Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Systematic replacement of lysine with glutamine and alanine in *Escherichia coli* malate synthase G: effect on crystallization

David M. Anstrom, Leslie Colip, Brian Moshofsky, Eric Hatcher and S. James Remington*

Institute of Molecular Biology, University of Oregon, USA

Correspondence e-mail: jremington@uoxray.uoregon.edu

Received 12 September 2005
Accepted 8 November 2005
Online 24 November 2005

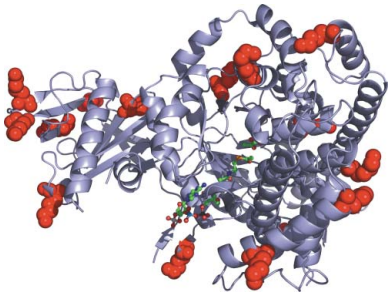
Two proposals recommend substitution of surface lysine residues as a means to improve the quality of protein crystals. In proposal I, substitution of lysine by alanine has been suggested to improve crystallization by reducing the entropic cost of ordering flexible side chains at crystal contacts. In proposal II, substitution of lysine by residues more commonly found in crystal contacts, such as glutamine, has been proposed to improve crystallization. 15 lysine residues on the surface of *Escherichia coli* malate synthase G, distributed over a variety of secondary structures, were individually mutated to both alanine and glutamine. For 28 variants, detailed studies of the effect on enzymatic activity and crystallization were conducted. This has permitted direct comparison of the relative effects of the two types of mutations. While none of the variants produced crystals suitable for X-ray structural determination, small crystals were obtained in a wide variety of conditions, in support of the general approach. Glutamine substitutions were found to be more effective than alanine in producing crystals, in support of proposal II. Secondary structure at the site of mutation does not appear to play a major role in determining the rate of success.

1. Introduction

Despite decades of intense effort, in most cases the primary impediment to X-ray crystallographic structure determination of proteins remains the difficulty of obtaining suitable crystals. Now that it has become routine to undertake major manipulations of the genes coding for proteins of interest, several new approaches to the crystallization problem have been advanced. In these proposals, protein engineering has been used with some success to improve crystal quality or to generate new crystal habits. Modifications have ranged from single or multiple mutations on the surface of the protein to N- or C-terminal deletions, loop deletions and the use of fusion proteins to promote crystallization (see reviews by Dale *et al.*, 2003; Derewenda, 2004*a*). The effect on crystallization of small affinity tags used in purification has also been investigated (Bucher *et al.*, 2002).

The substitution of even one residue can have a large effect on crystallization. For example, in human myoglobin a single Lys→Arg substitution was required to produce crystals suitable for diffraction studies (Hubbard *et al.*, 1990). In a second example, the K86Q mutation of human H ferritin was made to mimic crystal contacts previously observed in mouse and horse L-chain ferritins. This resulted in 1.9 Å diffraction, whereas crystals of wild-type human H ferritin were unsuitable for data collection (Lawson *et al.*, 1991). While these successes are very encouraging, they remain anecdotal. The general applicability of the approach needs to be systematically investigated. In the first more general study, the effects of surface mutations on protein solubility and crystallization properties were investigated for human thymidylate synthase (McElroy *et al.*, 1992). In that work, 12 different single-site mutations were made to the surface of the protein. Charged residues were replaced either with polar residues or with residues with the opposite charge, while nonpolar and uncharged residues were mutated to charged residues. Remarkably, three out of the 12 variants crystallized in space groups different from that of the wild-type protein.

To understand why such small changes were able to produce such drastic changes in crystallization behaviour, it may be useful to consider the properties of typical crystal contacts. Firstly, the area of



an individual crystal contact is usually small. According to one survey of 152 crystal structures, an average of 280 Å^2 is buried (roughly ten amino-acid residues) on each molecule making a pairwise crystal contact (Janin & Rodier, 1995). In a different study, 78 crystal structures were investigated and it was found that crystal contact areas were slightly smaller: 45% were less than 100 Å^2 and only 8% were larger than 500 Å^2. In terms of protein segments separated by at least five residues, 68% of crystal contacts were comprised of a single segment and 21% were comprised of two segments (Carugo & Argos, 1997).

Crystal contacts in the Protein Data Bank (Berman *et al.*, 2000) have also been investigated with respect to the prevalence of amino-acid type, the tendency of certain amino-acid side chains to form specific and non-specific interactions in the crystal contact and the effect of ionic strength (Dasgupta *et al.*, 1997; Iyer *et al.*, 2000). Crystal contacts tend to include polar interactions and to avoid inclusion of hydrophobic side chains. However, lysines are systematically under-represented. As arginine and glutamine are the two most favored residues in crystal interfaces, it was suggested that substitution of surface lysines with arginine and/or glutamine might aid in crystallization (Dasgupta *et al.*, 1997). A study which included the effects of ionic strength in the mother liquor showed that in addition to being the most favored overall at crystal contacts, arginine is more favored at low ionic strength and glutamine is more favored at high ionic strength (Iyer *et al.*, 2000). Taken together, the small size of crystal contacts and the relative frequency with which residues are present in crystal contacts make it quite plausible that a single substitution or a small number of substitutions (especially at adjacent sites in sequence) on the surface of a protein could substantially alter the shape and/or electrostatic potential of a crystal contact, leading to changes in crystallization behaviour (Hubbard *et al.*, 1990; Lawson *et al.*, 1991; McElroy *et al.*, 1992). It is not surprising that when known crystal contacts are specifically targeted for mutagenesis, the effect on crystallization can be dramatic (Camara-Artigas *et al.*, 2001; Charron *et al.*, 2002).

