Biochimica et Biophysica Acta 1864 (2016) 211–218

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

A computational analysis of *S*-(2-succino)cysteine sites in proteins

Gianluca Miglio [a,*], Alessandro Damiano Sabatino [b], Eleonora Veglia [a], Maria Teresa Giraudo [c], Marco Beccuti [d], Francesca Cordero [d]

[a] *Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy*
[b] *Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK*
[c] *Dipartimento di Matematica "Giuseppe Peano", Università degli Studi di Torino, Turin, Italy*
[d] *Dipartimento di Informatica, Università degli Studi di Torino, Turin, Italy*

ARTICLE INFO

Article history:
Received 8 August 2015
Received in revised form 19 October 2015
Accepted 13 November 2015
Available online 14 November 2015

Keywords:
Protein succination
S-(2-succino)cysteine sites
Cysteine reactivity
Fumaric acid

ABSTRACT

The adduction of fumaric acid to the sulfhydryl group of certain cysteine (Cys) residues in proteins via a Michael-like reaction leads to the formation of *S*-(2-succino)cysteine (2SC) sites. Although its role remains to be fully understood, this post-translational Cys modification (protein succination) has been implicated in the pathogenesis of diabetes/obesity and fumarate hydratase-related diseases. In this study, theoretical approaches to address sequence- and 3D-structure-based features possibly underlying the specificity of protein succination have been applied to perform the first analysis of the available data on the *succinate proteome*. A total of 182 succinated proteins, 205 modifiable, and 1750 non-modifiable sites have been examined. The rate of 2SC sites *per* protein ranged from 1 to 3, and the overall relative abundance of modifiable sites was 10.8%. Modifiable and non-modifiable sites were not distinguishable when the hydrophobicity of the Cys-flaking peptides, the acid dissociation constant value of the sulfhydryl groups, and the secondary structure of the Cys-containing segments were compared. By contrast, significant differences were determined when the accessibility of the sulphur atoms and the amino acid composition of the Cys-flaking peptides were analysed. Based on these findings, a sequence-based score function has been evaluated as a descriptor for Cys residues. In conclusion, our results indicate that modifiable and non-modifiable sites form heterogeneous subsets when features often discussed to describe Cys reactivity are examined. However, they also suggest that some differences exist, which may constitute the baseline for further investigations aimed at the development of predictive methods for 2SC sites in proteins.

1. Introduction

The post-translational modification (PTM) of proteins is a major mechanism for eukaryotic cells to regulate their complex physiological processes [1,2]. Moreover, the adduction of reactive compounds to proteins is a pivotal triggering event in chemical toxicity [3,4]. So far, more than 400 PTM types have been described, and a significant fraction of proteins is thought to be affected by PTMs [5,6]. However, the role of some types of modification remains to be fully understood.

The sulfhydryl group in the side chain of cysteine (Cys) residues in proteins is subject to several PTM types, including sulphenation, sulphination, nitrosylation, sulfhydration, glutathionylation, and adduction to endogenous and exogenous electrophiles [7–9]. The adduction of fumaric acid (FA) to the free sulfhydryl group of certain Cys residues in proteins via a Michael-like reaction leads to the formation of *S*-(2-succino)cysteine (2SC) sites. Originally detected in plasma proteins [10], including albumin, this post-translational Cys modification type (referred as protein succination) has been described in models of diabetes/obesity, fumarate hydratase (FH)-related diseases, and Leigh syndrome. Increased levels of succinated proteins have been found in murine 3T3-L1 adipocytes cultured in high glucose medium (30 mM, compared with a physiological level of 5 mM), as well as in tissues from streptozotocin-treated rats, *db*/*db* (leptin receptor deficient), *ob*/*ob* (leptin deficient), and diet-induced obese mice [11–14]. It has been proposed that nutrient excess results in elevated ATP:ADP, NADH:NAD^+, and mitochondrial membrane potential, and that the increased NADH:NAD^+ inhibits oxidative phosphorylation, leading to a persistent accumulation of mitochondrial intermediates including FA, which, in turn, causes protein succination [15]. A pronounced accumulation of succinated proteins even results from the loss of FH activity [16–20]. FH catalyses the reversible conversion of FA to malic acid in the tricarboxylic acid cycle. Importantly, germline loss-of-function mutations in *FH* are known to predispose affected individuals to

Abbreviations: Cys, cysteine; DAVID, Database for Annotation, Visualization and Integrated Discovery; FA, fumaric acid; FH, fumarate hydratase; GO, gene ontology; HLRCC, Hereditary Leiomyomatosis and Renal Cell Cancer; MC, modifiable cysteine; NMC, non-modifiable cysteine; PFM, positional frequency matrix; PTM, post-translational modification; ROC, receiver operator characteristics; 2SC, *S*-(2-succino)cysteine; SD, succination degree; TM, transition matrix.

