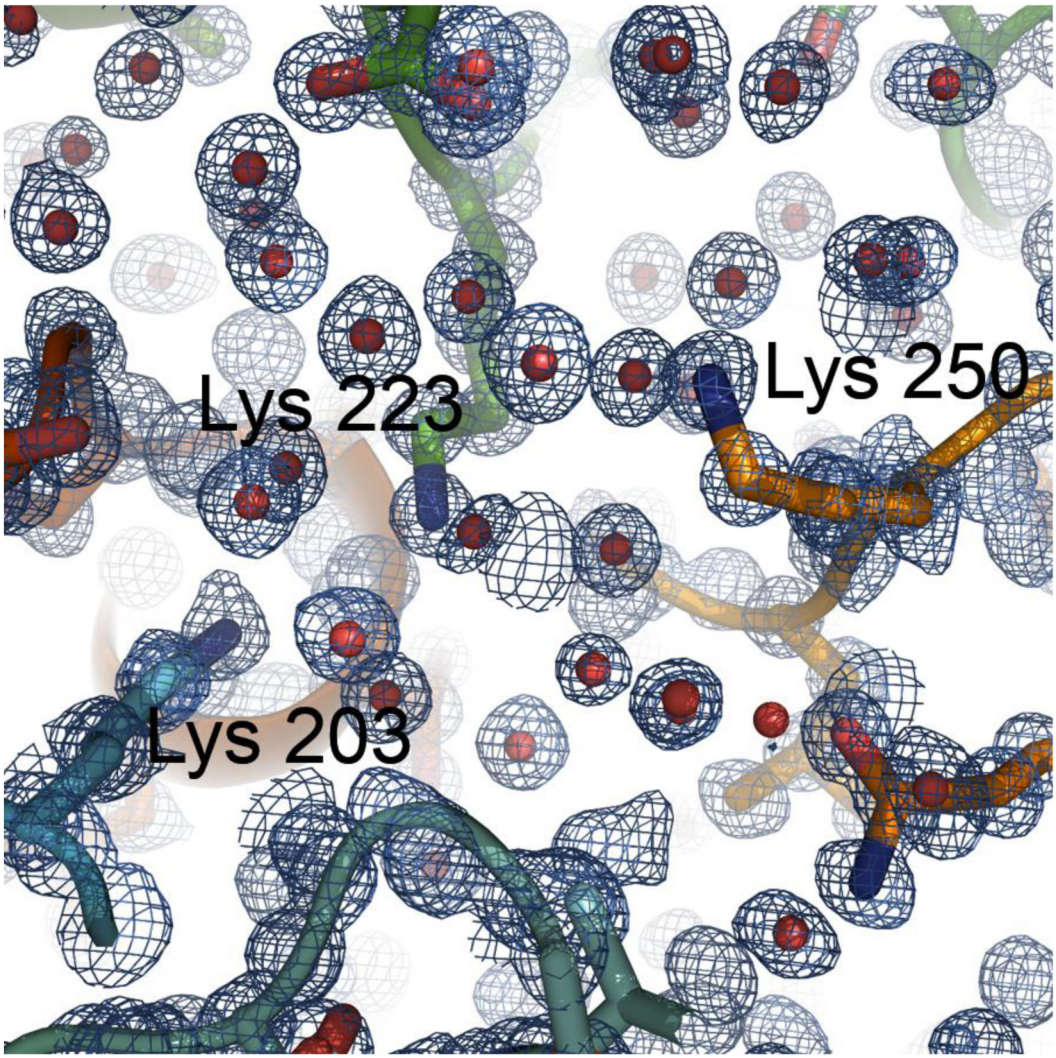


Fig. 3. Lysines sequestered at a crystal contacts

Three lysine residues, each from a different molecule, are sequestered at a crystal contact with a number or ordered water molecules; PDB code 1R6J, 0.72 Å resolution structure of the PDZ2 domain of synetenin.