In an alternative proposal, entropic effects have been considered. Crystallization is accompanied by a change in free energy and some portion arises from loss of entropy from reduced rotational and translational freedom of the protein as it joins the growing crystal (Vekilov, 2003). Additionally, there will be some contribution arising from a loss of conformational entropy of side chains, particularly those on the surface (Derewenda, 2004*b*). As some residues have higher side-chain flexibility than others (Avbelj & Fele, 1998, Creamer, 2000), a higher entropic cost may be paid to fix these residues in the protein crystal. Therefore, it has been argued that surface residues with high side-chain conformational entropy, such as lysine and glutamate, should be replaced by residues with little or none, such as alanine (Derewenda, 2004*b*). This method has been called the 'surface-entropy reduction approach' (Garrard *et al.*, 2001) and may even be applicable to proteins whose structures have not been solved. Lysine is buried only 6% of the time (Baud & Karlin, 1999), so any short protein segment with one or more lysines and/or other residues with high surface proclivity is very likely to be solvent-exposed. Thus, such segments are good candidates for substitution by alanine (Derewenda, 2004*b*). This approach is compatible with the observation that crystal contact patches are small and are frequently comprised of only one protein segment (Carugo & Argos, 1997; Janin & Rodier, 1995).

The arguments on which these proposals are based are to some extent incompatible. Glutamine and arginine have roughly the same side-chain conformational entropy as lysine (Avbelj & Fele, 1998; Creamer, 2000) and replacement of lysine with either of these residues may not lower the entropic cost of crystallization as suggested by the entropy-reduction approach. Conversely, the order of prevalence at crystal contacts is (Arg, Gln) > Ala > Lys (Dasgupta *et al.*, 1997), which suggests that replacement of lysine with arginine or glutamine may be more effective than replacement by alanine.

The most thoroughly studied protein with respect to the above proposals is the globular domain of human Rho-specific guanine nucleotide-dissociation inhibitor (RhoGDIΔ66N). Single and multiple substitutions of lysine to alanine (Longenecker, Garrard *et al.*, 2001), glutamic acid to alanine and aspartic acid (Mateja *et al.*, 2002) and lysine to arginine (Czepas *et al.*, 2004) mutants have been made and the variants subjected to crystallization experiments; in the case of suitably diffracting crystals the structure was solved. Not all mutation types were equally effective; for instance, no diffraction-quality crystals of any of the glutamic acid to aspartic acid mutants were obtained, while the E154A/E155A mutant produced crystals that diffracted to 1.2 Å. Lysine to alanine mutations also resulted in crystals that diffracted well, with variants producing crystals that diffracted to a maximum resolution of 2.0–2.8 Å. One lysine to arginine mutant resulted in a crystal that diffracted to 1.6 Å, while five other variants produced crystals that diffracted in the 2.3–4.0 Å resolution range.

We chose to investigate both approaches with malate synthase, for which it is very difficult to obtain diffraction-quality crystals. Malate synthase has been implicated as a virulence factor in several different pathogens, most notably *Mycobacterium tuberculosis* (Graham & Clark-Curtiss, 1999) and *Candida albicans* (Lorenz & Fink, 2001); hence it is a promising target for structure-based inhibitor design. The structure of *Escherichia coli* malate synthase isoform G (ECMSG), a monomeric enzyme of molecular weight 80 kDa, has been solved at high resolution (Howard *et al.*, 2000; Anstrom *et al.*, 2003). The crystals readily grow in needle form in two different crystal forms; however, it is extremely difficult to obtain crystals with a cross-section large enough for diffraction studies. Form I crystals belong to space group $P2_12_12_1$, with unit-cell parameters 73.8, 88.7 and 109.9 Å. They were obtained by hanging-drop vapor diffusion against high-salt well solution containing 1.8 *M* $(NH_4)_2SO_4$, 100 m*M* imidazole pH 8.0 and 50 m*M* sorbitol. The protein solution contained the enzyme at 24 mg ml^{-1}, 14 m*M* $MgCl_2$ and 7 m*M* DTT. Form II crystals also

