Journal of Computer-Aided Molecular Design, **16:** 501–509, 2002.
KLUWER/ESCOM
© 2002 *Kluwer Academic Publishers. Printed in the Netherlands.*

Theoretical descriptors for the quantitative rationalisation of plastocyanin mutant functional propertiess

F. De Rienzo[1,3], G.H. Grant[2] & M.C. Menziani[1,*]
[1]*Dipartimento di Chimica, Universita' di Modena e Reggio Emilia, Via Campi, 183, 41100 Modena, Italy;* [2]*Physical & Theoretical Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QZ, UK;* [3]*INFM-S, Modena, Italy*

Key words: functional properties, plastocyanin, point mutations, quantitative structure-activity relationships, redox-proteins, theoretical descriptors.

Summary

A quantitative rationalisation of the effect of specific amino acids on the recognition process and redox characteristics of plastocyanin towards cytochrome f, as determined by point mutation experiments, has been attempted in this study. To achieve this goal we derived theoretical descriptors directly from the three-dimensional structure of the plastocyanin mutants, in the same manner as it is usually done for small drug-like molecules. The protein descriptors computed can be related to: (a) the electrostatic and dipole-dipole interactions, effective at long distance; (b) polar interactions whose features are encoded by charged partial surface area descriptors; (c) the propensity of the surface residues to form hydrogen bonding interactions; and (d) dispersion and repulsive interactions. Moreover, an estimation of mutation-dependent variation of redox potential observed has been obtained by electrostatic free energy calculations. The quantitative structure-activity relationship (QSAR) models offer structural interpretation of the point mutation experiment responses and can be of help in the design of new protein engineering experiments.

Introduction

In recent years, much research has been devoted to unravel the relationship between molecular structure and function. This has been especially so in the fields of pharmaceutical chemistry where quantitative structure-activity relationships (QSAR) have been applied to series of small organic molecules to develop medicinally active compounds [1]. The possibility to understand the mutual influences of structure and property, and to obtain meaningful quantitative rationalisations of the variation of the property of interest, depends primarily on the availability of descriptors able to decipher the encrypted code of the molecular structure. Currently, computational techniques can routinely be employed to generate sets of descriptors for appropriate representation of small molecules. Therefore the likelihood of pinpointing the subtle causes of the observed specific activity (binding affinity, selectivity, efficacy etc.) has been greatly enhanced [2]. Applications of QSAR techniques to peptide and protein systems are less well developed. With the overwhelming growth in information about protein function provided by structural genomics and proteomics, one can potentially change the QSAR perspective from ligands to proteins. Sound statistical models with good predictive ability have been reported in the literature. These make use of amino acid scores derived from the application of principal component analysis (PCA) to descriptor matrices of different types, such as empirical scales, 3D descriptors, interaction property descriptors, etc. [3] However, the descriptors employed in all these approaches are computed on isolated amino acids or on simplified

*Author for correspondence: Dipartimento. di Chimica, Universita' degli Studi di Modena e Reggio Emilia, Via Campi 183, 41100 Modena, Italy; e-mail: menziani@unimo.it

model systems and the physical interpretation of the correlations obtained is often unclear.

We have recently shown [4,5] that descriptors, originally derived to rationalise quantitatively the reactivity determinants of small molecules such as drugs, can be defined and computed on the 3D structures of series of protein mutants. These are successful in elucidating the causes that determine variations in molecular properties which are crucial for the protein function. In particular, the changes in the thermodynamics of spinach plastocyanin Cu(II) reduction, induced by point mutations of residues lying in close proximity to the protein copper centre were analysed by a QSAR approach [4]. The correlations found were informative as to how electrostatic and solvation effects control the $E^{\circ\prime}$ values in this species through the combined effects on the reduction enthalpy and entropy. Moreover, a comparative analysis of the computed molecular electrostatic potential of plastocyanin mutants and the wild type (wt) protein was found to interpret the variation of relative rate constants experimentally measured [5]. In the present work, we focus on the quantitative rationaliszation of the role of acidic residues on the eastern site of plastocyanin (pc) in the overall electron transfer from cytochrome f (cytf) [6,7].

Several site-directed mutagenesis experiments have been performed on this system. In particular, evidence for the participation of residues from both of the acidic patches at the 'eastern' site (Figure 1) of plastocyanin in the interaction with cytochrome f has been provided by perturbation of the putative association interface [8].

The performance of the descriptors computed on the 3D structure of the wt and mutant proteins has been evaluated with respect to the intrinsic nature of the functional data available. These refer to binding constants, redox potentials and overall second-order rate constants for reduction of plastocyanin variants by cytochrome f. Moreover, the problem of correctly reproducing the changes of the wt plastocyanin redox potential caused by mutations has been addressed using continuum electrostatics methods that allow the calculation of the pH-dependent electrostatic free energy differences between the two oxidation states of the protein [9].