* Corresponding author at: Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, 10125 Turin, Italy.
E-mail address: gianluca.miglio@unito.it (G. Miglio).

http://dx.doi.org/10.1016/j.bbapap.2015.11.003
1570-9639/

multiple cutaneous and uterine leiomyomas, as well as to Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC) [21]. Protein succination has been implicated in the FH-related carcinogenesis [22–24], albeit its exact role remains to be fully elucidated [25–27]. Finally, increased protein succination has also been described in the brainstem of Ndufs4 knockout mice (a model of Leigh syndrome) [28], thus indicating a potential role for this PTM type in the pathogenesis of this mitochondrial disorder. Collectively, these findings indicate the potential relevance of protein succination and suggest that a better understanding of its role could be achieved by a more comprehensive investigation of the *succinate proteome*.

Many 2SC sites have been identified in previous studies [29–31]. However, given the abundance of Cys residues in proteomes [32], it is reasonable that a large proportion of succinated proteins and 2SC sites still remains to be identified. Experimental identification of modified sites is typically a labour-intensive and expensive approach. Computational prediction of these sites could be a convenient and high-throughput strategy to generate helpful information for subsequent experimental verification. Methods for computationally predicting redox-active Cys residues with specific modifications have already been proposed [8,33–37]. However, to date no method has been developed to predict 2SC sites in proteins, and the *succinate proteome* has never been studied in order to evaluate the specificity of the protein succination.

Therefore, toward the long-term goal of developing computational methods to predict 2SC sites in proteins, in this study several theoretical approaches to address sequence- and 3D-structure-based features that may lie behind the specificity of protein succination have been applied to analyse the available data (published or extracted from other open sources) on the *succinate proteome*.

2. Material and methods

2.1. Protein data collection

Data sets of proteins containing 2SC sites at established locations were built by collecting data from published reports. The suite tool Database for Annotation, Visualization and Integrated Discovery (DAVID; version 6.7, http://david.abcc.ncifcrf.gov/) was used to analyse different annotations (e.g., for cellular component, biological process, molecular function). The sequence of the succinated proteins were retrieved from the UniProt resource (http://www.uniprot.org/). Available 3D-protein structures were retrieved from the Protein Data Bank repository (http://www.rcsb.org/pdb/home/home.do), and visualized in VegaZZ (http://nova.disfarm.unimi.it/cms/index.php?Software_projects:VEGA_ZZ; version 3.0.3).

2.2. Analysis of 2SC sites

Data sets of succinated proteins were tested for parameters often discussed in regard to post-translational Cys modifications. From the sets of protein sequences, all fragments having a Cys residue in the central position were extracted. Each sequence fragment represents a peptide of x amino acids upstream and downstream of the Cys residue and length $= 1 + 2x$ amino acids, as described in previous studies on other PTM types [35–40]. Each fragment was assigned to a set according to the type of Cys residue: modifiable Cys (MC) or non-modifiable Cys (NMC).

The hydrophobicity of the Cys-flanking peptides was analysed using an in-house Python (https://www.python.org/, version 3.4) programme, implementing the standard Kyte–Doolittle scale [41] and adopting $x = 6$ amino acids, as previously described [38,39].

The acid dissociation constant (pK_a) of the free sulfhydryl group in the side chain of Cys residues were calculated as described by Marino and Gladyshev [39] with the PropKa implementation in VegaZZ.

The solvent accessibility area of the sulphur atom in the side chain of Cys residues was determined with the web tool DSSP (http://swift.cmbi.ru.nl/gv/dssp/), by default settings.