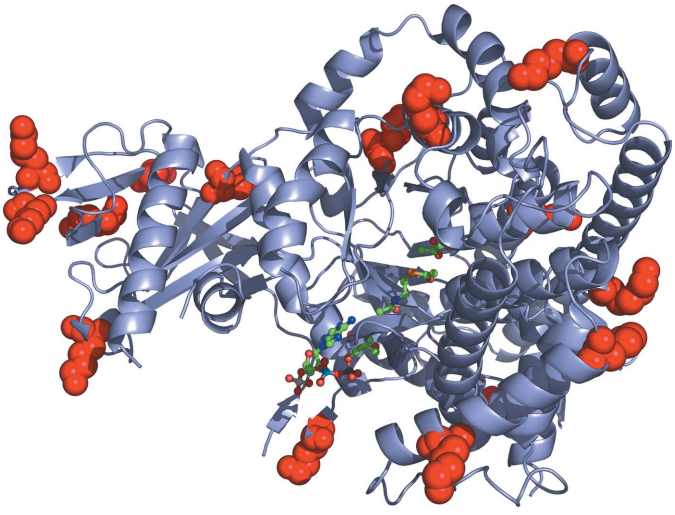


Figure 1
Ribbon diagram of *E. coli* malate synthase G with mutated lysine side chains colored red. Coenzyme A and pyruvate bound at the active site are shown in ball-and-stick representation. This figure was prepared using *PyMOL* (DeLano, 2002).

Table 1
Revised Fast Screen (RFS) crystallization conditions.

Condition No.	Final pH	Salt	Buffer	Precipitant
1	4.6	0.2 *M* $CaCl_2$	0.1 *M* acetate pH 4.6	30% MPD
2	7.3	—	0.1 *M* HEPES pH 7.5	40% MPD
3	6.4	0.2 *M* magnesium acetate	0.1 *M* PIPES pH 6.5	30% MPD
4	6.4	0.2 *M* sodium acetate	0.1 *M* citrate pH 5.6	30% MPD
5	6.1	—	0.1 *M* Tris pH 8.5	1.8 *M* ammonium phosphate
6	6.1	—	0.1 *M* citrate pH 5.6	1.5 *M* ammonium phosphate
7	6.5	—	—	2 *M* sodium/potassium phosphate pH 6.5
8	6.9	0.2 *M* NaCl	—	2 *M* sodium/potassium phosphate pH 7.0
9	6.7	—	0.1 *M* PIPES pH 6.5	1.4 *M* sodium acetate
10	7.5	—	0.1 *M* imidazole pH 7.5	1 *M* sodium acetate/10% PEG 1550
11	7.7	—	0.1 *M* HEPES pH 7.5	1.4 *M* sodium citrate
12	7.6	—	0.1 *M* HEPES pH 7.5	1.5 *M* Li_2SO_4
13	7.6	—	0.1 *M* HEPES pH 7.5	0.8 *M* sodium/potassium tartrate
14	7.5	—	—	4 *M* sodium formate
15	5.0	—	0.1 *M* acetate pH 4.6	1.8 *M* sodium formate
16	8.0	—	0.1 *M* Tris pH 8.5	2 *M* ammonium sulfate
17	6.6	—	0.1 *M* PIPES pH 6.5	2 M ammonium sulfate
18	7.5	—	0.1 *M* HEPES pH 7.5	2 *M* ammonium sulfate/5% PEG 400
19	4.4	—	0.1 *M* sodium acetate pH 4.6	2 *M* ammonium sulfate
20	7.3	0.2 *M* $MgCl_2$	0.1 *M* HEPES pH 7.5	30% PEG 4000
21	8.7	0.2 *M* Li_2SO_4	0.1 *M* Tris pH 8.5	30% PEG 4000
22	5.5	0.2 *M* ammonium acetate	0.1 *M* acetate pH 4.6	30% PEG 4000
23	6.4	0.2 *M* ammonium acetate	0.1 *M* citrate pH 5.6	30% PEG 4000
24	4.8	0.2 *M* ammonium sulfate	0.1 *M* acetate pH 4.6	30% PEG 4000
25	7.0	0.2 *M* ammonium sulfate	0.1 *M* HEPES pH 7.0	30% PEG 4000
26	6.9	—	0.1 *M* cacodylate pH 6.5	30% PEG 4000
27	8.7	0.2 *M* sodium acetate	0.1 *M* Tris pH 8.5	30% PEG 4000
28	6.4	—	0.1 *M* citrate pH 5.6	20% PEG 4000/10% 2-propanol
29	7.5	—	0.1 *M* HEPES pH 7.5	20% PEG 4000/5% 2-propanol
30	6.7	0.14 *M* sodium citrate	0.1 *M* citrate pH 5.6	20% PEG 8000/10% 2-propanol
31	7.3	0.2 *M* $MgCl_2$	0.1 *M* HEPES pH 7.5	20% PEG 8000/10% 2-propanol
32	7.1	0.2 *M* sodium citrate	0.1 *M* HEPES pH 7.0	20% PEG 8000/10% 2-propanol
33	6.5	0.2 *M* magnesium acetate	0.1 *M* PIPES pH 6.5	20% PEG 8000
34	6.6	0.2 *M* sodium acetate	0.1 *M* PIPES pH 6.5	30% PEG 8000
35	6.4	0.2 *M* ammonium sulfate	0.1 *M* PIPES pH 6.5	30% PEG 8000
36	8.5	—	0.1 *M* Tris pH 8.5	15% PEG 8000
37	8.0	—	0.05 *M* potassium phosphate pH 7.5	25% PEG 8000
38	6.4	0.2 *M* zinc acetate	0.1 *M* PIPES pH 6.5	20% PEG 8000
39	6.5	0.2 *M* calcium acetate	0.1 *M* PIPES pH 6.5	20% PEG 8000
40	7.7	0.2 *M* NaCl	0.1 *M* potassium phosphate pH 7.5	25% PEG 8000
41	6.25	0.5 *M* Li_2SO_4	—	15% PEG 8000
42	7.2	0.2 *M* $CaCl_2$	0.1 *M* HEPES pH 7.5	25% PEG 400
43	8.9	0.2 *M* sodium citrate	0.1 *M* Tris pH 8.5	30% PEG 1550
44	7.4	0.2 *M* $MgCl_2$	0.1 *M* HEPES pH 7.5	30% PEG 1550
45	6.5	—	0.1 *M* PIPES pH 6.5	30% PEG 1550
46	8.2	0.2 *M* ammonium acetate	0.1 *M* Tris pH 8.5	20% 2-propanol
47	4.6	0.2 *M* $CaCl_2$	0.1 *M* acetate pH 4.6	20% 2-propanol
48	7.4	0.14 *M* sodium citrate	0.1 *M* HEPES pH 7.5	20% 2-propanol