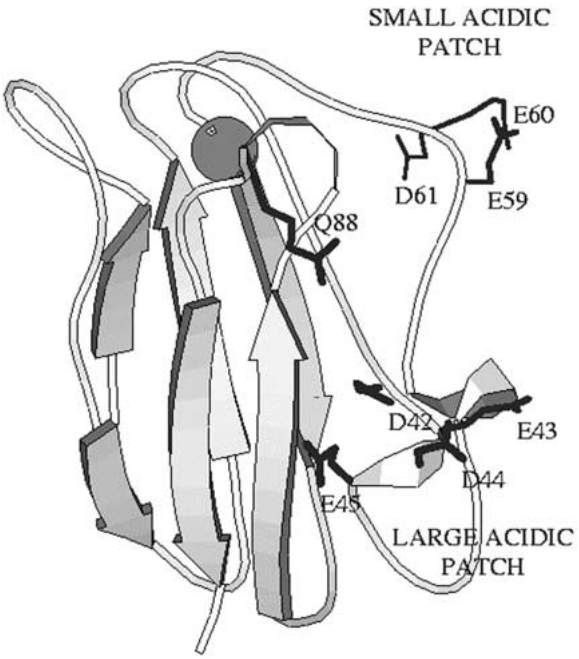

Figure 1. Mapping of the small and large acidic patch residues on the 3D structure of plastocyanin. The 'small acidic patch' consists of residues 59-61, located near the Cu-site, and the 'large acidic patch' of residues 42-45, located in the β5-β6 loop region. The picture was drawn using MOLSCRIPT [25].

Computational procedures

The crystallographic structure of spinach plastocyanin [10] was retrieved from the Brookhaven Protein Data Bank (PDB entry: 1ag6, oxidised state, pH 4.4). Mutants were obtained by substituting the appropriate residue in the wild type structure using the Biopolymer module of the InsightII software package (Accelrys, San Diego, CA., U.S.A.). Side chain spinning and Karplus rotamer library, available in the side chain modelling palette of the program Quanta (Accelrys, San Diego, CA., U.S.A.), have been used to ensure that the side chain conformation was energetically reasonable.

Molecular electrostatic potential and Similarity Indices (SIs) calculations

The University of Houston Brownian Dynamics (UHBD) program [11] was used to compute the electrostatic potentials of the proteins. The N- and C-terminal residues were treated as charged. The ionizable residues were considered in their usual protonation states at pH 7. The OPLS [12] non-bonded parameter sets for atomic charges and radii were as-

signed to the protein residues. The charge parameters were assigned to the copper site (Cu atom and its ligands) using a simple distribution algorithm [13]. A grid of dimensions 110 × 110 × 110 $Å^3$ was used, together with a 1.0 Å grid spacing, for the computation of the electrostatic potential via a finite difference solution of the linearised Poisson-Boltzmann equation. The grid was centred on the global centre of mass of the superimposed structures (superimposition was done using a standard least squares fit of Cα atoms). The dielectric constants of the solvent and the protein were set to 78 and 2, respectively. The dielectric boundary was determined from the van der Waals surface of the protein, and dielectric boundary smoothing [14] was implemented. The molecular electrostatic potential grids were computed at an ionic strength of 100 mM.

The protein interaction property similarity analysis (PIPSA) program was used for the similarity index calculations [15]. For each pair of proteins a and b, the molecular electrostatic potentials ϕ_a(i,j,k) and ϕ_b(i,j,k), computed on a three-dimensional grid, were compared by calculating the Hodgkin similarity index (SI) [16].

$$\mathrm{SI}(a, \mathrm{WT}) = 2(\mathbf{M}_{\boldsymbol{a}}, \mathbf{M}_{\mathrm{WT}})/(\mathbf{M}_{\boldsymbol{a}}{}^2 + \mathbf{M}^2_{\mathrm{WT}}) \quad (1)$$

where $(\mathbf{M}_{\boldsymbol{a}}, \mathbf{M}_{\mathrm{WT}}), \mathbf{M}_{\boldsymbol{a}}{}^2, \mathbf{M}^2_{\mathrm{WT}}$ are scalar products that represent the sum of products of values calculated for a given molecular electrostatic potential on grid points (i,j,k):

$$(\mathbf{M}_{\mathbf{a}}, \mathbf{M}_{\mathrm{WT}}) = \Sigma_{\mathrm{i,j,k}} \phi_a(\mathrm{i, j, k}) \phi_{\mathrm{WT}}(\mathrm{i, j, k}) \quad (2)$$

The SI values, which were computed as described in the paper by De Rienzo et al. [13] lie in the range −1 to 1, with values close to 1 implying that the proteins have highly similar potentials and values near to −1 implying that the potentials have inverted distributions. The details of the method have been described by Blomberg et al. [15].

