BMC Biotechnology

RESEARCH ARTICLE Open Access

Surface patches on recombinant erythropoietin predict protein solubility: engineering proteins to minimise aggregation

M. Alejandro Carballo-Amador[1], Edward A. McKenzie[2], Alan J. Dickson[3] and Jim Warwicker[2*]

Abstract

Background: Protein solubility characteristics are important determinants of success for recombinant proteins in relation to expression, purification, storage and administration. *Escherichia coli* offers a cost-efficient expression system. An important limitation, whether for biophysical studies or industrial-scale production, is the formation of insoluble protein aggregates in the cytoplasm. Several strategies have been implemented to improve soluble expression, ranging from modification of culture conditions to inclusion of solubility-enhancing tags.

Results: Surface patch analysis has been applied to predict amino acid changes that can alter the solubility of expressed recombinant human erythropoietin (rHuEPO) in *E. coli*, a factor that has importance for both yield and subsequent downstream processing of recombinant proteins. A set of rHuEPO proteins (rHuEPO E13K, F48D, R150D, and F48D/R150D) was designed (from the framework of wild-type protein, rHuEPO WT, via amino acid mutations) that varied in terms of positively-charged patches. A variant predicted to promote aggregation (rHuEPO E13K) decreased solubility significantly compared to rHuEPO WT. In contrast, variants predicted to diminish aggregation (rHuEPO F48D, R150D, and F48D/R150D) increased solubility up to 60% in relation to rHuEPO WT.

Conclusions: These findings are discussed in the wider context of biophysical calculations applied to the family of EPO orthologues, yielding a diverse range of calculated values. It is suggested that combining such calculations with naturally-occurring sequence variation, and 3D model generation, could lead to a valuable tool for protein solubility design.

Keywords: Protein solubility, Protein aggregates, Inclusion bodies, Erythropoietin, Solubility prediction, Protein expression

Background

Biological systems have evolved by orchestration of molecular interactions, with proteins as key elements. In the circulatory system 2.4 million red blood cells are replaced every second in human adults [1]. This requires stable and efficient regulation to fulfil this demand. Erythropoietin (EPO), the main glycoprotein behind this task, regulates the growth and proliferation of red blood cell progenitors [2]. EPO is one of the top-selling therapeutics [3], providing therapy for millions of patients. There remains a continued demand to make EPO at large-scale and to increase the economic efficiency and, hence, availability to patients. Under this scheme, EPO has been a successful example of a biosimilar available in the market [3, 4]. Human erythropoietin (HuEPO) consists of 166 amino acid residues, in a structure that includes three N-linked glycosylation sites (N24, N38 and N83), an O-linked glycosylation (S126) and two disulphide bonds (C7-C161 and C29-C33) [5]. These complex post-translational modifications (PTMs) are the main challenge for expression of HuEPO in heterologous expression systems [6], and in the

* Correspondence: j.warwicker@manchester.ac.uk
[2]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
Full list of author information is available at the end of the article

cost-efficient *E. coli* system none of these PTMs are effectively incorporated in the cytoplasmic environment. However, the activity of a non-glycosylated version of HuEPO expressed in *E. coli* has been proved in in vitro cell proliferation assays [7]. Using bacteria to produce recombinant proteins efficiently entails a significant challenge since cysteine mispairing may lead to misfolding and low yields [8]. In order to overcome this challenge, engineered strains of *E. coli* have been developed such as the SHuffle system [9]. In this *E. coli* strain the cytoplasmic environment is altered by the overexpression of DsbC disulphide bond isomerase, and by deletion of two reductases (glutaredoxin [gor] and thioredoxin [trxB]).

Glycosylation of HuEPO contributes ~ 40% of the overall molecular mass, improving stability and solubility of the molecule as a therapeutic [10–12]. Lack of glycosylation in recombinant HuEPO (rHuEPO) derived from *E. coli* leads to aggregation during expression and, potentially, during subsequent purification, storage and delivery [13]. This protein aggregation phenomenon during expression in *E. coli* leads to incorporation into inclusion bodies (IBs), from the interactions of partially folded, misfolded or unfolded recombinant proteins in the cytoplasm [8]. Surface charge engineering has been illustrated by mutation of the three N-glycosylation sites on HuEPO to lysine (N24K, N38K and N83K), increasing net charge. This engineering decreased IB formation and facilitated the purification of protein to provide the rHuEPO crystal structure [13, 14].

Protein aggregation can involve chemical aggregation, such as disulphide bond formation, and/or physical aggregation, such as non-covalent interactions between hydrophobic surfaces [15]. Several approaches have been undertaken to diminish hydrophobic patches on the surface to prevent protein aggregation [16–23]. In this context, improved expression, stability and solubility of rHuEPO and granulocyte colony-stimulating factor (G-CSF) has been generated by application of the in vitro ribosome display technique, in combination with three parallel selection pressures (reducing agent, elevated temperature and hydrophobic interaction chromatography matrices) [24]. In the case of rHuEPO, a variant encoding four mutations resulted in a form that was less prone to aggregation [24]. Furthermore, the application of fusion tags to improve rHuEPO solubility has been successful [7, 25]. Some of these fusion partners, including NusA and maltose-binding protein (MBP), have large negatively-charged areas that may be involved in promotion of folding of the target protein by limiting protein aggregation [26]. Engineering of negatively-charged areas on protein surfaces is gaining strength as an approach for increasing solubility [27–31].

Here we report a novel experimental approach targeted at improvement of rHuEPO solubility for expression in *E. coli*, following the observation that soluble expression of proteins is inversely correlated with the size of the largest positively-charged patch on the protein surface [29, 32]. This result is based on data from cell-free expression [33], and here we test the hypothesis by mapping surface charge of rHuEPO, focusing on modulation of positively-charged patches through mutagenesis. A set of mutants has been generated, ranging from more (rHuEPO E13K) to less positively-charged surface patches (rHuEPO F48D, R150D and F48D/R150D), compared to natural (wild type) rHuEPO (rHuEPO WT). Experimental results support the prediction, i.e. largest positively-charged patch size correlated with the degree of protein aggregation in the cytoplasm of *E. coli*. Further application of this approach, particularly in the context of natural protein surface variation, could improve the rational design of proteins with enhanced solubility in cytoplasmic expression.