belong to space group $P2_12_12_1$, but with unit-cell parameters $a = 73.9$, $b = 107.4$, $c = 205.0$ Å. They were obtained from roughly physiological salt concentrations using hanging drops containing 7 µl protein solution consisting of 28 mg ml^{-1} protein, 8.3 m*M* $MgCl_2$, 8.3 m*M* Tris pH 7.9, 4.2 m*M* DTT, 50 m*M* pyruvate and 4.2 m*M* acetyl coenzyme A added to 7 µl of well solution containing 50 m*M* HEPES pH 7.0, 150 m*M* sodium acetate, 5 m*M* $MgCl_2$ and 16% PEG 8000. The packing of the molecules seems to be completely different in the two crystals.

In an attempt to improve the crystallization of the enzyme, we selected 15 of the surface lysines in ECMSG (Fig. 1) and mutated each position individually to both alanine and glutamine. We then investigated the relative effect of these two mutation types on crystallization and enzymic activity. Alanine was chosen to investigate the surface-entropy reduction approach (Derewenda, 2004*b*), while glutamine was chosen as it is favored in crystal contacts (Dasgupta *et al.*, 1997), particularly at high ionic strength (Iyer *et al.*, 2000). While the effects of lysine to arginine substitutions have been studied systematically (Czepas *et al.*, 2004), the effects of lysine to glutamine substitution have not. Our systematic study has the advantage of allowing direct comparison of the effects of the same two substitutions at 15 different surface locations on a large monomeric enzyme.

2. Materials and methods

2.1. Mutagenesis, protein expression and purification

Mutations were introduced into the cDNA of *E. coli* malate synthase G (Howard *et al.*, 2000) using either the QuikChange site-directed mutagenesis kit (Stratagene) or two-stage PCR (Wang & Malcolm, 1999) with primers purchased from Integrated DNA Technologies. All mutations were verified by DNA sequencing. Single colonies of each mutant were picked and a 10 ml overnight culture grown in LB broth was then used to inoculate 1 l SLBH. Cultures were grown to $OD_{600} \simeq 1.0$ before induction with 50 µg ml^{-1} IPTG (final concentration) at 293 K overnight. Cultures were harvested by centrifugation and pellets were resuspended in 25 ml lysis buffer (50 m*M* HEPES pH 7.9, 0.3 *M* NaCl, 10% glycerol, 10 m*M* $MgCl_2$, 2 m*M* β-mercaptoethanol) with EDTA-Free Complete Mini protease-inhibitor tablets (Roche) as per the manufacturer's instructions, lysed by sonication and the cellular debris was pelleted by centrifugation. The supernatant was passed through a 5 ml Ni–NTA Agarose (Qiagen) column pre-equilibrated with lysis buffer, washed with 10 ml lysis buffer and then with 5 ml 20 m*M* imidazole, 50 m*M* HEPES pH 7.9, 0.3 *M* NaCl, 10 m*M* $MgCl_2$, 2 m*M* β-mercaptoethanol and eluted with 20 ml 100 m*M* imidazole, 50 m*M*

Table 2
Revised Fast Screen crystallization hit count by position.