2.3. Positional frequency matrices and transitional matrices

The amino acid composition of the Cys-flanking fragments was determined and examined. In particular, Cys-centred peptides were staked and aligned (Cys in the central position was the unique constraint). Then, MC and NMC sets were pre-processed to exclude redundant fragments (fragments with identical amino acid sequence appear once). Finally, Positional Frequency Matrices (PFM_{MC} and PFM_{NMC}) were generated (see supplemental material Table S1, and Table S2). PFM_{MC} and PFM_{NMC} are $22 \times (1 + 2x)$ matrices, where 22 denotes the 20 natural amino acids in proteins, plus X (any unknown amino acid), and O (empty positions due to the proximity of the central Cys residue to either the *N*- or *C*-terminal), while $1 + 2x$ amino acids is the window size. The matrix entries, PFM_{MCij} and PFM_{NMCij}, report the relative frequency of the amino acid i in the position j, computed considering all segments in the MC or NMC subset. MatCompare [42] was used to quantify the similarity between PFM_{MC} and PFM_{NMC}. The amino acid composition of the Cys-flanking peptides were also used to generate two transition matrices, TM_{MC} and TM_{NMC}, that model the probabilities of finding a given amino acid at a certain position depending on the nearest previous amino acid in the sequence according to a simple first-order stationary Markov chain model. TM_{MC} and TM_{NMC} are three-dimensional arrays $22 \times 22 \times 2x$, where 22 is the number of residues (as previously defined), and $2x$ is the number of positions. Moreover, the first position of the protein segments was associated with a frequency vector. To verify the similarity between the TMs a specific statistical test was used to find out possible significant differences between couples of corresponding estimated transition probabilities (see below).

2.4. A sequence-based computational descriptor for cysteine residues

The PFM matrices were used to scan fragment sequences using the sliding-window approach. An absolute score associated with each sequence was computed, by adding the corresponding frequency values in the PFM considered. In particular, for each sequence two scores were computed:

$$\text{Score\_MC} = \sum_{j=-x}^{-1} \text{PFM}_{\text{MC}_{ij}} + \sum_{j=1}^{x} \text{PFM}_{\text{MC}_{ij}}, \quad (1)$$

$$\text{Score\_NMC} = \sum_{j=-x}^{-1} \text{PFM}_{\text{NMC}_{ij}} + \sum_{j=1}^{x} \text{PFM}_{\text{NMC}_{ij}}, \quad (2)$$

where the index i corresponds to the specific amino acid found in the j-th position of the sequence. Then, a 'Succination Degree' (SD) was defined as:

$$\text{SD} = \frac{\text{Score\_MC}}{\text{Score\_NMC}}. \quad (3)$$

The TM matrices, with their corresponding frequency vectors, were also used to compute score values using Eqs. (1) and (2) instantiated on TM_{MC} and TM_{NMC} matrices. SD values were then computed using Eq. (3), but substituting Score_MC and Score_NMC with such new scores.

To evaluate the ability of the PFM-based and TM-based SD to describe modifiable and non-modifiable sites, first the two sets of fragments were randomly split (SD values were normally distributed; data not shown) into training and test subsets, consisting in 20% and 80% of the total number of fragments, respectively. Fragments belonging to the test subset were then examined by drawing Receiver Operator

Characteristics curves using GraphPad Prism 5 (http://www.graphpad.com/).

3. Results

3.1. Overview of the reference data sets of proteins containing 2SC sites

Sets of proteins containing 2SC sites at established locations have been described in three reports: (*i*) 94 proteins in embryonic fibroblasts and renal cysts from FH-deficient mice have been reported by Ternette et al. [29], called set *A*; (*ii*) 28 proteins in 30 mM glucose-treated murine adipocytes have been described by Merkley et al. [30], namely set *B*; and (*iii*) 60 proteins in FH-mutant tumour and two HLRCC-derived cancer cell lines (UOK262 and NCC-FH-1) have been reported by Yang et al. [31], called set *C*.

To determine the biological meaning behind these sets of proteins, the annotations for each modified protein were analysed, after a pre-processing step to exclude redundancy due to the presence of orthologues. A wide distribution was determined when Gene Ontology (GO) annotation terms for cellular components were examined (Table S3). Mitochondria-associated terms (i.e., GO:0005739 ~ *mitochondrion*, GO:0044429 ~ *mitochondrial part*, GO:0031966 ~ *mitochondrial membrane*) were commonly detected in the three data sets, indicating that reported succinated proteins are predominately mitochondrial. However, proteins with exclusively cytosol- or nucleus-related annotations were also represented (Table S3).

Protein succination is not limited to proteins of a specific biological process or molecular function (Table S4 and Table S5). Indeed, for example, enzymes involved in carbohydrate, amino acid, nucleobase/nucleoside/nucleotide and fatty acid metabolism, RNA processing, proteins synthesis, iron, zinc, and copper ion homeostasis, cytoskeleton organization were also represented in the three sets of proteins.