Results

Redesign of rHuEPO WT for altered charge surface

A published algorithm [29] was used to identify amino acids of rHuEPO for which mutation could be predicted to alter solubility (Table 1). The method is based on an observed correlation between positive charge patches and insolubility [29] for data derived in a cell-free expression system [33]. Design for improved solubility, therefore, involved identification and reduction of the larger positively-charged patch. For the protein variants shown in Table 1, the substitution R150D gave a lowered positive patch size and was predicted to be more soluble than rHuEPO WT. In contrast, substitution E13K had an increased positive patch compared with wild type and was predicted to generate a less soluble product. Both of these sites lie on the protein surface. A third site was chosen to introduce a negative charge (F48D), rather than make a charge swap, and decrease the size of the largest positive patch. It was recognized that although F48 was also on the surface, this mutation might present

Table 1 Predicted solubilities of recombinant human erythropoietin. Solubility profile was defined as described in Chan et al. [29]. Positive patch sizes are divided by that best separating soluble and insoluble datasets [33], above 1.0 implies predicted insolubility

Protein	Pos patch ratio to threshold	Prediction
rHuEPO wild-type	1.49	Insoluble
rHuEPO F48D	0.75	Soluble
rHuEPO R150D	0.61	Soluble
rHuEPO F48D/R150D	0.47	Soluble
rHuEPO E13K	2.47	Insoluble

a more challenging mutation structurally since the phenylalanine ring covers in part the hydrophobic sidechains of V46 and L155. Figure 1 shows the single site mutations and one double site mutation employed in this study, and their charge surfaces in comparison with wild type rHuEPO. Amino acid conservation analysis [34–36] showed that R150 is relatively conserved across evolution, but E13 and F48 showed less conservation (see Additional file 1: Figure S1). Previous site-directed mutagenesis of these residues had shown no alteration of the folded state of rHuEPO structure [37].

Soluble expression of rHuEPO is modulated in line with the predictions

The expression and solubility of rHuEPO variants were studied under low induction conditions (Fig. 2). The total expression of rHuEPO was approximately equivalent across WT and mutant constructs, particularly in the SHuffle system (Fig. 2d). Protein solubility, assessed as the ratio of EPO detected in soluble and total EPO fraction, agrees with the prediction for all 4 mutant constructs in the SHuffle system, but for only 2 of the 4 variants in the BL21 system, the exceptions being those involving mutation at F48 i.e. rHuEPO F48D and F48D/R150D (Fig. 2c).

That the two-way charge swap tests (R150D and E13K) match predictions in both expression systems used is encouraging in terms of applying a correlation [29] learned from a large dataset, albeit in cell-free expression [33]. With regard to the general mechanism through which charge may play a role in determination of solubility, there is an increasing body of work that indicates that negatively-charged residues (rather that positively-charged residues) are more favourable for protein solubility [27, 30, 38–41]. Whatever the molecular mechanism by which the engineering of charged patches alters solubility, it may be associated with attainment of the native, or a near native state. This would be consistent with the observation here that charge-based predictions are matched for SHuffle but not with the BL21 system. Expression in the SHuffle system will, in relative terms, favour correct formation of the two disulphide bonds in the folded recombinant protein. The F48D mutation, apart from introducing negative charge, may expose more non-polar surface than it removes, due to the sidechains of V46 and L155 that lie beneath F48 (Fig. 3). It would be expected that the BL21 strain should be more susceptible than the SHuffle strain to this exposure, since it would be less able to refold partially unfolded protein. This difference may underpin the solubility data for F48D and F48D/R150D (Fig. 2c). Although the aspartic acid sidechain introduced in the F48D mutant could in principle alter pH-dependent properties, protein expression and measurements are made at neutral pH, for which a negative charge will be carried by the F48D carboxylate group.

Bioinformatics and solubility engineering

Having found that charge surface properties, and charge contributions to folded state stability, are factors that commonly arise in experimental studies aimed at improvement of EPO production, we assessed how these properties varied between EPO orthologues. The largest positively-charged patches, and the predicted contribution of ionisable group charge interactions to folded state stability at pH 7, were calculated for 115 EPO homologues found through a BLAST [42] search, and passed through a comparative modelling pipeline. Positive patches are distributed over a surprisingly large range (viewed as

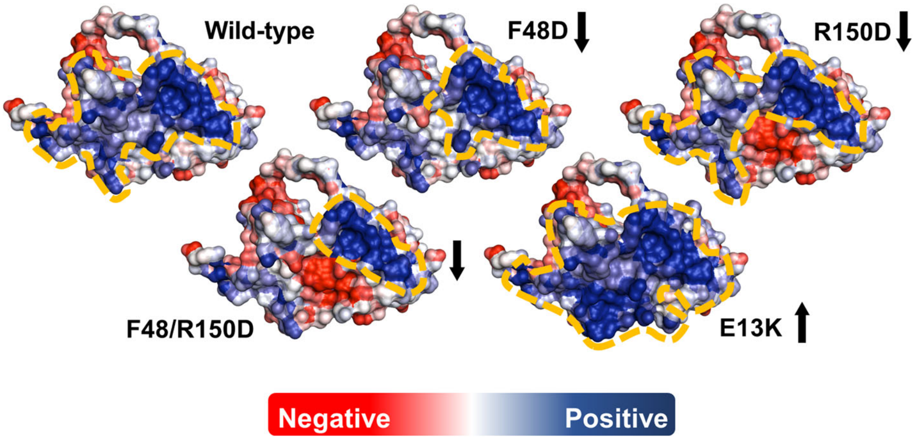

Fig. 1 HuEPO wild-type and variants surface illustration showing the electrostatic potential patches [29]. Amino acids in positive patches are represented by blue, non-charged patches by white and negatively charged by red colour, respectively, with dashed yellow contours drawn in to delineate the largest positive patches