				Hits				
				Alanine		Glutamine		
Position	Mutated	Secondary structure†	Exposure‡	277 K	RT	277 K	RT	Total
17	A, Q	Helix	I	0	2	9	2	13
76	—	Coil	E	—	—	—	—	—
78	A, Q	Helix	I	8	4	9	3	24
82	—	Helix	I	—	—	—	—	—
192	A, Q	Sheet	I	4	2	3	4	13
196	A, Q	Coil	E	6	3	14	5	28
203	A, Q	Coil	E	0	1	5	1	7
206	A, Q	Sheet	E	4	3	9	3	19
232	A, Q	Sheet	I	7	1	8	4	20
250	A, Q	Helix	E	7	1	8	5	21
283	—	Helix	B	—	—	—	—	—
301	A, Q	Sheet	E	4	2	9	6	21
304	—	Disordered	NA	—	—	—	—	—
312	—	Sheet	E	—	—	—	—	—
376	—	Helix	E	—	—	—	—	—
379	—	Coil	I	—	—	—	—	—
390	—	Sheet	B	—	—	—	—	—
392	—	Coil	B	—	—	—	—	—
404	A, Q	Helix	E	7	4	8	5	24
421	—	Sheet	B	—	—	—	—	—
473	—	Helix	I	—	—	—	—	—
477	—	Helix	I	—	—	—	—	—
483	Q	Helix	E	—	—	5	5	10
501	Q	Coil	E	—	—	2	1	3
506	—	Sheet	B	—	—	—	—	—
522	—	Helix	B	—	—	—	—	—
619	—	Sheet	E	—	—	—	—	—
653	A, Q	Helix	E	11	3	5	5	24
665	A, Q	Helix	E	8	3	11	4	26
691	—	Helix	I	—	—	—	—	—
702	A, Q	Coil	E	7	3	4	3	17
720	—	Helix	I	—	—	—	—	—
Total§				73	32	109	56	270

† Secondary-structural classification was performed using the program *DSSP* (Kabsch & Sander, 1983). Both 3_{10}- and α-helices are denoted as Helix; Coil corresponds to all secondary-structural elements except helices and sheets. Residue 304 was disordered in all structures. ‡ Disordered lysines were automatically modeled by the *Swiss-Pdb Viewer* (Guex & Peitsch, 1997) and lysine side-chain solvent accessibility was calculated using the *CCP*4 program *AREAIMOL*, which was subsequently classified as B for buried, I for intermediate and E for exposed after Rost & Sander (1994). § Wild-type *E. coli* malate synthase G produced seven crystals at 277 K and five at room temperature and is not included in the table.

HEPES pH 7.9, 0.3 *M* NaCl, 10 m*M* $MgCl_2$, 2 m*M* β-mercaptoethanol. The protein was then dialyzed for 24 h with two buffer changes against 1 l 50 m*M* HEPES pH 7.9, 0.1 *M* NaCl, 10 m*M* $MgCl_2$, 2 m*M* DTT. Following concentration to 30 mg ml^{-1}, protein was then further purified by FPLC using a HiLoad 16/60 Superdex 200 gel-filtration column (Amersham Biosciences). The enzymatic activity of each mutant was assayed as previously described (Anstrom *et al.*, 2003).

2.2. Crystallization and data collection

Mutant and wild-type proteins were crystallized by hanging-drop vapour diffusion in 24-well VDX plates (Hampton). Two crystallization screens were used: the Revised Fast Screen (RFS, Table 1), which was developed in-house and is based on the Hampton Crystal Screen and the sparse-matrix screen described by Jancarik & Kim (1991), and the Core Screen from the *Thermotoga maritima* structural genomics project (Page *et al.*, 2003). Protein concentration was 16 $\pm$ 1 mg ml^{-1} for room-temperature trays and 18 $\pm$ 1 mg ml^{-1} for 277 K trays and was in a buffer containing 50 m*M* HEPES pH 7.9, 0.1 *M* NaCl, 10 m*M* $MgCl_2$, 2 m*M* DTT, 0.5 *M* L-malate and 40 m*M* coenzyme A. Drop size was 2 µl with a 1:1 mixture of protein and well solutions. Trays were manually evaluated periodically for no less than one month. Crystals of large enough size to be suitable for X-ray diffraction experiments were flash-frozen and assessed using a 6 kW rotating-anode X-ray source and an R-AXIS IV image-plate detector.

2.3. Crystallization scoring

All crystal trays were observed using a standard light microscope and scored periodically for a period of not less than one month. A condition was considered a 'hit' or positive if the drop contained microcrystals, single crystals or clusters of crystals which typically were needles.

3. Results

A total of 28 single-lysine surface mutations (13 Lys→Ala, 15 Lys→Gln) were made to ECMSG at 15 different sites (Fig. 1 and Table 2). These sites were selected to more or less evenly cover the surface of the enzyme, but are not close to the substrate-binding pockets so as to not interfere with enzyme function. To ensure that the mutations did not significantly alter the properties of the enzyme activity, assays were conducted on each mutant as described previously (Anstrom *et al.*, 2003). Activity was comparable to or slightly greater than the wild type in all cases (data not shown). Although other workers have proposed to only substitute lysines located on exposed turns (Derewenda, 2004*b*), in our study some mutation sites were located on helices (seven positions) and β-strands (four positions) in addition to turns (four positions).