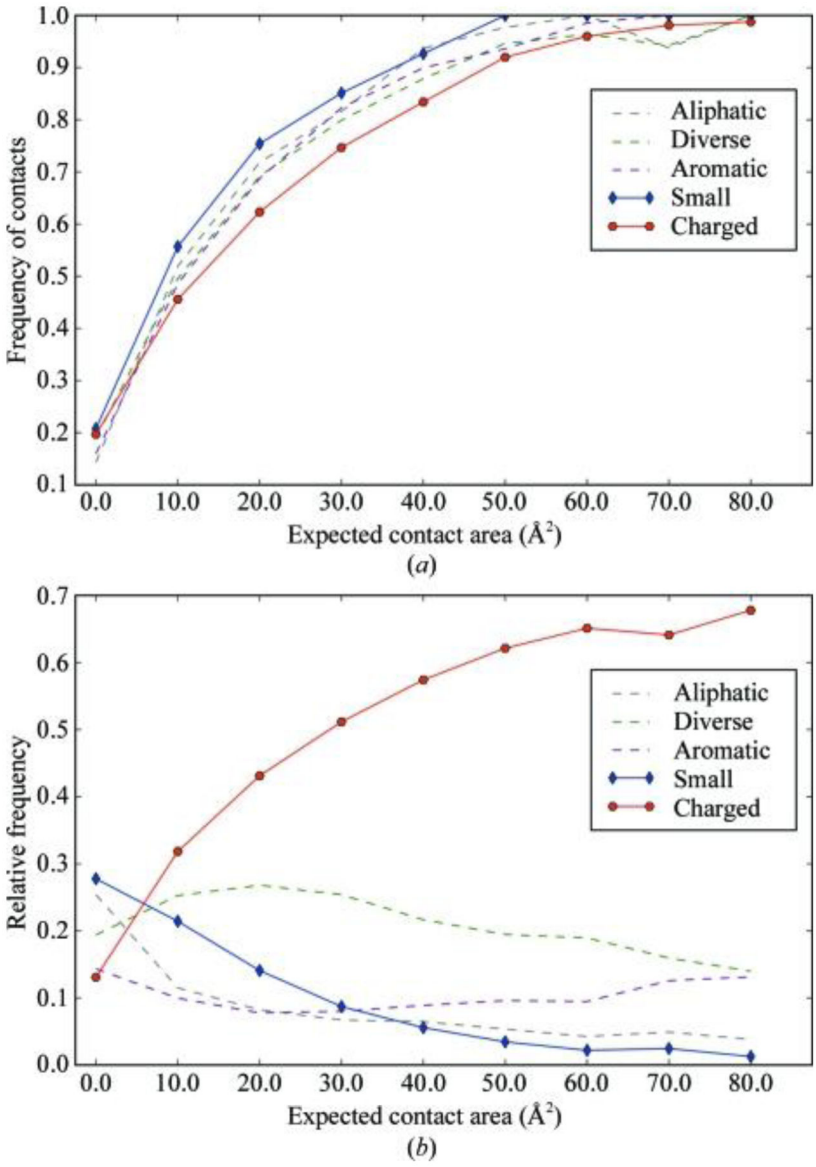


Fig. 4. Non-random composition of crystal contact surfaces

(*a*) Relative frequencies of five categories of amino acids, *i.e.* aliphatic (Val, Leu, Ile), aromatic (Trp, Phe, Tyr, His), small (Ala, Gly, Ser, Thr, Cys), charged (Lys, Arg, Glu, Asp) and other (Asn, Gln, Met, Pro), binned as a function of rECA. The relative frequency in each bin is the ratio of the number of residues of a given type to the total number of residues. (*b*) The fraction of residues involved in crystal contacts as a function of rECA plotted for the five categories as defined above. rECA is the residue expected contact area. For details see Cieslik & Derewenda (2009) The role of entropy and polarity in intermolecular contacts in protein crystals. Acta Cryst D 65.