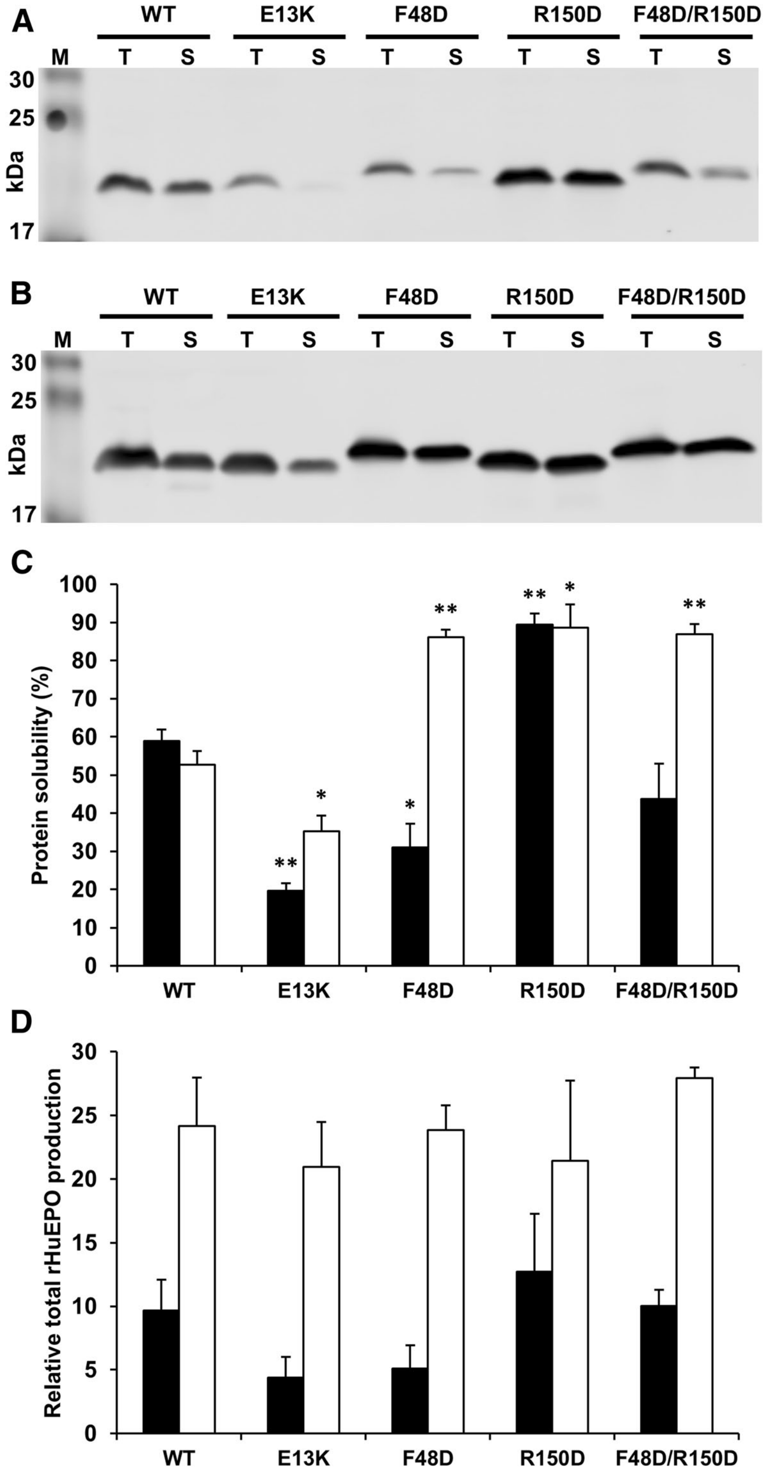

Fig. 2 Western blot of rHuEPO expression and solubility. (**a**-**b**) Equal volumes of total protein and soluble fraction were probed with Mouse anti-polyHis antibody and imaged with the Odyssey Imaging System for BL21 (DE3) pLysS (**a**) and SHuffle (**b**) strains. (**c**) Experimental solubility was determined by the distribution of rHuEPO between soluble and inclusion body fraction in *E. coli*. (**d**) Relative total rHuEPO production (arbitrary units). Error bars represent the + SEM for measurements in triplicate; statistical significance was calculated using a two-sided unpaired t-test (**P* < 0.05, ** *P* < 0.01). BL21 (DE3) pLysS (■); SHuffle (□)