Redox potential calculations

The procedure PKALK [9] was used to perform electrostatic free energy calculations of the wt plastocyanin and its mutants. A detailed description of the calculations is given by Scheafer et al. [9] Briefly, the program includes interfaces with the CHARMM program (version 22) [17] for coordinate generation and charge and radius assignment according to the all-hydrogen parameter set, the UHBD program [11] for the calculation of electrostatic free energies, and the Monte Carlo program, [18] for evaluating statistical ensemble averages of the titrating system. In the finite-difference Poisson-Boltzmann (FDPB) calculations the dielectric constants for the protein interior and the solvent were set to 20 [19] and 80, respectively. An ionic strength of 145 mM and a temperature of 293 K were used. A Stern (ion exclusion) layer of 2 Å was used.

The electrostatic free energy is calculated by integration of the titration curve. The program computes the pK_a values of each titratable group in the protein taking into account its environment. The PKALK procedure also allows the pK_as of titrating residues to be evaluated from the average protonation states of each site at a given pH; an average protonation of 0.5 is the equivalent of half-neutralisation. Since the free energy calculations are based on ensemble averages of protonation states at each pH, it is not necessary to define these explicitly. The calculation was performed for both the reduced and oxidised states of the protein.

Assuming that there are no relevant conformational changes in the protein structure within the range of pH considered for the titration curve, the difference in the electrostatic free energy of the two states, at a given pH, can be considered to be related to the redox potential by Nernst equation ($\Delta G = -nFE$). With this assumption, which is not true at pH<3.8, when the bonding His87 gets protonated and flips away from the copper, [20] the values of the redox potentials obtained at pH>3.8 (i.e. pH 6) can be directly compared to those from experiments carried out at the same pH.

The titratable sites are, besides the C and N terminus, Asp, Glu,, Arg, Lys, His and Tyr. However, for some residues the calculations do not correctly reproduce the experimental behaviour. In fact, very large changes to the standard pK_a value (about 10.3), which do not reproduce the real effects, are computed for the Tyr residues buried in the protein interior. When considering a static structure, deprotonation will be disfavoured if the surroundings cannot provide stabilisation of the charged form. This is likely to be the case for buried Tyr residues where the neutral, protonanted form will be involved in internal hydrogen bonding networks. Deprotonation produces a negative charge; experimentally this would most likely be accompanied by a change in the protein structure that would allow solvent access to the ion: in the calculation the position of the equilibrium would shift to allow the site to remain protonated. Since the standard pK_a of these groups is beyond the pH range over which experiments were carried out we assume that

they are in their protonated form throughout and thus exclude them from the titration calculation. Therefore, only Tyr83, which is completely exposed to the solvent, was treated as a titratable residue, while Tyr70 and Tyr80, which are only partially exposed and completely buried, respectively, were considered as non-titratable residues.

The treatment of the histidines is also atypical. In fact, the only two histidines (H37 and H87) in the pc structure are Cu-ligand residues: their Nδ atoms coordinate the metal ion. Therefore, both of them were considered as non titratable residues. This is of course an approximation required by the method. In fact, the reduced form of pc undergoes structural changes that involve the coordination of His87 to the copper. More precisely, at low pH values this histidine, which is solvent exposed, gets protonated and flips away from the Cu, loosing its coordination and inactivating the protein [20]. This important conformational change cannot be taken into account with the PKALK program, thought it influences the redox potential of plastocyanin. However, this will only affect energies calculated below or around the pK_as of the histidines; at pH = 6 it is reasonable to assume that they will be in their unprotonated form.