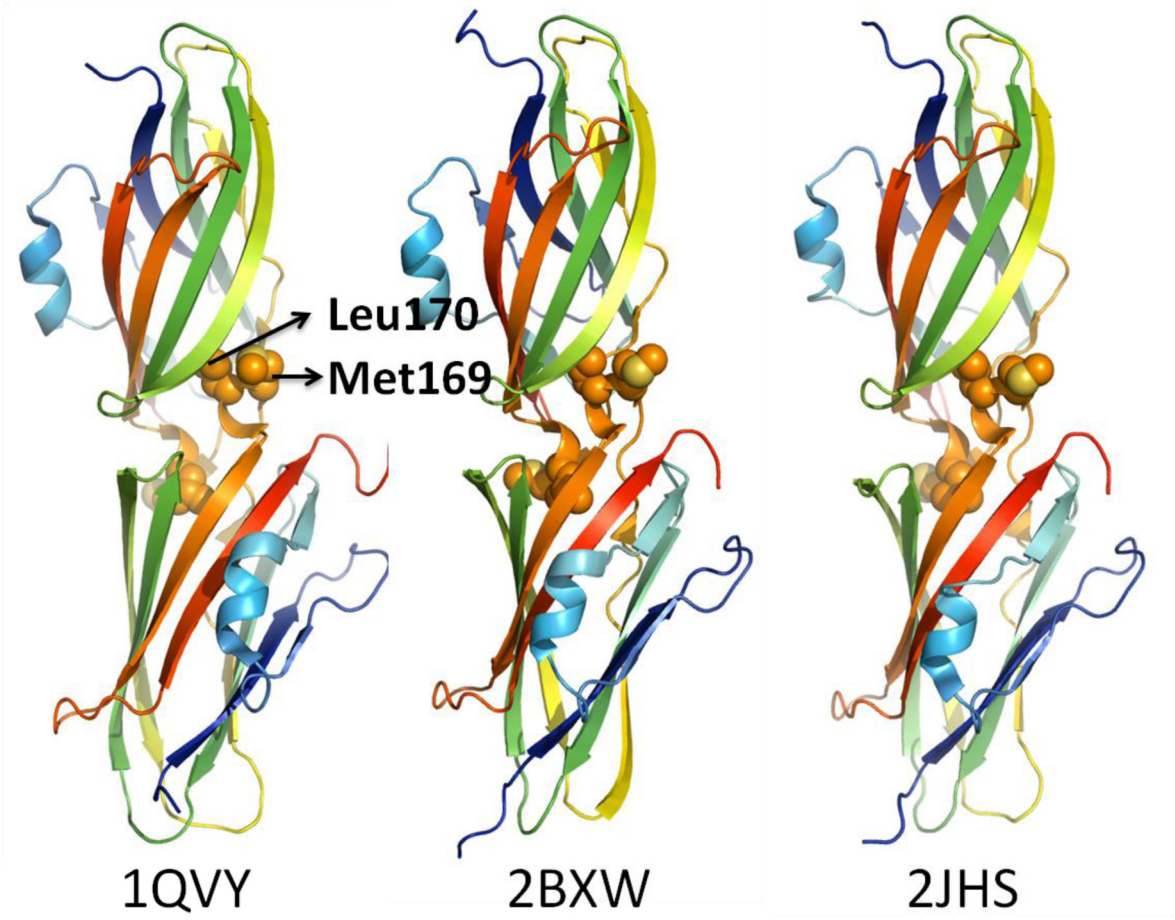


Fig. 5. Reproducibility of a homodimeric crystal contact in RhoGDI, independent of crystal forma and conditions

1QVY is a mutant containing a non-crystallographic dimer, crystallized from sodium formate and $(NH_4)_2SO_4$; 2BXW shows a crystallographic dimer, obtained from sodium citrate, with propanol and PEG; 2JHS shows a non-crystallographic dimer in crystals obtained from $(NH_4)_2SO_4$ and sodium citrate.