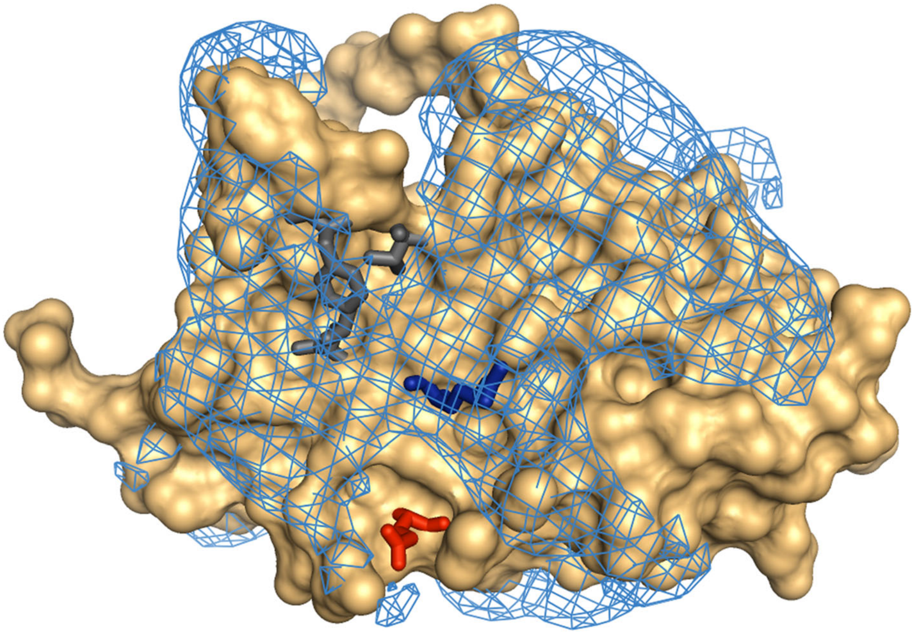

Fig. 3 Mutated residues on the surface of HuEPO. Molecular surface is shown (orange), with positive electrostatic field contoured at 30 mV (blue mesh). Residues E13 (red), F48 (grey) and R150 (blue) are shown in sticks, rather than surface representation. Also drawn are the non-polar residues V46 and L155 that are covered by F48 in the WT structure

a distribution or as individual examples, Fig. 4). Since it is not an evolutionarily conserved property, positive patch size presumably only becomes relevant for protein expression at the higher levels of over-expression, in comparison with expression in nature. As a general consideration, for over-expression, the range of values permissible naturally could present an important design feature for protein production (especially in relation to development of novel format species). A heat map, clustered according to pairwise sequence identity in Clustal [43], shows that positive patch sizes tend to cluster together, with occasional larger transitions (Fig. 4a). This indicates that a small number of mutations, or even a single mutation as in the current study, can significantly alter the charge distribution of a protein.

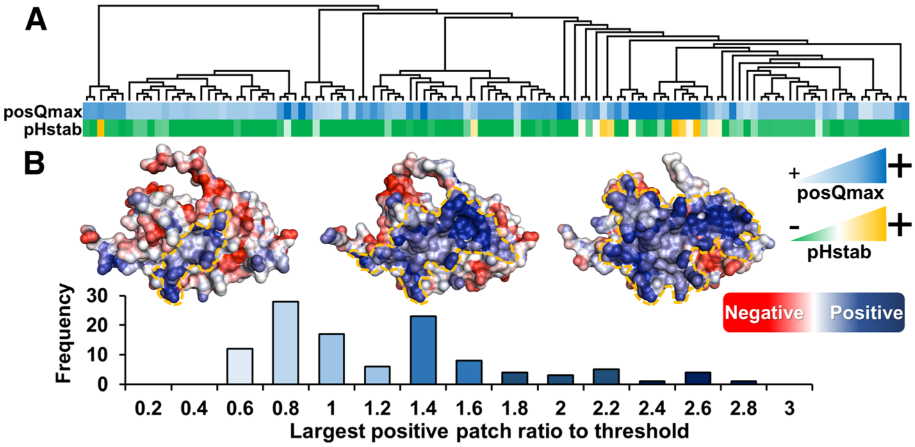

Fig. 4 Positively-charged patches and predicted stability in EPO orthologues. (**a**) A sequence-based phylogenetic tree is combined with positive patch and stability calculations. Colour coding for posQmax varies from lighter to darker blue as the calculated largest positive patch increases for an EPO orthologue (same colour code for histogram in panel **b**). Predicted pH-dependent contribution to folded state stability (pHstab) varies from yellow (positive, unfavourable) to green (negative, favourable). (**b**) Distribution of largest positive patch ratios to the threshold for the 115 EPO orthologues, showing that surface charge changes substantially. HuEPO is located roughly in the centre of the distribution. Frequency is the number of EPO orthologues having a largest positive patch ratio to threshold, in the given x-axis bin

Discussion

When expressed in *E. coli* rHuEPO WT tended to form insoluble protein aggregates in inclusion bodies [14]. The use of sub-optimal temperature and lower inducer concentrations has been argued to be a more appropriate approach to assess protein folding (and hence solubility) of recombinant proteins in *E. coli* [44]. Low induction conditions (decreased temperature, lower IPTG challenge) decreased cell growth and the elongation rate of translation [45] and induced chaperone activity and protein folding capability [46]. These responses led to better folding and less degradation [47] and offered a more refined system to investigate the consequences of variant structures on the expression and solubility of recombinant protein expression in *E. coli.*

Our understanding of IB formation and subsequent recovery of native protein is increasing [48, 49]. Important factors include the effects of a particular protein on metabolic burden in *E. coli* [50], chaperone and codon optimisation [51, 52], particular genetic loci [53], and solubility tag and medium engineering of the expression system [54]. However, there are still unknowns in what dictates IB formation for particular proteins. With regard to solubility in cell-free expression [33], a model was put forward [29] as one potential explanation of the correlation, involving charge interactions between the positive charge surface of a protein and large concentrations of a biological macro-anion i.e. mRNA. The current work does not address the underlying model. Other studies have highlighted the importance of charge properties in the production of EPO, from the incorporation of three lysines to remove the N-linked glycosylation sites and modify pI [14], to recent work in which the contribution of positively-charged patches was highlighted [55]. In the latter case, it was reasoned that improved solubility properties associated with increased ionic strength could be a result of lessening of repulsion within clusters of basic sidechains. The positively-charged region is the same as that studied here. However, the rationale for modification of aggregation properties is different, i.e. native state stabilization [55] compared with reduction of self-association (current work). The systems are quite different, purified protein [55] compared with cytoplasmic expression in *E. coli* (current work), and the suggested molecular mechanisms in each case remain unproven at this stage. It is intriguing that the same positively-charged region on EPO has featured in a predictive study (current work) and in a post-hoc rationalization [55]. Further, this patch has also been implicated in a selection mutagenesis screen for EPO variants with improved stability [24]. Ribosome display led to a variant that incorporated mutations at 4 sites, and exhibited a decreased aggregation in accelerated shelf-life studies. One of the mutations in that variant was G158E, introducing a negative charge into the largest positively-charged patch [24].