Definition of the descriptors used in the selected QSAR models

SI. Electrostatic potential similarity index computed on the plastocyanin and its mutants. The index sqrt(2-2SI) = $|\mathbf{M}_a - \mathbf{M}_{wt}|/sqrt[(\mathbf{M}_a^2 + \mathbf{M}_{wt}^2)/2]$, was used. Assuming that $\mathbf{M}_a^2 \approx \mathbf{M}_{wt}^2 \approx \mathbf{M}^2$, (which is reasonable, since only few mutations were introduced at the same time), then sqrt(2-2SI) = $|\mathbf{M}_a - \mathbf{M}_{wt}|/(\mathbf{M})$ and provides a measure of the similarity between the magnitudes and distributions of the molecular electrostatic potentials of the two proteins compared [13]. This index ranges from 0 to 2, with values near to 0 implying that the protein MEPs are highly similar and values near to 2 implying that MEPs are anticorrelated. When the index equals 1, the proteins have orthogonal MEPs.
D, and Dz. Dipole moments (D) of the proteins and their components along the z axis, (Dz) computed by means of the QUANTA program. The z axis coincides with the main axis of the elongated protein, with its origin at the geometric centre, pointing towards the copper site.
PNSA-1. Partial negative surface area [21] of the protein, computed on the 3D structures as PNSA-1= $\Sigma(-SA_i)$, where $(-SA_i)$ is the surface area contribution of the ith negative atom.
DPSA-1. Difference in the charged partial surface area [21] of the protein (DPSA-1 = PPSA-1 – PNSA-1). PPSA-1, the partial positive surface area of the proteins, is computed, in analogy with DPNSA-1, which is described above, as PPSA = $\Sigma(+SA_i)$, where $(+SA_i)$ is the surface contribution of the ith positive atom.
WHASA-3. Weighted hydrogen bond acceptor surface area [21] of protein, computed as [$\Sigma(HASA_i)$ Qi MMSA]/1000, where $HASA_i$ is the hydrogen acceptor surface area contribution of the ith atom, Q_i is its partial atomic charge, and MMSA is the total mutated molecular surface area.
IC. Information content defined on the basis of Shannon information theory [22]. A statistical treatment of chemical structures by means of information theory relies on the premise that each structure contains a limited set of elements (n) that can be decomposed into disjoint subsets n_i (i = 1, 2, ..., k) by means of selected equivalence relations defined on the set: IC = $-\Sigma_i p_i \log_2 p_i$ bit, where $p_i = n_i/n$ is the probability that a randomly selected element will be in the ith subset. The logarithm is taken at a basis 2 in order to measure the information in bits. IC is only evaluated for the mutated residue, therefore it is independent of the whole three-dimensional structure of the protein. It can easily be computed *a priori* from the secondary structure. According to its definition, this index encodes the branching ratio, unsaturation, and constitutional diversity of the amino acid side chains.
$\mathbf{V_{out}}$. Excess volume of the protein caused by mutation, computed by the Molecular Similarity facility of the Quanta program by superimposing each mutant on the wild type and comparing their total volume.
$(\mathbf{\Delta E_m})_c$. Redox potential differences are computed between the redox potential value calculated for each mutant and for the wild type. These values have been computed as differences between the electrostatic free energy of the reduced and oxidised forms [9]. Since we are interested in reproducing the experimental values of redox potentials which were measured at pH 6 [8], the descriptors were computed at this pH value.

Results and discussion

The key steps in the redox reaction between pc and cytf can be summarised as: (1) the formation of an initial complex of a purely electrostatic nature; (2) the

rearrangement necessary to attain a configuration of the complex optimal for electron transfer, and (3) the electron transfer itself.

The binding constants K_A [8] for the interaction of wt and mutant plastocyanins with cytochrome f, and the redox potentials (E_m), [8] listed in Table 1, describe respectively the first and the last process. The second order rate constant (k_2, Table 1) [8] does not provide direct information about the mechanism, but it is often used in comparative studies for estimating the effects of mutations in plastocyanin and/or cytochrome f on the overall electron transfer rate [23]. Therefore, we performed comparative analyses by computing $\ln(K_A/\ln K_{Awt})$, E_m-E_{mwt} and $\ln(k_2/k_{2wt})$, that is we considered the relative variation of the experimental parameters caused by mutations of the wt protein sequence. The significant intercorrelation observed (Table 2) between $\ln(k_2/k_{2wt})$ and $\ln(K_A/K_{Awt})$ ($r^2 = 0.97$) suggests that, for this series of protein mutants, the driving force of the whole process is the recognition step and that the rate of intracomplex electron transfer is little affected by mutations [8].

Spinach plastocyanin is strongly acidic, with negatively charged residues concentrated on the so-called 'eastern site' of the proteins, where Tyr83, one of the putative entry/exit points for electrons, is situated [24]. This acidic site consists of two separated clusters: the 'small acidic patch' of residues 59-61, located near the Cu-site, and the 'large acidic patch' of residues 42-45, located in the β5-β6 loop region (Figure 1). The electron-donor partner of pc, cytochrome f, has a ridge of positively charged residues located in the region where the heme is bound. Therefore the acidic area in plastocyanin is likely to be necessary for recognition of the basic patch in cytochrome f, as it has been shown by the NMR-based structures of the spinach plastocyanin/turnip cytochrome f complex obtained by Ubbink et al. [26,27]. However, rearrangement from the configuration of the initial electrostatic complex may be important for efficient electron transfer [28].

The values of the molecular descriptors involved in the correlations that better model the variations in the experimental data available for this set of plastocyanin mutants are listed in Table 1. All these descriptors are computed using the 3D structures of the wild-type protein and its mutants, as described in the Methods section. As can be noted from the correlation matrix reported in Table 2, in some cases significant correlations occur among the descriptors computed, therefore multiple correlation analysis can not be attempted.

Table 1. Experimental second-order rate constant at ph 6 ($k_{2\_6.0}$), binding constant (K_A) and redox potential (E_m values) [8] and theoretical molecular descriptors computed on the Plastocyanin and its mutants.