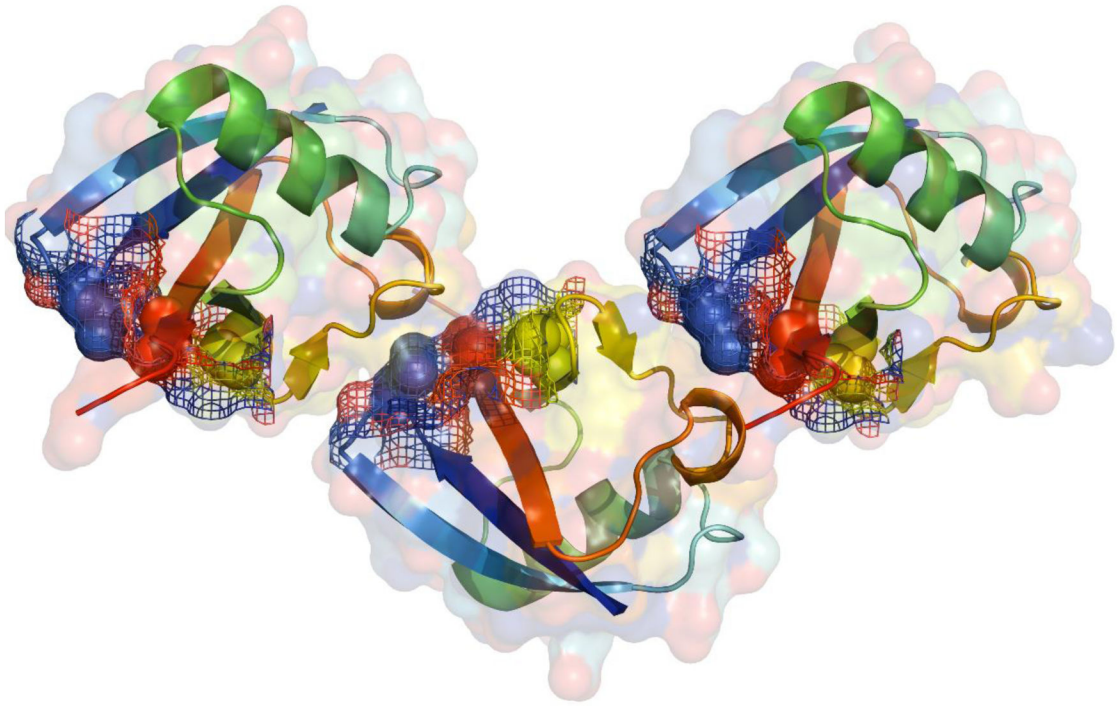


Fig. 6. A minor crystal contact in ubiquitin, mediated by a now recognized active surface

The three functional amino acids are Leu8, Ile44 and Val70. The contact making surface is highlighted as a mesh. PDB code 1UBQ.