The ionic strength dependence study of Banks (2015) implied that charge-charge interactions are delicately balanced in EPO (at pH 6.9), and thus predicted ionisable group contributions to folded state stability are also shown in the heat map format (at pH 7.0). It is relatively rare for proteins to have a predicted contribution (or indeed a measured contribution) that is net unfavourable as shown for a few examples here (Fig. 4a). This is likely to be correlated with the observation that increasing ionic strength improves the folded state stability for EPO, since unfavourable charge-charge interactions will be diminished [55]. Again, a relatively simple bioinformatics calculation shows molecular detail that is likely to underpin observed changes in EPO production, and furthermore it can guide design for improved production.

Importantly, no account has been taken of glycosylation in these interpretations, since it is difficult to model the conformations of these components with respect to the protein. The net negative charge carried by the glycosyl groups is emphasized by the change in pI of EPO from 9.2 to 4.4 upon glycosylation [10]. Interactions between sialic acid groups and positive charges on the protein are likely to improve the stabilizing component of the charge-charge balance. It should be emphasized that when charge-charge interactions contribute relatively little in net terms at neutral pH, this is generally the result of favourable and unfavourable terms cancelling, not due to a complete absence of ionisable group interactions. Such a situation lends itself to modification by relatively small changes in amino acid sequence (depending on location of those changes in the structure). In the heat map for EPO homologue stability contributions from charge-charge interactions (Fig. 4a), similar contributions generally cluster together (within the sequence-based phylogenetic tree), but there are also some abrupt changes, emphasizing the sequence-sensitivity. In general terms, the least stable contributions tend to group with the most positive patches (and vice versa). This may support the rationalization given for improved resistance to aggregation as ionic strength is increased for pure EPO [55], in that larger positively-charged patches are predicted to associate with less stable folded proteins.

Conclusions

Solubility prediction for proteins has been the focus of several groups in the last 25 years [38–41, 56–59]. Here, a method developed in our group [29] has been tested. Mutations of rHuEPO gave experimental results in line with predictions, excepting F48D in one of the two expression systems used. The F48D mutation stands apart from the other (charge swap) variants, in that it is likely to alter the hydrophobicity. It is therefore plausible that

F48D leads to reduced folding efficiency, a suggestion consistent with an improved solubility in the SHuffle system, compared with reduced solubility in the BL21 strain. Further purification might allow a better structural understanding of these mutations, however, native purification has not been possible [7, 11, 13, 14]. Finding that the region of EPO targeted here also appears in selection mutagenesis for improved stability [24] and a recent study of ionic strength effects [55], bioinformatics was employed to reveal the extent of variation in EPO homologues. Interestingly, both the largest positively-charged patch and the predicted contribution of ionisable group interactions to stability (pH 7) vary substantially through evolution, suggesting that the solubility of these EPO homologues, if over-expressed, would be divergent. Indeed, even small changes in amino acid sequence can lead to relatively large changes in solubility. Whilst it may not always be feasible to engineer a human protein across species (e.g. considering immunogenicity of a therapeutic protein), biophysical calculations for a set of homologues could guide design of protein with enhanced stability and solubility for large-scale expression, if incorporated at a sufficiently early point in the design cycle.

Methods

rHuEPO solubility profile and mutant design

The PyMOL Molecular Graphics System version 1.3 [60] and Swiss-PdbViewer 4.0.1 [61] were used to analyse rHuEPO structural and sequence features. Protein solubility predictions were calculated using an algorithm developed in our group [29] (see Additional file 2: Table S1 and Additional file 3: Table S2). The algorithm computes structured-based parameters, including the sizes of positively- and negatively-charged patches, when the electrostatic potential field is contoured at + 25 mV or − 25 mV. It also gives the size of the largest patch for which all points are between − 25 mV and + 25 mV (effectively non-charged). The principal prediction [29] is generated from the ratio of the largest positively-charged patch to a threshold, but a supplemental prediction from the combination of positive and non-charged patches is also made. Thresholds were calculated from a dataset of experimental solubilities determined for cell-free expression of *E. coli* proteins [33], as the value of a parameter that best separates less and more soluble proteins. Proteins with larger positive patches are predicted as less soluble, and have a ratio to threshold above 1.0. Where the ratio to threshold is below 1.0, a protein is predicted as soluble. Current work concentrates on positive patches (posQ), since this structure-based feature gives the best separation [29]. Substitutions to modify posQ, based on the protein data bank [62] file 1EER [63] were carried out in Swiss-PdbViewer, and the resulting structures analysed for modification of the largest positive patch (Table 1). This in silico mutational screening gave the following candidate mutations: rHuEPO E13K, rHuEPO F48D, rHuEPO R150D and the double mutant rHuEPO F48D/R150D. Aspartic acid was selected for the introduction of negative charge, in preference to glutamic acid, due to its shorter side chain, lowering the possibility of non-specific interactions with surrounding side chains. Calculations were performed using a modified crystal structure of an analogue rHuEPO taken from the 1EER PDB entry in order to maintain consistency with our experimental and native rHuEPO cDNA (K24 N, K38 N, K83 N, N121P and S122P).

Bioinformatics analysis of rHuEPO surface and stability

Multiple sequence alignments and surface mapping coloured by residue conservation were performed using ConSurf with default parameters [34–36], using the UniRef90 database [64], which removes redundancy at 90% sequence identity (see Additional file 1: Figure S1). A separate structural study of EPO homologues was made with 115 models generated from a Clustal alignment [43] using a sidechain replacement method [65] for comparative modelling. Input to the Clustal alignment was from a search for EPO orthologues with BLAST [42], followed by manual checking of EPO annotation. Patch calculations were made based on the comparative models, to give a view of EPO variation over species.

The set of EPO comparative models was also used to estimate the contribution of ionisable groups (i.e. the pH-dependent stability term) to folded state stability at pH 7. A Debye-Hückel model for interactions between ionisable group charges, at 0.15 M ionic strength was used, with Monte Carlo sampling of protonation states to derive the pH-dependent term. This modelling follows previous methodology [66], and includes subtraction of an estimated unfolded state set of interactions to arrive at the predicted pH-dependent term for folded state compared with unfolded state.