	Experimental data						Theoretical descriptors						
	$10^{-6}k_{2\_6.0}$ ($M^{-1}s^{-1}$)	K_A (M^{-1})	E_m (mV)	$\ln(k_2/k_{2wt})$[a]	$\ln(K_A/K_{Awt})$[b]	(ΔE_m)[c]	sqrt(2-2SI)[d]	D[d] (Debye)	Dz[d] Debye)	Vout[d] (A^3)	IC[d]	PNSA-1[d] (A^2)	DPSA-1[d] (A^2)	WHASA-3[d] (A^2)	(ΔE_m)[e]
1 **wt**	185.0	7100	381	0.000	0.000	0.00	0.000	318.10	198.00	0.00	90.295	1425.53	819.17	−3360.00	0.00
2 **q88e**	220.0	15040	375	0.173	0.751	−6.00	0.210	365.70	−249.40	4.13	87.921	1499.78	893.46	−3842.70	−11.59
3 **d42n**	76.50	2650	387	−0.883	−0.986	6.00	0.210	282.70	−173.70	3.00	102.363	1398.51	793.03	−3206.40	5.79
4 **e43n**	56.10	2860	384	−1.193	−0.909	3.00	0.224	259.90	−181.80	7.12	99.758	1287.22	680.86	−2744.80	1.62
5 **e43k**	29.30	660	387	−1.843	−2.376	6.00	0.480	203.00	−167.60	21.00	102.485	1285.47	530.83	−2677.90	3.20
6 **e43qd44n**	26.90	700	386	−1.928	−2.317	5.00	0.443	215.10	−152.90	13.00	112.798	1260.59	554.23	−2616.70	4.03
7 **e59ke60q**	13.80	312	406	−2.596	−3.125	25.00	0.645	164.00	52.99	25.88	119.012	1048.77	382.79	−2453.80	12.23
8 **e59ke60qe43n**	5.560		405	−3.505		24.00	0.828	103.90	−36.76	33.00	222.898	1010.46	244.48	−2054.40	12.87

[a] k_2 is the rate constant of each plastocyanin variant in the set; $k_2$wt is the rate constant of the wt plastocyanin.
[b] K_A is the association constant of each plastocyanin variant in the set; K_A is the association constant of the wt plastocyanin.
[c] Experimental differences between the redox potential of the wt and mutant plastocyanins.
[d] Theoretical descriptors whose meaning is explained in the details in the methods section
[e] Theoretical descriptor: differences between the computed redox potential values of the wt and mutant plastocyanins.

Table 2. Correlation matrix of the experimental data and molecular descriptors reported in Table 1.

	$\ln(k_2/k_{2wt})$	$\ln(K_A/K_{Awt})$	$(\Delta E_m)e$	D	Dz	sqrt(2-2SI)	Vout	IC	PNSA-1	DPSA-1	WHASA-3
$\ln(k_2/k_{2wt})$	1.000										
$\ln(K_A/K_{Awt})$	0.985	1.000									
$(\Delta E_m e$	−0.899	−0.839	1.000								
D	0.992	0.991	−0.902	1.000							
D_z	−0.930	−0.882	0.989	−0.934	1.000						
sqrt(2-2SI)	−0.950	−0.884	0.856	−0.921	0.871	1.000					
Vout	−0.941	−0.886	0.866	−0.934	0.876	0.978	1.000				
IC	−0.805	−0.937	0.741	−0.790	0.791	0.800	0.779	1.000			
PNSA-1	0.960	0.906	−0.953	0.959	−0.973	−0.905	−0.919	−0.772	1.000		
DPSA-1	0.985	0.968	−0.905	0.988	−0.937	−0.948	−0.966	−0.799	0.974	1.000	
WHASA-3	−0.962	−0.933	0.837	−0.974	0.884	0.838	0.855	0.730	−0.936	−0.950	1.000

Table 3. Regression models obtained for the plastocyanin variants by using the molecular descriptors listed in Table 1. The values in parentheses represents the errors on the regression coeficients, *n* is the number of proteins, *R* is the correlation coefficient, *s* is the standard deviation, and *F* is the value of the Fisher ratio.