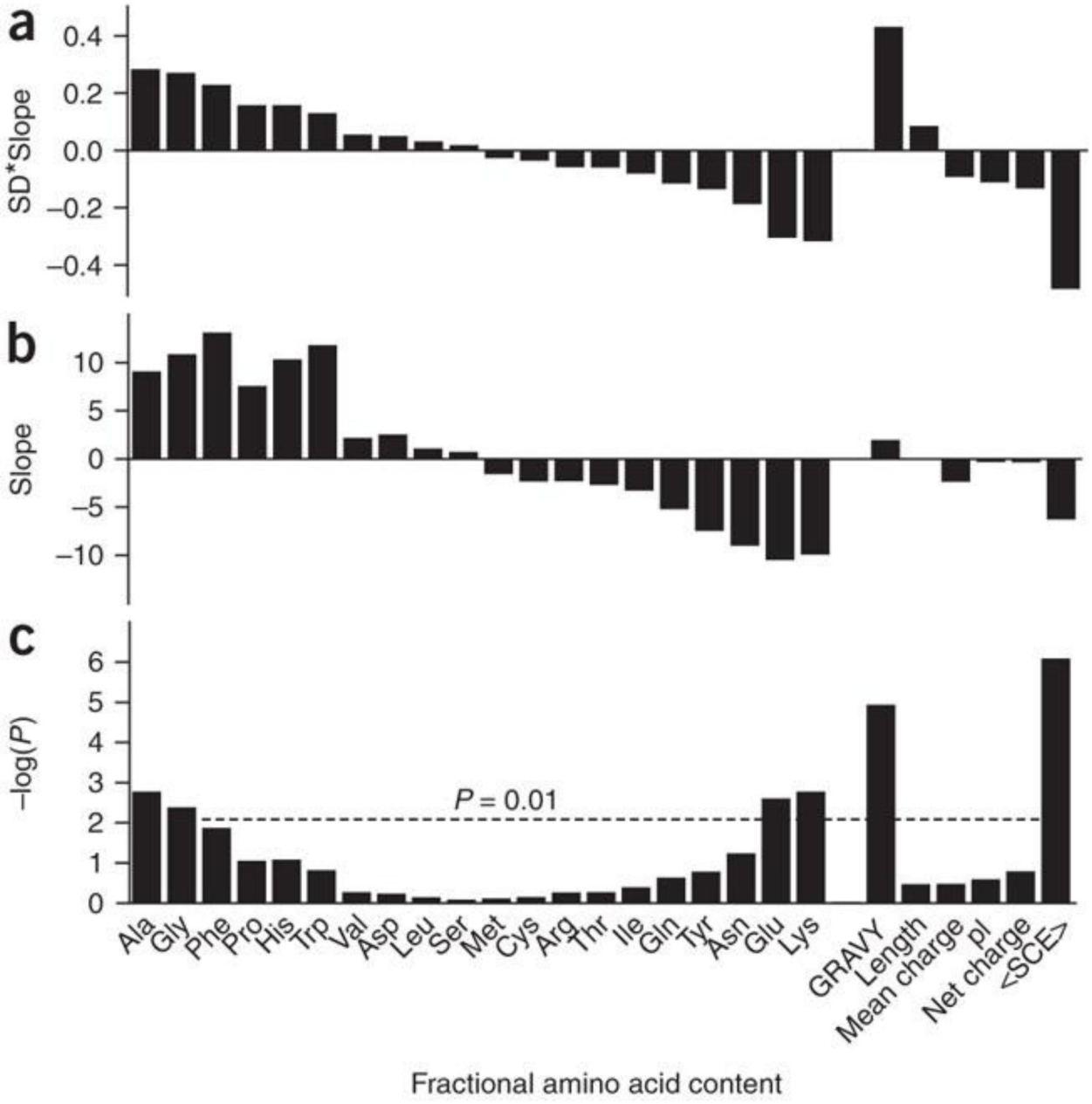


Fig. 7. Logistic regressions based on success in crystal structure determination (that is, PDB deposition) performed on a dataset of 679 proteins from the NESG protein expression and crystallization pipeline. Variables evaluated included the fractional content of each amino acid, mean residue hydrophobicity (GRAVY), chain length, mean charge (fraction arginine+lysine +asparagine+glutamatic acid), pI, mean net charge and ⟨SCE⟩

(a) Predictive value of each parameter, which is defined as the product of its logistic regression slope and the s.d. of its distribution in the dataset. **(b)** Logistic regression slope. **(c)** Negative log of logistic regression *P*-value. (From Price *et al.* (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol 27 (1):51–57; Copyright permission pending)

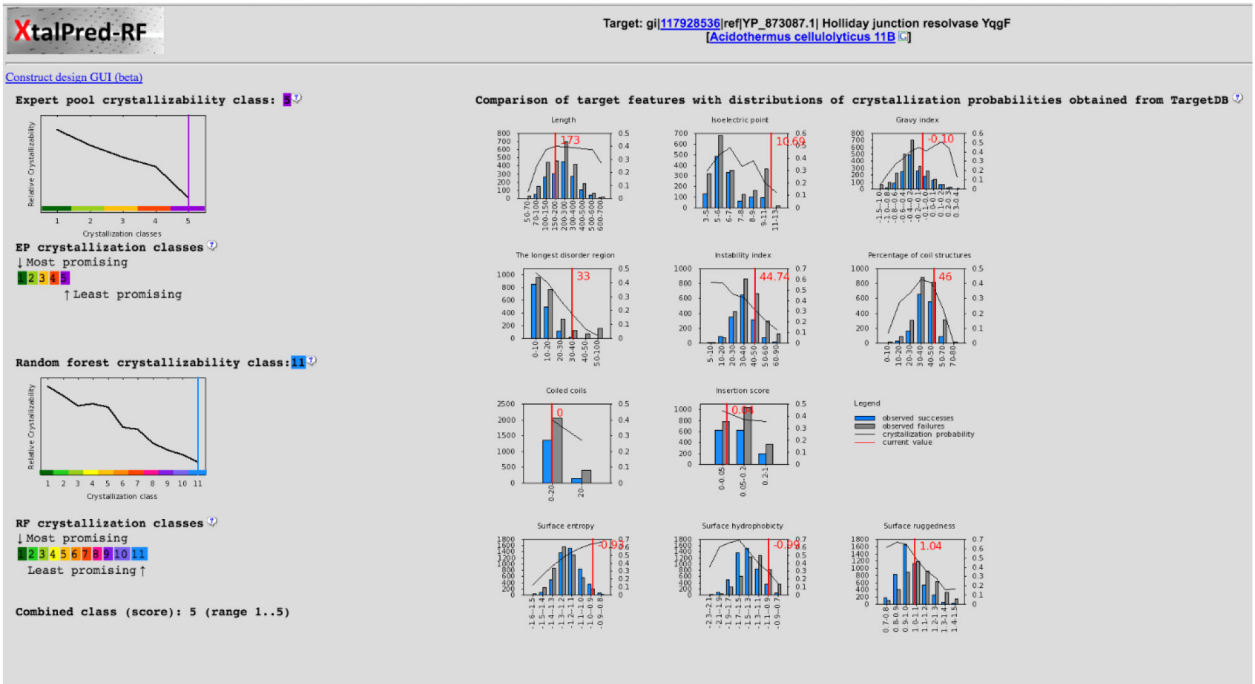


Fig. 8. Typical output from XTALPRED

The program analyzes various biophysical parameters and displays these values agains statistical data for other proteins and correlation with crystallizability. Random forest scoring puts the protein into 14 categories from the easiest to most difficult for crystallization.