Construction of rHuEPO mutants and expression vectors

Human erythropoietin cDNA was amplified from a pre-existing mammalian expression vector by applying primers containing the restriction sites 5′-*BamHI* and 3′-*EcoRI*. The PCR fragment (lacking signal peptide) was subcloned into a pHis vector. This plasmid is a modified version of the commercial pET-16b vector (Novagen). The gene sequence for each plasmid was as follows: 5′-6xHis-Thrombin cleavage site-*BamHI*-rHuEPO-*EcoRI*-3′. rHuEPO mutations were introduced using the GENEART Site-Directed Mutagenesis System with the enzyme AccuPrime *Pfx* (Invitrogen).

Protein expression and solubility assay

The bacterial cell lines used in this study were *Escherichia coli* BL21 (DE3) pLysS and SHuffle (New England BioLabs). Bacterial strains were transformed with the pHis-rHuEPO plasmids. Transformed cells were grown overnight in 5 ml working volume of Luria-Bertani (LB) medium (10 g tryptone, 5 g yeast extract, 5 g NaCl) containing 100 μg/ml ampicillin at 37 °C with shaking at 220 rpm. In addition, BL21 (DE3) pLysS were grown in the presence of chloramphenicol (50 μg/ml) in order to preserve the pLysS plasmid. On the following day, 1 ml of pre-culture was transferred to 50 ml 2% (v/v) LB supplemented with 2% (w/v) glucose with 100 μg/ml ampicillin in 250 ml shake flasks. Experiments were performed in triplicate. Shake flasks were incubated at a constant temperature of 25 °C, with shaking at 180 rpm. Bacteria were grown to an OD_{600} of approximately 0.6–0.8. Protein expression was induced by the addition of IPTG (0.05 mM, final). After 5 h, cultures were centrifuged at 6500 g for 15 min at 4 °C. Bacterial pellets were suspended in 5 ml of lysis buffer (25 mM Tris pH 7.5, 150 mM NaCl, 1% [v/v] Triton X-100) and were stored at − 20 °C until future use. The cells were disrupted by six sonication cycles of 30 s at 20% amplitude and then allowed to cool for 30 s on ice water bath. Separation of soluble and total fractions was performed by centrifugation at 18,000 g for 30 min at 4 °C of 1 ml of each sample from the whole cell lysate. The supernatants were collected and handled as the soluble fraction. An additional 1 ml of each lysate sample was processed as the total fraction, and rHuEPO solubility was calculated by densitometric ratio of protein detected in soluble and total fractions.

Proteins in soluble and total fractions were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) with 12% (w/v) acrylamide using the Mini-PROTEAN Tetra Cell (BioRad). Samples containing equal volumes (20 μl) of protein were subjected to heat at 95 °C for 5 min in 6x denaturing buffer (375 mM Tris pH 6.8, 12% [w/v] SDS, 60% [v/v] glycerol, 0.06% [w/v] bromophenol blue, 5.5% [v/v] β-mercaptoethanol). Separated proteins were transferred to nitrocellulose membranes using a transblot semi-dry transfer cell (Bio-Rad) at 15 V for 45 min. Membranes were blocked overnight for non-specific binding in blocking buffer (5% [w/v] skimmed milk in TBS-Tween pH 7.4) at 4 °C with shaking. For detection of rHuEPO, a mouse anti-polyHis antibody (Sigma) was diluted 1:5000 in blocking buffer solution and the membrane was incubated in this solution for 2 h at room temperature with agitation. After three washes in TBS-Tween (5 min each time), samples were incubated with an IR-labelled secondary Donkey anti-Mouse IgG antibody (LI-COR) diluted 1:15000 in blocking buffer solution at room temperature for 45 min. Following incubation, the secondary antibody was removed and the membrane was washed three times. For IR detection, blots were imaged with the Odyssey Imaging System. Bands were quantified in Image Studio Lite software (LI-COR).

Additional files

Additional file 1: Figure S1. Multiple alignment of HuEPO. (A) A surface map is coloured by residue conservation scores [34–36]. The image was rendered using PyMOL [60]. (B) Panel shows the same color-coding for conservation show in panel (A), but here applied to the amino acid sequence of rHuEPO. (PDF 1525 kb)

Additional file 2: Table S1. Positively-charged patches size profile of rHuEPO WT from the charged patch calculator. Complete screening of posQ ratio scores for the modified rHuEPO WT (PDB: 1EER) is shown. The largest positive patches are represented by blue (ratio > 1.0). Those proteins with ratio above 1.0 are predicted as insoluble and below 1.0 as soluble. The three targeted residues in this study are highlighted in red. Ratio: largest positively-charged patch (posQ) value from the charged patch calculator [29]. Charge patches: HYD, hydrophobic (non-charged); NEG, negatively-charged; POS, positively-charged. (PDF 478 kb)

Additional file 3: Table S2. Summary of the solubility screening of rHuEPO. Left column shows the complete mutational screening of all positive charge amino acids (i.e. arginine and lysine) within the largest positively-charged patch (posQ) for aspartic acid (D). Next two columns summarize a set of substitutions of any amino acid in the posQ for D. The column on the right shows all the negative charge residues (i.e. aspartic and glutamic acid) within the posQ for arginine or lysine. Those proteins with posQ ratio above 1.0 are predicted as insoluble and below 1.0 as soluble. Selected proteins for further site-directed mutagenesis are highlighted in red. (PDF 168 kb)

Abbreviations

EPO: Erythropoietin; HuEPO: Human erythropoietin; IBs: Inclusion bodies; posQ: positive patches; PTMs: Post-translational modifications; rHuEPO: Recombinant human erythropoietin; WT: Wild-type protein

Acknowledgements

We thank Professor Stephen High and Dr. E. Ceh-Pavia for inspiring discussions.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) via a BioProNet award (BB/L013770/1), and a PhD studentship by CONACyT to MAC (309167). The funders had no role in the study design, or in the collection, analysis and interpretation of data, or in writing the manuscript.