For each mutant and the wild-type enzyme, RFS trays (Table 1) were set up at both room temperature and 277 K. Core Screen (Page *et al.*, 2003) trays were also set up for 23 mutants and the wild-type enzyme at room temperature. All trays were manually checked for crystals over the course of at least one month. The rate at which the Core Screen produced hits was very low and then only in conditions similar to those found in the RFS. In addition, this screen contains a high proportion of toxic reagents, so it was abandoned early in the investigation. It is interesting to note that in the *T. maritima* structural genomics project, the largest protein successfully crystallized using the Core Screen was 93.5 kDa (Page *et al.*, 2003), comparable in size to ECMSG.

Overall, there were a total of 282 unique crystallization hits from 2784 unique crystallization experiments. Each mutant and the wild-type protein produced at least one crystal hit (Table 2). While several conditions produced no crystals for any protein, no single condition produced crystals for every protein. The condition with the greatest number of hits was RFS condition 26 (0.1 *M* cacodylate pH 6.5, 30% PEG 8000) at 277 K, followed by RFS condition 14 (4 *M* sodium formate) at room temperature, which produced 22 and 19 hits, respectively. In all but two cases the hits produced were either microcrystals, long thin needles or needle clusters; in no case was a single needle larger than 20 µm across. The two exceptions were K196Q, which produced a cluster of thin stacked plates at 277 K in RFS condition 22 (30% PEG 4000, 0.1 *M* acetate pH 4.6, 0.2 *M* ammonium acetate), and K301A, which produced a highly twinned mass of bars at 277 K in RFS condition 1 (30% MPD, 0.1 *M* acetate pH 4.6, 0.2 *M* $CaCl_2$). In the latter two cases, a piece of each crystal was flash-frozen in mother liquor and on the home source diffracted to no better than 15 Å resolution. Diffraction may have been negatively affected by handling and/or flash-freezing. It is important to note that since we were unable to characterize the unit cells of the crystals obtained in this experiment, some of them may have been

Table 3
Revised Fast Screen crystallization hit count by precipitant and mutation type at both 277 K and room temperature.

	Salt	Small organic†	Large organic‡	Total
Wild type	5	0	7	12
Alanine	26	3	76	105
Glutamine	61	3	101	165
Total	92	6	184	282
Conditions in RFS	15	11	22	48

† PEG 1550 and smaller. ‡ PEG 4000 and 8000.

similar or identical to one of the two crystal forms that had already been analyzed at high resolution.

By mutation type, glutamine is preferred over alanine. Assuming a Gaussian distribution of errors, the preference for glutamine is significant at better than the 99% confidence level (three times the standard error). Glutamine substitutions produced 165 hits from 1492 experiments with 15 mutants (success rate 11.1% ± standard error 0.8%; Table 3), whereas alanine substitutions produced 105 hits from 1292 experiments with 13 mutants (success rate 8.1 ± 0.7%). Discounting 13 hits obtained for the two sites for which the results of only glutamine substitutions were analyzed, the success rate for glutamine increases slightly to 152/1292 or 11.8 ± 0.8%. Alanine and glutamine mutations at the same site also do not always crystallize under the same conditions.

In terms of crystallization conditions, alanine and glutamine mutations as a whole produce more crystallization hits at 277 K than at room temperature, as did the wild-type enzyme (Table 2). Hits for the wild-type enzyme and both alanine and glutamine mutations cover nearly the entire pH range of the RFS. When the RFS is broken down into three groups of precipitants, high-salt, low-molecular-weight organics (*e.g.* PEG 1550 and smaller) and high-molecular-weight organics (PEG 4000 and 8000), then there is a noticeable difference between alanine and glutamine mutations. Alanine substitutions produced three times as many crystals in large organic precipitants than in salts, whereas glutamine substitutions showed a slight preference for large organic precipitants. Few crystals were obtained using low-molecular-weight organic precipitants (Table 3).

If the crystallization hits in this study are broken down by secondary-structure type, mutations on helices produced 142 positives from 13 mutants (11 ± 0.8% success rate), 73 positives were produced from eight β-sheet mutants (9 ± 1% success) and 55 positives were produced from seven mutants situated on turns or coils (8 ± 1% success). While one could argue that there may be a slight bias favouring helical sites, in our opinion the most important conclusion that one can derive from this result is that it would be unwise to exclude any candidate site from mutagenesis on the basis of secondary-structure considerations alone.

For all the substitutions, the impact of side-chain disorder, degree of surface exposure and number of crystal contacts on number of crystallization hits was investigated, but no readily discernable trends were detected (data not shown).