1) $\ln(k_2/k_{2wt}) = 0.005(\pm0.0004)$DPSA-1 $-4.824(\pm0.257)$
n=8, $R^2 = 0.969$, F=189.58, $s^2 = 0.056$

2) $\ln(k_2/k_{2wt}) = -4.392(\pm0.591)$sqrt(2-2SI) $+0.202(\pm0.270)$
n=8, $R^2 = 0.902$, F=55.17, $s^2 = 0.18$

3) $\ln(k_2/k_{2wt}) = -0.0021(\pm0.0002)$WHASA-3 $-7.577(\pm0.724)$
n=8, $R^2 = 0.925$, F=73.60, $s^2 = 0.139$

4) $\ln(k_2/k_{2wt}) = -0.098(\pm0.014)$Vout $-0.155(\pm0.253)$
n=8, $R^2 = 0.885$, F=46.02, $s^2 = 0.212$

5) $\ln(K_A/K_{Awt}) = -0.019(\pm0.001)$D $-6.371(\pm0.311)$
n=7, $R^2 = 0.983$, F=283.18, $s^2 = 0.040$

6) $\ln(K_A/K_{Awt}) = -0.117(\pm0.019)$IC $-10.673(\pm2.009)$
n=7, $R^2 = 0.877$, F=35.74, $s^2 = 0.286$

7) $\Delta Em = 0.151(\pm0.009)$Dz $+30.807(\pm1.497)$
n=8, $R^2 = 0.979$, F=280.62, $s^2 = 2.955$

8) $\Delta Em = -0.061(\pm0.008)$PNSA-1 $+85.517(\pm10.08)$
n=8, $R^2 = 0.909$, F=59.72, $s^2 = 12.887$

The linear regression models reported in Table 3 have been chosen on the basis of the best statistical parameters exhibited in the correlations obtained. They involve molecular descriptors that can be related to: (a) the electrostatic (eq. 2, Table 3) and dipole-dipole (eqs 5, and. 7, Table 3) interactions, effective at long distance; (b) polar interactions whose features are encoded by charged partial surface area (CPSA) descriptors [21] (eqs 1, and 8, Table 3); (c) the propensity of the surface residue to form hydrogen bonding interactions (eqs 3, Table 3); (d) dispersion (eq. 6, Table 3) and repulsive interactions (eq. 4, Table 3).

Mutation of the wt protein to negatively charged or neutral residues induces modifications of the electrostatic potential distribution and of the dipolar characteristics of the protein. As previously pointed out, [5] the correlation obtained with the SI descriptor (eq. 2 Table 3) shows that the overall rate constants for the electron transfer reactions between wt or mutant pcs and cytf are affected mainly by electrostatic interactions, at least when mutations are introduced in the eastern site region. In fact, mutation of one or more negatively charged residue(s) into a positively charged or polar residue in the plastocyanin eastern site produces changes in the electrostatic properties of the protein, reducing its ability to attract the cytochrome f basic patch and consequently reducing the overall reaction rate. This index is very useful to synthesise numerically the great amount of information obtained by the visual inspection and comparison of molecular properties, such as electrostatic potentials and shapes, but suffers from the problem of the choice of the reference protein. We addressed this problem in details in a recently published paper [5]. Briefly, the obvious choice is to compare the mutants with respect to the wild type protein so that the effect of point mutations on the protein surface can be readily evaluated. However, in the set of mutants used, the Q88E mutant has a reaction rate constant (k_2Q88E) that is, with respect to the experimental errors, comparable to (at pH 7.5) or slightly higher than (at pH 6) that of the wt protein. k_2Q88E is not correctly predicted by any of the regression models: it is always underestimated. By choosing the Q88E mutant as the

reference protein, in analogy with the small molecule approach, where the molecule with the highest activity is chosen as the reference, a significant improvement in the correlation statistics is indeed observed ($\ln (k_2/k_{2Q88E}) = -4.257(\pm 0.336)\mathrm{sqrt}(2\text{-}2SI_{Q88E}) + 0.548(\pm 0.184)$; n=8, $R^2 = 0.964$, F=160.41, $s^2 = 0.06$). [5] The direct consequence of this choice is that the model should be rebuilt whenever further experiments reveal new mutants with improved activity.

Confirmation of the importance of mutation-induced charge distribution modification to the association process and the redox potential is provided by eq. 5 and eq. 7 (Table 3), which involve the total dipole moment descriptor (D) and its component along the z axis (Dz), respectively. In both cases 98% of the variation in the experimental data is explained. The interpretation of eq. 5 is straightforward, i.e. the association of plastocyanin and cytochrome f is favoured by mutational enhancement of the protein polarity. In contrast, the effect on the redox potential might be expected to be more complex. In fact, small but significant experimental differences were observed in the redox potential of the small acidic patch mutants (E59K/E60Q and E59K/E60Q/E43N) and the Q88E mutant while the E_m values of the large acidic patch mutants (D42N, E43N, E43K, and E43Q/D44N) were not significantly different to that of the wt protein [8].

The positive end of the dipole moment vector in the wild type protein makes an angle of about -120 degrees with the z axis, which extends from the protein centre of mass towards the copper site. The regression analysis shows that small redox potential values are associated with more negative values of the z dipole moment component resulting from a mutational shift of the centre of gravity of the positive charge of the protein moving away from the copper site towards the opposite end of the barrel-like protein. Therefore, the introduction of a negative charge (Q88E) produces stabilisation of the oxidised state, while mutations that render the protein globally more neutral stabilise the reduced state to an extent that depends on the position of the mutated residues with respect to the copper site.