Protein domains often dictate cellular localization and function. When the InterPro annotation terms for protein domains were examined, most proteins were not captured. However, some annotation terms were overrepresented, and two (IPR000504: *RNA recognition motif*, *RNP-1*, and IPR012677: Nucleotide-binding, α-β plait) were significantly enriched (Table S6). Collectively, these data indicate that protein succination is not limited to proteins of a specific biological function or structure.

The mean frequency of Cys residues in proteins of the set *A*, *B*, and *C* were 2.1 ± 0.2%, 4.7 ± 1.6%, and 2.0 ± 0.1%, respectively (Fig. 1A). Due to their high Cys content, cornifin-A, metallothionein-1, and -2 were classified as outliers ($P < 0.01$; Grubb's test). The rate of 2SC sites *per* protein ranged from 1 to 3, with more than 80% of proteins having a single 2SC site (Fig. 1B). The rate of modified Cys residues was 10.9%, 3.7%, and 11.2%, respectively. However, having a single 2SC site and 523 Cys residues, zonadhesin markedly contributes to the low abundance of 2SC sites in the set *B*; excluding this protein from the analysis, the rate of 2SC sites in the set *B* and the overall relative abundance of modifiable sites rise to 9.6% and 10.8%, respectively.

Collectively, these findings indicate that, despite the heterogeneity in almost all annotations examined, which is consistent with the idea that succination is a non-specific PTM type, only certain Cys residues can react with FA to form 2SC sites, thus suggesting the existence of features of Cys residues that govern the specificity of this PTM type.

3.2. Analysis of the 2SC site features

A new protein data set consisting of 178 proteins defined as the union among the three published sets of proteins was studied (Fig. 2). The original MC and NMC sets included 205 and 1750 fragments, respectively. After exclusion of redundant fragments, they include 190 and 1524 fragments, respectively.

3.2.1. Hydrophobicity of the Cys-flaking peptides

Peptides flanking Cys residues were first studied for the hydrophobicity, as defined by the Kyte–Doolittle scale, where positive values correspond to a hydrophobic and negative to a hydrophilic environment [41]. As shown in Fig. 3, computed values were normally distributed and, compared to the NMC residues, the average hydrophobicity of the MC residues was slightly, albeit significantly, lower (0.058 ± 0.015 and −0.165 ± 0.041, respectively; $P < 0.001$, *Z*-test). The large overlap between the two sets of sites indicates that hydrophobicity computed by the Kyte–Doolittle scale could not be a reliable parameter to distinguish MC from NMC residues.

3.2.2. Amino acid composition of the Cys-flanking peptides

From a descriptive analysis, several differences were highlighted when the composition of the Cys-flanking peptides were compared (Fig. 4). For example, residues with bulky side chains (e.g., valine, phenylalanine, and leucine) were underrepresented, while residues with polar chains (e.g., serine, and threonine) were overrepresented. MatCompare was used to compute the Kullback–Leibler divergence between the PFM_{MC} and PFM_{NMC}. No statistically significant divergence

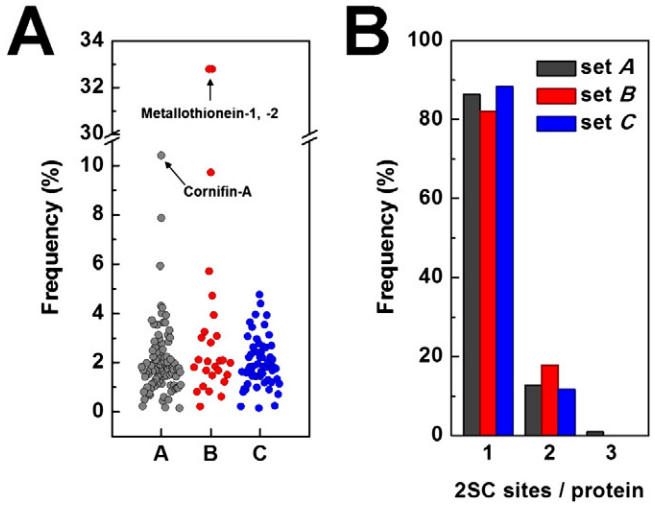

Fig. 1. Overview of protein datasets. Occurrence frequency of Cys residues (A), and of 2SC sites *per* protein (B) was determined for each protein. In addition, outliners ($P < 0.01$, Grubb's test for outliers) are highlighted.