4. Discussion

There are two competing proposals regarding the most appropriate substitution of surface lysine residues for improving protein crystallization. The first is the surface-entropy reduction approach (Derewenda, 2004*b*), in which single or preferably multiple adjacent lysines and/or other flexible residues are mutated to alanine in order to reduce the entropic cost of crystallization. The second proposes to replace residues that are uncommon in crystal contacts, such as lysine, with residues more frequently found in crystal contacts, such as arginine or glutamine (Dasgupta *et al.*, 1997). The two proposals are incompatible as alanine has neutral preference for crystal contacts (Dasgupta *et al.*, 1997), while arginine and glutamine have roughly the same side-chain entropy as lysine (Avbelj & Fele, 1998; Creamer, 2000).

In this study, 15 lysines on the surface of *E. coli* malate synthase G were individually mutated to alanine and glutamine. Of the 28 variants investigated in detail, no mutant produced crystals amenable to high-resolution X-ray structural analysis. However, some generalizations can be made. The first is that glutamine mutations produced significantly more crystal hits than did alanine mutations (11 ± 0.8% *versus* 8 ± 0.7%, respectively). Secondly, the distribution of hits by precipitant for the two types of mutations is different: alanine substitutions crystallized in large organic precipitants at a rate of three to one over salts, while this preference is much less pronounced in glutamine substitutions (Table 3). Thirdly, secondary structure does not appear to play a large role in the frequency of hits produced by Lys→Ala and Lys→Gln mutations (Table 2), so no candidate site should be excluded on the basis of secondary structure alone. Finally, although this result may be specific to malate synthase, crystal setups at 277 K were more successful than those kept at room temperature. These results have broad implications for improvement of protein crystallization by alteration of surface characteristics in general.

The finding that glutamine mutations were more successful than alanine appears to support the proposal that surface lysines should be substituted with residues more frequently found in crystal contacts such as glutamine or arginine (Dasgupta *et al.*, 1997) rather than alanine (Derewenda, 2004*b*). As glutamine and arginine have similar side-chain conformational entropy to lysine (Avbelj & Fele, 1998; Creamer, 2000), it seems likely that the predominant effect of the Lys→Gln substitution is to alter the shape and charge characteristics of a surface patch, which might allow crystal contacts that were not permitted in the wild-type protein. This may be a general result. It has previously been argued that the improved crystallization of Lys→Ala mutants may arise from a change in surface properties, rather than entropic effects (Qiu & Janson, 2004).

In fact, in the majority of cases where surface mutagenesis has led to a structure determination, the sites of mutation were located at crystal contacts (Buschiazzo *et al.*, 2002; Czepas *et al.*, 2004; Derewenda *et al.*, 2004; Devedjiev *et al.*, 2004; Hubbard *et al.*, 1990; Janda, Devedjiev, Cooper *et al.*, 2004; Janda, Devedjiev, Derewenda *et al.*, 2004; Lawson *et al.*, 1991; Longenecker, Garrard *et al.*, 2001; Longenecker, Lewis *et al.*, 2001; Mateja *et al.*, 2002; Munshi *et al.*, 2003; Prag *et al.*, 2003). Similar generalizations based on a smaller number of crystal structures have been made previously (Derewenda, 2004*b*; Qiu & Janson, 2004).

Secondary structure does not seem to have a profound influence on the success of lysine→glutamine or alanine substitutions. This suggests that there is no reason to limit mutations to surface loops or coiled regions, as previously proposed (Derewenda, 2004*b*), so long as the substitutions do not greatly destabilize the secondary or tertiary structure. This greatly increases the number of available sites to mutate, as helices, sheets and coils are present at crystal interfaces at roughly the same frequency (36, 38 and 26%, respectively; Dasgupta *et al.*, 1997).

On a final note, our systematic study was based on knowledge of the three-dimensional structure of *E. coli* malate synthase G (Anstrom *et al.*, 2003; Howard *et al.*, 2000). This allowed selection of sites that were distant from the active site and located on the surface

of the molecule. Although lysine is buried only 6% of the time (Baud & Karlin, 1999), out of a total of 32 lysines in malate synthase two that are in close proximity (lysines 390 and 392, Table 2) are in fact buried. In those cases where the benefit of hindsight is not available, modification of one or more buried lysines (or glutamates *etc.*) may have a detrimental effect on the activity or stability of the protein, which can be assayed by a variety of techniques. In recognition of this, the effect of mutations on the stability of RhoGDIΔ66N (Czepas *et al.*, 2004; Mateja *et al.*, 2002) was assayed by thermal denaturation. The use of dynamic light scattering has also been suggested as a general approach to assess changes in structure or stability (Zulauf & D'Arcy, 1992). Various activity assays have been used to assess the functionality of modified proteins (Buschiazzo *et al.*, 2002; Dale *et al.*, 1999; D'Arcy *et al.*, 1999; Jenkins *et al.*, 1995; Longenecker, Garrard *et al.*, 2001; Prag *et al.*, 2003).

This work was supported by a grant from the National Science Foundation (MCB-0111053 to SJR) and a graduate training grant from the National Institutes of Health (GM-07759 to the Institute of Molecular Biology).