In addition to the electrostatic effects underlined by the Dz index, the extent of the partial negative surface area (PNSA-1) of the protein variants (eq. 8) seems to play an important role on the relative stability of the oxidised and reduced forms of the protein mutants, probably by influencing the orientation of the water molecules in the network which is established in proximity of the copper site.

Specific and non-specific polar interactions are represented by the hydrogen bonding accessible surface and charged contact surface area descriptors, which are computed for the proteins. In particular, the difference in partial positive and negative surface area of the proteins (DPSA-1), a descriptor that models polar intermolecular interactions, yields a very good linear relationship with k_2 (eq. 1, Table 3), while the contribution of the hydrogen-bonding potential of the proteins for the interaction of the plastocyanin variants under study with the cytochrome f is taken into account in eq. 3, Table 3.

Finally, detrimental aspects for the interaction of pc mutants with their redox partner are: a) the increased capability of these mutants to establish dispersion interactions with the rest of the protein, and, therefore, to delocalisze the positive or polar charge introduced, as shown by the negative slope of the correlations between $\ln(k_A/k_{Awt})$ and the information content (IC) descriptor (eq. 6, Table 3), and b) the repulsive interactions between the protein partners as codified by V_{out}, (eq. 4, Table 3), the mutation induced excess volume of the proteins with respect to the volume of the wt plastocyanin.

The results obtained by a different approach based on computation of the relative electrostatic free energy differences between the redox states of the wild type plastocyanin and its mutants are reported in Figure 2. A satisfactory prediction is obtained by considering that the midpoint redox potential experimental values were accurate to ± 3mV [8].

In general, it can be observed that the variation of the redox potential of the wt pc is underestimated when mutations introduce one or more positive charges (E43K, E59K/E60Q, E59K/E60Q/E43N), while the opposite might be envisaged when mutations introduce additional negative charges in the eastern site (Q88E). Given the flexible nature of such groups it might be expected that a more accurate description of these residues that would incorporate more conformational flexibility could be advantageous. The largest error (about 50%) in the estimation of the relative variation of redox potentials due to mutations was observed for the small acidic patch mutants (E59K/E60Q, E59K/E60Q/E43N). This discrepancy might be explained by considering that the residues E59 and E60 lie in close proximity to the Cu-site and can, therefore, be important for the stability of the redox site. Mutations of these residues might, for instance, induce modifications of the active site conformation or increase its flexibility. These effects,

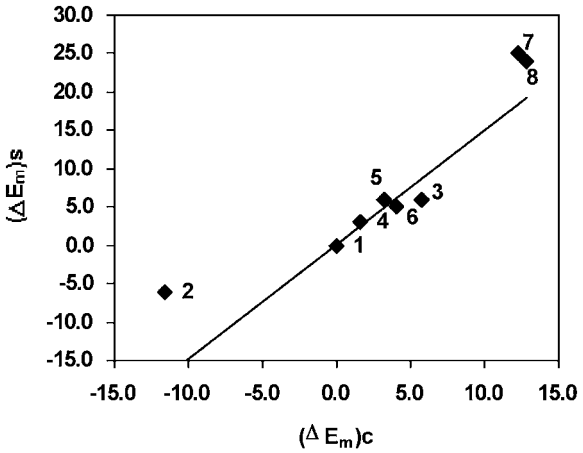

Figure 2. Experimental versus computed midpoint redox potential differences for plastocyanin and its mutants. The correlation obtained by considering all the proteins in Table 1 is: $(\Delta E_m)e = 1.311(\pm 0.235)\ (\Delta E_m)c + 3.263(\pm 1.883)$, n=8, $r^2 = 0.838$, F = 31.02, $s^2 = 22.875$. The correlation obtained by omitting the Q88E mutant is: $(\Delta E_m)e = 1.983(\pm 0.193)\ (\Delta E_m)c - 1.400(\pm 1.417)$, n=7, $r^2 = 0.955$, F = 105.74, $s^2 = 5.661$.

which might affect the entropy of the system and therefore influence the experimental redox properties of the site, are not taken into account while computing the $(\Delta E_m)c$ descriptor. A second area concerns, as hinted above, the behaviour of the Cu-bonding histidines. Their explicit incorporation in the calculation, again through the consideration of alternative conformations, should be explored.

Conclusions

Key features for plastocyanin interaction with cytochrome f are depicted by the molecular descriptors involved in the quantitative structure-affinity/function relationship analysis carried out in this work.

They provide quantitative models of the effects of mutations on electrostatically-influenced protein-protein reaction rate, binding constant and redox potential, confirming the hypothesis that the association (molecular recognition) process is a primary determinant of the overall electron-transfer process between cytf and pc, at least *in vitro*.