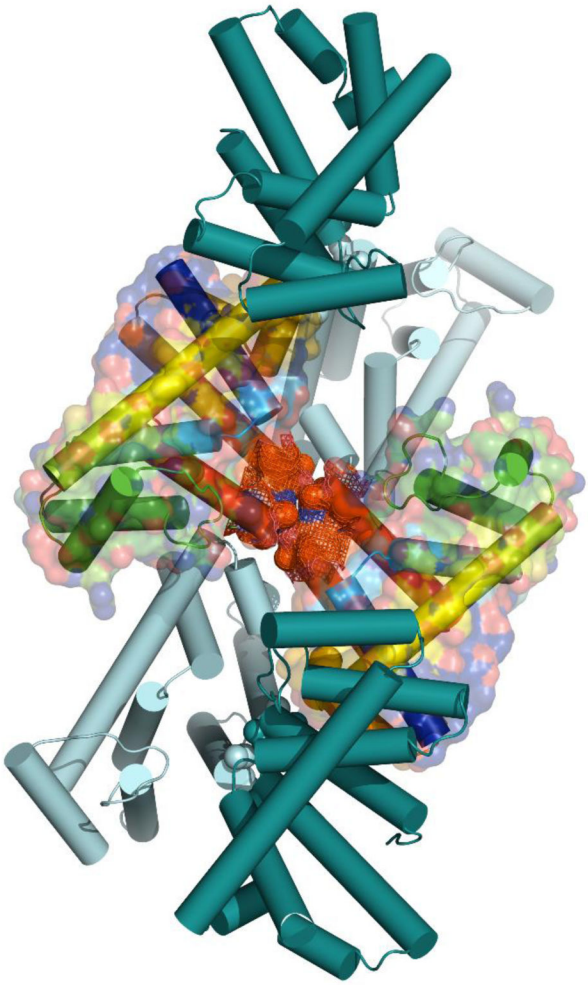


Fig. 9. A homotypic (homodimeric) contact mediated in the crystals of the RGS-like domain by the engineered patch created by mutations K463A, E465A and E466A (spheres)

The contact making surface is highlighted as a mesh. One dimer is shown in color, with full surface. The dimers below and above, arranges along a six-fold screw axis, are shown in cyan and green. PDB code 1HTJ.