References

Anstrom, D. M., Kallio, K. & Remington, S. J. (2003). *Protein Sci.* **12**, 1822–1832.

Avbelj, F. & Fele, L. (1998). *J. Mol. Biol.* **279**, 665–684.

Baud, F. & Karlin, S. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 12494–12499.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bucher, M. H., Evdokimov, A. G. & Waugh, D. S. (2002). *Acta Cryst.* D**58**, 392–397.

Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. & Alzari, P. M. (2002). *Mol. Cell*, **10**, 757–768.

Camara-Artigas, A., Magee, C. L., Williams, J. C. & Allen, J. P. (2001). *Acta Cryst.* D**57**, 1281–1286.

Carugo, O. & Argos, P. (1997). *Protein Sci.* **6**, 2261–2263.

Charron, C., Kern, D. & Giegé, R. (2002). *Acta Cryst.* D**58**, 1729–1733.

Creamer, T. P. (2000). *Proteins*, **40**, 443–450.

Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J. & Derewenda, Z. S. (2004). *Acta Cryst.* D**60**, 275–280.

D'Arcy, A., Stihle, M., Kostrewa, D. & Dale, G. (1999). *Acta Cryst.* D**55**, 1623–1625.

Dale, G. E., Kostrewa, D., Gsell, B., Stieger, M. & D'Arcy, A. (1999). *Acta Cryst.* D**55**, 1626–1629.

Dale, G. E., Oefner, C. & D'Arcy, A. (2003). *J. Struct. Biol.* **142**, 88–97.

Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. (1997). *Proteins*, **28**, 494–514.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.

Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K. M., Evdokimov, A. G., Derewenda, Z. S. & Waugh, D. S. (2004). *Structure*, **12**, 301–306.

Derewenda, Z. S. (2004*a*). *Methods*, **34**, 354–363.

Derewenda, Z. S. (2004*b*). *Structure*, **12**, 529–535.

Devedjiev, Y., Surendranath, Y., Derewenda, U., Gabrys, A., Cooper, D. R., Zhang, R. G., Lezondra, L., Joachimiak, A. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **343**, 395–406.

Garrard, S. M., Longenecker, K. L., Lewis, M. E., Sheffield, P. J. & Derewenda, Z. S. (2001). *Protein Expr. Purif.* **21**, 412–416.

Graham, J. E. & Clark-Curtiss, J. E. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 11554–11559.

Guex, N. & Peitsch, M. C. (1997). *Electrophoresis*, **18**, 2714–2723.

Howard, B. R., Endrizzi, J. A. & Remington, S. J. (2000). *Biochemistry*, **39**, 3156–3168.

Hubbard, S. R., Hendrickson, W. A., Lambright, D. G. & Boxer, S. G. (1990). *J. Mol. Biol.* **213**, 215–218.

Iyer, G. H., Dasgupta, S. & Bell, J. A. (2000). *J. Cryst. Growth*, **217**, 429–440.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Janda, I., Devedjiev, Y., Cooper, D., Chruszcz, M., Derewenda, U., Gabrys, A., Minor, W., Joachimiak, A. & Derewenda, Z. S. (2004). *Acta Cryst.* D**60**, 1101–1107.

Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnicki, J., Cooper, D. R., Graf, P. C., Joachimiak, A., Jakob, U. & Derewenda, Z. S. (2004). *Structure*, **12**, 1901–1907.

Janin, J. & Rodier, F. (1995). *Proteins*, **23**, 580–587.

Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6057–6061.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W. V. & Harrison, P. M. (1991). *Nature (London)*, **349**, 541–544.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst.* D**57**, 679–688.

Longenecker, K. L., Lewis, M. E., Chikumi, H., Gutkind, J. S. & Derewenda, Z. S. (2001). *Structure*, **9**, 559–569.

Lorenz, M. C. & Fink, G. R. (2001). *Nature (London)*, **412**, 83–86.

McElroy, H. E., Sisson, G. W., Schoettlin, W. E., Aust, R. M. & Villafranca, J. E. (1992). *J. Cryst. Growth*, **122**, 265–272.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K. L., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst.* D**58**, 1983–1991.

Munshi, S., Hall, D. L., Kornienko, M., Darke, P. L. & Kuo, L. C. (2003). *Acta Cryst.* D**59**, 1725–1730.

Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* D**59**, 1028–1037.

Prag, G., Misra, S., Jones, E. A., Ghirlando, R., Davies, B. A., Horazdovsky, B. F. & Hurley, J. H. (2003). *Cell*, **113**, 609–620.

Qiu, X. & Janson, C. A. (2004). *Acta Cryst.* D**60**, 1545–1554.

Rost, B. & Sander, C. (1994). *Proteins*, **20**, 216–226.

Vekilov, P. G. (2003). *Methods Enzymol.* **368**, 84–105.

Wang, W. & Malcolm, B. A. (1999). *Biotechniques*, **26**, 680–682.

Zulauf, M. & D'Arcy, A. (1992). *J. Cryst. Growth*, **122**, 102–106.