Availability of data and materials

Data, analysis algorithms, and materials are available from the corresponding author on reasonable request.

Authors' contributions

MAC carried out the experiments, analysed all data, and wrote the first draft of the manuscript. EAM participated in its design and supervision. AJD and JW conceived the study, participated in its design and supervision, and finalised the manuscript with MAC. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Facultad de Ciencias, Universidad Autónoma de Baja California, Km. 103 Carretera Tijuana–Ensenada, Pedregal Playitas, 22860 Ensenada, Baja California, Mexico. [2]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. [3]Faculty of Science and Engineering, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK.

Received: 23 January 2019 Accepted: 30 April 2019
Published online: 09 May 2019

References

1. Sackmann E: Biological membranes architecture and function. In: Handbook of Biological Physics. Edited by Lipowsky R, Sackmann E, vol. Volume 1. Amsterdam: Elsevier; 1995: 1–63.
2. Krantz SB. Erythropoietin. Blood. 1991;77(3):419–34.
3. Walsh G. Biopharmaceutical benchmarks 2018. Nat Biotechnol. 2018;36:1136.
4. Goldsmith D, Dellanna F, Schiestl M, Krendyukov A, Combe C. Epoetin Biosimilars in the treatment of renal Anemia: what have we learned from a decade of European experience? Clinical drug investigation. 2018;38(6):481–90.
5. Lai PH, Everett R, Wang FF, Arakawa T, Goldwasser E. Structural characterization of human erythropoietin. J Biol Chem. 1986;261(7):3116–21.
6. Skibeli V, Nissen-Lie G, Torjesen P. Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. Blood. 2001; 98(13):3626–34.
7. Jeong TH, Son YJ, Ryu HB, Koo BK, Jeong SM, Hoang P, Do BH, Song JA, Chong SH, Robinson RC, et al. Soluble expression and partial purification of recombinant human erythropoietin from E. coli. Protein Expr Purif. 2014;95: 211–8.
8. Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des. 1998;3(1):9–23.
9. Lobstein J, Emrich C, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Factories. 2012;11(1): 56.
10. Davis JM, Arakawa T, Strickland TW, Yphantis DA. Characterization of recombinant human erythropoietin produced in Chinese hamster ovary cells. Biochemistry. 1987;26(9):2633–8.
11. Narhi LO, Arakawa T, Aoki KH, Elmore R, Rohde MF, Boone T, Strickland TW. The effect of carbohydrate on the structure and stability of erythropoietin. J Biol Chem. 1991;266(34):23022–6.
12. Banks DD. The effect of glycosylation on the folding kinetics of erythropoietin. J Mol Biol. 2011;412(3):536–50.
13. Cheetham JC, Smith DM, Aoki KH, Stevenson JL, Hoeffel TJ, Syed RS, Egrie J, Harvey TS. NMR structure of human erythropoietin and a comparison with its receptor bound conformation. Nat Struct Biol. 1998;5(10):861–6.
14. Narhi LO, Arakawa T, Aoki K, Wen J, Elliott S, Boone T, Cheetham J. Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of Escherichia coli-derived erythropoietin. Protein Eng. 2001;14(2):135–40.
15. Wang W. Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm. 2005;289(1–2):1–30.
16. Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R. Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. Proc Natl Acad Sci U S A. 1995;92(13):6057–61.
17. Li Y, Yan Y, Zugay-Murphy J, Xu B, Cole JL, Witmer M, Felock P, Wolfe A, Hazuda D, Sardana MK, et al. Purification, solution properties and crystallization of SIV integrase containing a continuous core and C-terminal domain. Acta Crystallogr D Biol Crystallogr. 1999;55(Pt 11:1906–10.
18. Das D, Georgiadis MM. A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase. Protein Sci. 2001; 10(10):1936–41.
19. Slovic AM, Summa CM, Lear JD, DeGrado WF. Computational design of a water-soluble analog of phospholamban. Protein Sci. 2003;12(2):337–48.
20. Fan D, Li Q, Korando L, Jerome WG, Wang J. A monomeric human apolipoprotein E carboxyl-terminal domain. Biochemistry. 2004;43(17):5055–64.
21. Lawson AJ, Walker EA, White SA, Dafforn TR, Stewart PM, Ride JP. Mutations of key hydrophobic surface residues of 11 beta-hydroxysteroid dehydrogenase type 1 increase solubility and monodispersity in a bacterial expression system. Protein Sci. 2009;18(7):1552–63.
22. Andersen TCB, Lindsjø K, Hem CD, Koll L, Kristiansen PE, Skjeldal L, Andreotti AH, Spurkland A. Solubility of recombinant Src homology 2 domains expressed in E. coli can be predicted by TANGO. BMC Biotechnol. 2014;14(1):3.
23. Jetha A, Thorsteinson N, Jmeian Y, Jeganathan A, Giblin P, Fransson J. Homology modeling and structure-based design improve hydrophobic interaction chromatography behavior of integrin binding antibodies. mAbs. 2018;10(6):890–900.
24. Buchanan A, Ferraro F, Rust S, Sridharan S, Franks R, Dean G, McCourt M, Jermutus L, Minter R. Improved drug-like properties of therapeutic proteins by directed evolution. Protein Eng Des Sel. 2012;25(10):631–8.
25. Ahn JH, Keum JW, Kim DM. Expression screening of fusion partners from an E. coli genome for soluble expression of recombinant proteins in a cell-free protein synthesis system. PLoS One. 2011;6(11):e26875.
26. Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expr Purif. 2004;36(2):207–16.
27. Trevino SR, Scholtz JM, Pace CN. Amino acid contribution to protein solubility: asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. J Mol Biol. 2007;366(2):449–60.
28. Perchiacca JM, Ladiwala AR, Bhattacharya M, Tessier PM. Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. Protein Eng Des Sel. 2012;25(10):591–601.
29. Chan P, Curtis RA, Warwicker J. Soluble expression of proteins correlates with a lack of positively-charged surface. Sci Rep. 2013;3:3333.
30. Chong SH, Ham S. Interaction with the surrounding water plays a key role in determining the aggregation propensity of proteins. Angew Chem Int Ed Engl. 2014;126(15):4042–5.
31. Laber JR, Dear BJ, Martins ML, Jackson DE, Divenere A, Gollihar JD, Ellington AD, Truskett TM, Johnston KP, Maynard JA. Charge shielding prevents aggregation of supercharged GFP variants at high protein concentration. Mol Pharm. 2017;14(10):3269–80.
32. Hussain H, Fisher DI, Roth RG, Mark Abbott W, Carballo-Amador MA, Warwicker J, Dickson AJ. A protein chimera strategy supports production of a model "difficult-to-express" recombinant target. FEBS Lett. 2018;592(14): 2499–511.
33. Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T, Taguchi H. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. Proc Natl Acad Sci U S A. 2009; 106(11):4201–6.
34. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010;38(Web Server issue):W529–33.
35. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr J Chem. 2013;53(3–4):199–206.
36. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016; 44(W1):W344–50.
37. Elliott S, Lorenzini T, Chang D, Barzilay J, Delorme E. Mapping of the active site of recombinant human erythropoietin. Blood. 1997;89(2):493–502.
38. Kuntz ID. Hydration of macromolecules. III. Hydration of polypeptides. JACS. 1971, 93(2):514–6.
39. Collins KD, Washabaugh MW. The Hofmeister effect and the behaviour of water at interfaces. Q Rev Biophys. 1985;18(4):323–422.
40. Collins KD. Charge density-dependent strength of hydration and biological structure. Biophys J. 1997;72(1):65–76.
41. Trevino SR, Scholtz JM, Pace CN. Measuring and increasing protein solubility. J Pharm Sci. 2008;97(10):4155–66.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
43. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