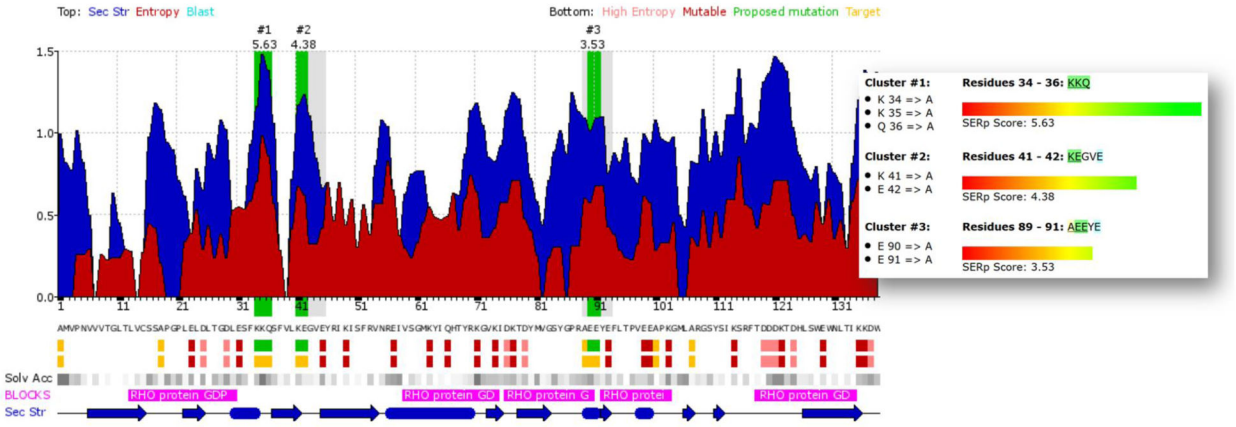


Fig. 10. Typical output from the SERp server

The program calculates a range of parameters, most important of which is the sliding window of side chain enetropy and secondary structure prediction. The program identified loops likely to be solvent exposed and suggests mutations in those loops where excess conformational entropy is likely to prevent specific contacts that may facilitate crystallization. Several variants, each with 2–3 mutations are suggested and scored. The sequence shown is that of RhoGDI; the inset shows a Table appearing elsewhere in the output scoring the suggested variants. All of the variants suggested here by the server are known to crystallize; the third variant yields crystals diffracting to 1.25 Å resolution.

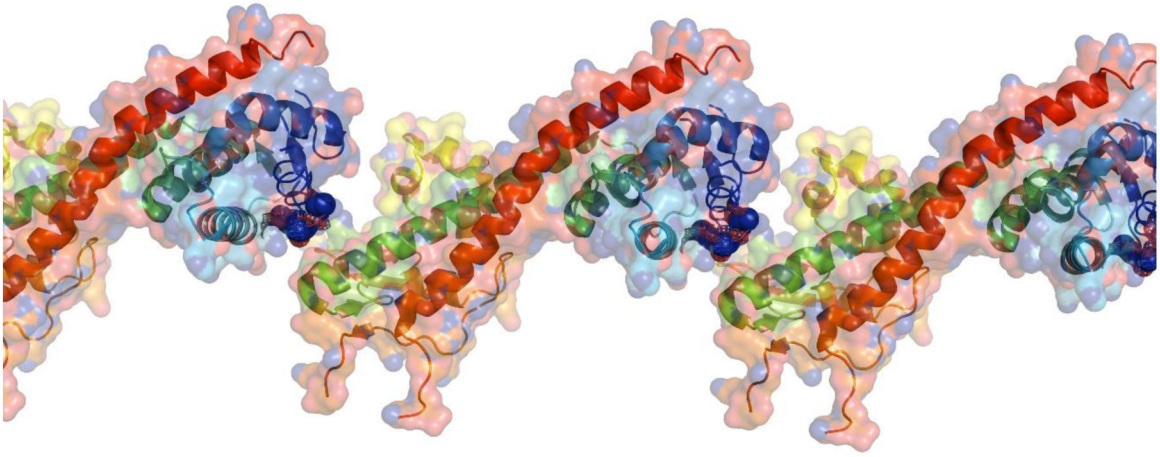


Fig. 11. A heterotypic contact mediated in the crystals of the *Yersinia pestis* V-antigen mediated is one molecule by a patch containing the mutations K40A, D41A, K42A (deep blue; shown as spheres)

The contact making surface is highlighted as a mesh. PDB code 1R6F.