Other factors that intervene in the interaction process and affect the redox properties of the partners, such as short-range attractive and repulsive forces and hydrogen-bonding potentials, are codified by CPSA and volume descriptors computed on the 3D structure of the proteins or by topological descriptors computed on the protein secondary structure.

In addition, we have shown, for the first time, that electrostatic free energy calculations can successfully reproduce the mutation-dependent variation of redox potentials observed for blue-copper proteins.

The efficacy of the method presented here needs to be verified for other systems, however, the models obtained can be used to guide interesting point mutations of residues governing the interaction between the two redox partners. These studies would further validate the present models and would reveal whether they are accurate enough to be used for protein engineering.

Acknowledgments

We thank Michael Schaefer (Universite Louis Pasteur, Strasbourg) for the PKALK suite of programs.

References

1. Livingstone, D.J. J. Chem. Inf. Comput. Sci., 40 (2000) 195–209.
2. Karelson, M., Lobanov, V.S., Katritzky, A.R. Chem. Rev., 96 (1996) 1027–1043.
3. Sandberg, M., Eriksson, L., Jonsson, J., Sjostrom, M., Wold, S. J. Med. Chem., 41 (1998) 2481–2491.
4. Battistuzzi, G., Borsari, M., Loschi, L., Menziani, M.C., De Rienzo, F., Sola, M. Biochemistry, 40 (2001) 6422–6430.
5. De Rienzo, F., Gabdoulline, R.R., Menziani, M.C., De Benedetti P.G. and Wade, R.C. Biophys. J., 81, (2001) 3090–3104.
6. Adman, E.T. Adv. Prot. Chem., 42 (1992) 145–197, and referencs therein.
7. Sykes, A.G. Adv. Inorg. Chem., 36 (1994) 377–408.
8. Kannt, A., Young, S., Bendall, D.S. Biochim. Biophys. Acta., 1277 (1996) 115–126.
9. Schaefer, M., Sommer, M., Karplus, M. J. Phys. Chem. B, 101, (1997) 1663–1683.
10. Xue, Y., Okvist, M., Young, S. Prot. Sci, 7 (1998) 2099–2105.
11. Madura, J.D., Briggs, J.M., Wade, R.C., Davis, M., Luty, B.A., Ilin, A., Antonsiewicz, J., Bagheri, M.K., Scott, B., McCammon, J.A. Comp. Phys. Comm., 91 (1995) 57–95.
12. Jorgensen, W.L., Tirado-Rives, J. J. Am. Chem. Soc., 110 (1988) 1657–1666.
13. De Rienzo, F., Gabdoulline, R.R., Menziani, M.C. Wade, R.C. Prot. Sci., 9 (2000) 1439–1454.
14. Davis, M.E., McCammon, J.A. J. Comp. Chem., 12 (1991) 909–912.
15. Blomberg, N., Gabdoulline, R.R., Nilges, M., Wade, R.C. Proteins, 37 (1999) 379–387.
16. Hodgkin, E.E., Richards, W.G. Int. J. Quant. Chem. Quant. Biol. Symp., 14 (1987) 105–110.
17. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan,S., Karplus, M. J. Comput. Chem., 4 (1983) 187–217.

18. Beroza, P., Fredkin, D.R., Okamura, M.Y., Feher, G. Proc. Natl. Acad. Sci. USA, 88 (1991) 5804–5808.
19. Antosiewicz, J., McCammon, J.A., Gilson, M.K. J. Mol. Biol., 238 (1994) 415–436.
20. Redinbo, M.R., Yeates, T.O., Merchant, S. J. Bioenerg. Biomemb., 26 (1994) 49–66.
21. Stanton, D.T., Jurs, P.C. Anal. Chem. 62 (1990) 2323–2329.
22. Kier, L.B. J. Pharm. Sci., 69 (1980) 807–811.
23. Hope, A.B. Biochim. Biophys. Acta., 1456 (2000) 5–26.
24. Chen, L., Durley, R.C.E., Mathews, F.S., Davidson, V.L. Science 264 (1994) 86–90.
25. Kraulis, P.J. J. Appl. Crystallogr., 24 (1991) 946–950.
26. Ubbink, M., Ejdeback, M., Karlsson, B.G., Bendall, D.S. Structure, 6 (1998) 323–335.
27. Ejdeback, M., Bergkvist, A., Karlsson, B.G., Ubbink, M. Biochemistry, 39 (2000) 5022–5027.
28. Bendall, D.S., Wagner, M.G., Schlarb, B.G., Soellick, T.R., Ubbink, M., Howe. C.J. *In* Peschek et al. (eds), The Phototrophic prokaryotes. Kluwer Academic/Plenum Publishers, New York. 1999, pp. 315–328.