44. Sevastsyanovich Y, Alfasi S, Overton T, Hall R, Jones J, Hewitt C, Cole J. Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. FEMS Microbiol Lett. 2009;299(1):86–94.
45. Farewell A, Neidhardt FC. Effect of temperature on in vivo protein synthetic capacity in Escherichia coli. J Bacteriol. 1998;180(17):4704–10.
46. Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN, Timmis KN. Chaperonins govern growth of Escherichia coli at low temperatures. Nat Biotechnol. 2003;21(11):1266–7.
47. Chesshyre J, Hipkiss A. Low temperatures stabilize interferon α-2 against proteolysis in Methylophilus methylotrophus and Escherichia coli. Appl Microbiol Biotechnol. 1989;31(2):158–62.
48. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Factories. 2015;14(1):1–10.
49. Qi X, Sun Y, Xiong S. A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. Microb Cell Factories. 2015;14(1):1–12.
50. Rahmen N, Fulton A, Ihling N, Magni M, Jaeger K-E, Büchs J. Exchange of single amino acids at different positions of a recombinant protein affects metabolic burden in Escherichia coli. Microb Cell Factories. 2015;14(1):1–18.
51. Itkonen JM, Urtti A, Bird LE, Sarkhel S. Codon optimization and factorial screening for enhanced soluble expression of human ciliary neurotrophic factor in Escherichia coli. BMC Biotechnol. 2014;14(1):92.
52. Wang Y, Li Y-Z. Cultivation to improve in vivo solubility of overexpressed arginine deiminases in Escherichia coli and the enzyme characteristics. BMC Biotechnol. 2014;14(1):1–10.
53. Pandey N, Sachan A, Chen Q, Ruebling-Jass K, Bhalla R, Panguluri KK, Rouviere PE, Cheng Q. Screening and identification of genetic loci involved in producing more/denser inclusion bodies in Escherichia coli. Microb Cell Factories. 2013;12(1):1–12.
54. Zhou K, Zou R, Stephanopoulos G, Too H-P. Enhancing solubility of deoxyxylulose phosphate pathway enzymes for microbial isoprenoid production. Microb Cell Factories. 2012, 11(1):1–8.
55. Banks DD. Nonspecific shielding of unfavorable electrostatic intramolecular interactions in the erythropoietin native-state increase conformational stability and limit non-native aggregation. Protein Sci. 2015;24(5):803–11.
56. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational Design of Protein Mutants with enhanced solubility. J Mol Biol. 2015;427(2): 478–90.
57. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein–sol: a web tool for predicting protein solubility from sequence. Bioinformatics. 2017;33(19):3098–100.
58. Matsui D, Nakano S, Dadashipour M, Asano Y. Rational identification of aggregation hotspots based on secondary structure and amino acid hydrophobicity. Sci Rep. 2017;7(1):9558.
59. Wolf Pérez AM, Sormanni P, Andersen JS, Sakhnini LI, Rodriguez-Leon I, Bjelke JR, Gajhede AJ, De Maria L, Otzen DE, Vendruscolo M, Lorenzen N. In vitro and in silico assessment of the developability of a designed monoclonal antibody library. mAbs. 2018:1–13.
60. Schrödinger LLC: The PyMOL Molecular Graphics System, Version 1.3r1. In.; 2010.
61. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997; 18(15):2714–23.
62. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–42.
63. Syed RS, Reid SW, Li C, Cheetham JC, Aoki KH, Liu B, Zhan H, Osslund TD, Chirino AJ, Zhang J, et al. Efficiency of signalling through cytokine receptors depends critically on receptor orientation. Nature. 1998;395(6701):511–6.
64. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007;23(10):1282–8.
65. Cole C, Warwicker J. Side-chain conformational entropy at protein–protein interfaces. Protein Sci. 2002;11(12):2860–70.
66. Chan P, Warwicker J: Evidence for the adaptation of protein pH-dependence to subcellular pH. BMC Biol 2009, 7:69–69.