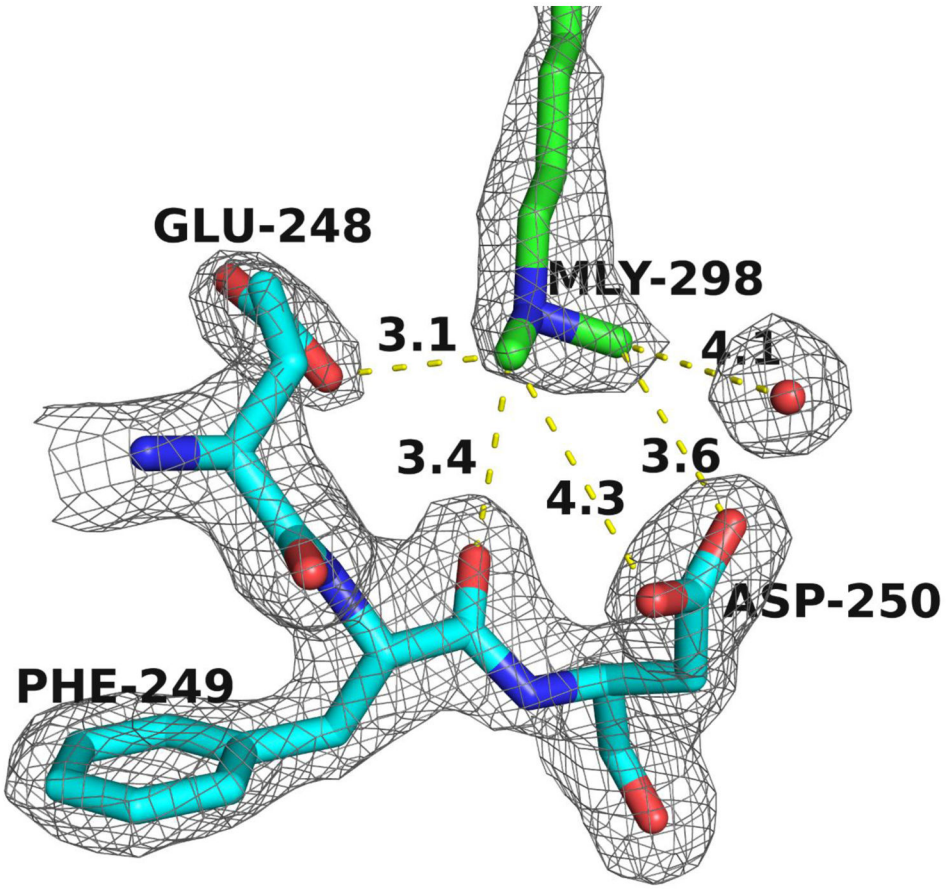


Fig. 12. A sticky patch mediated by di-methylated lysine

A crystal contact between two molecules of the *E. coli* RNA polymerase alpha subunit C-terminal domain (PDB 3K4G). Distances shown are in Å units.

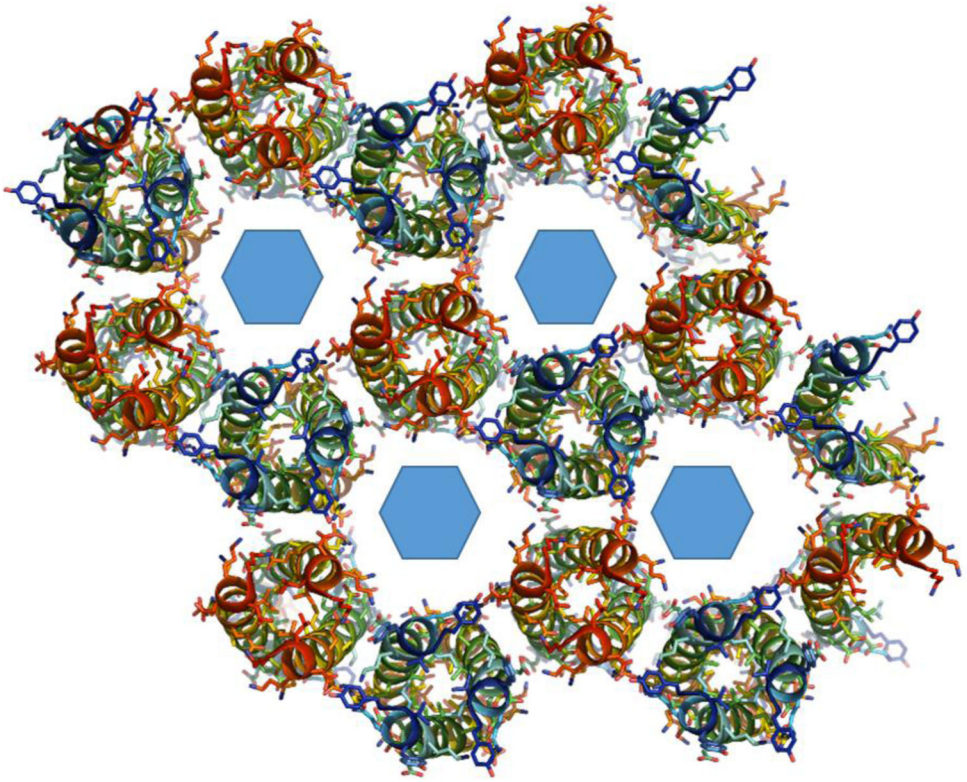


Fig. 13. A synthetic protein designed to crystallize in P6

A single of a designed helical protein forming a trimer, and then assembling into the P6 lattice. The six-fold axes are marked with hexagons. Three-fold axes are running down the trimers. (PDB code 3V86).