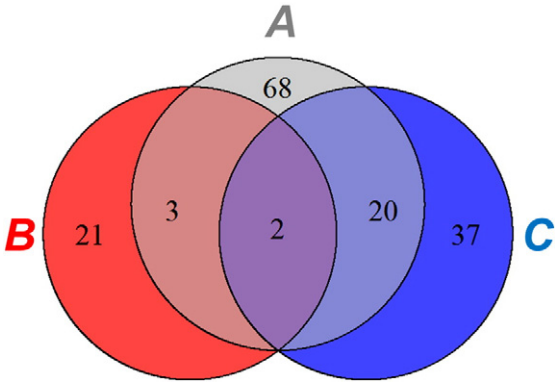

Fig. 2. The succinate proteome. After a preprocessing task to exclude proteins with either ambiguous location or unusual abundance of the 2SC sites (albumin from set *A*, triosephosphate isomerase and zonadhesin from set *B*, cytochrome *c*-type heme lyase from set *C*), sets *A* (93 proteins), *B* (26 proteins), and *C* (59 proteins) were compared and merged to generate set *D* (178 proteins). The Venn diagram displays the composition of set *D*. Intersections indicate the presence of proteins shared (identical or orthologues) by the three original sets ($A \cap B = 5$ proteins; $A \cap C = 22$ proteins; $A \cap B \cap C = 2$ proteins).

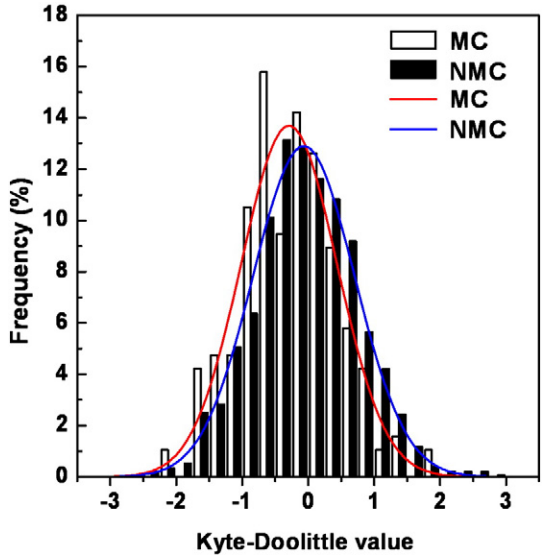

Fig. 3. Hydrophobicity of the Cys-flanking peptides. After a preprocessing task to exclude redundancy, Kyte–Doolittle values ($x = 6$ positions) were computed for modifiable ($n = 190$) and non-modifiable ($n = 1524$) Cys residues in data set *D*. Data were normalized and analysed by a Gaussian fitting. Their distributions are shown both as row data (bars) and after fitting (lines).

Table 1
P-values computed using the Kullback–Leibler divergence between the PFM_{MC} and PFM_{NMC}.

Position	*P*-value
−12	0.932
−11	0.653
−10	0.250
−9	0.173
−8	0.733
−7	0.303
−6	0.190
−5	0.083
−4	**0.013**
−3	**0.019**
−2	0.096
−1	NA
0	NA
1	**0.021**
2	0.550
3	**0.037**
4	0.385
5	**0.049**
6	0.665
7	0.469
8	0.807
9	0.264
10	**0.035**
11	0.407
12	0.600

NA = not applicable

was determined when the whole matrices were compared. However, significant divergences were determined when distinct *P*-values were computed for each single position (Table 1).

Furthermore, to verify the similarity among the transition matrices TM_{MC} and TM_{NMC}, binomial tests were performed to find out possible differences between any two corresponding estimated transition probabilities. The positions where differences have been detected at a significance level less than 0.1 together with the two amino acids involved in the transition from the given site to the next are shown in Table 2.

Together these findings indicate that the amino acid composition of the Cys-flanking peptides could be a reliable parameter to distinguish MC from NMC residues.

3.2.3. Secondary structures

The environment surrounding Cys residues in proteins, either with established crystal or NMR structures, or that could be modelled by standard homology modelling approaches using Swiss Model server, was examined. Given the limited number of available 3D-structures, in order to maximize the chance to discover any significant differences, the analysis was focused on subsets of hydrophilic and hydrophobic Cys residues, and representative results have been shown in Table 3. Among the 20 Cys residues described, 12 (five MC and seven NMC) were in a α-helices, 1 in a β-strand (one NMC), and 7 (five MC and two NMC) in a segment with no secondary structure assigned. These data suggest

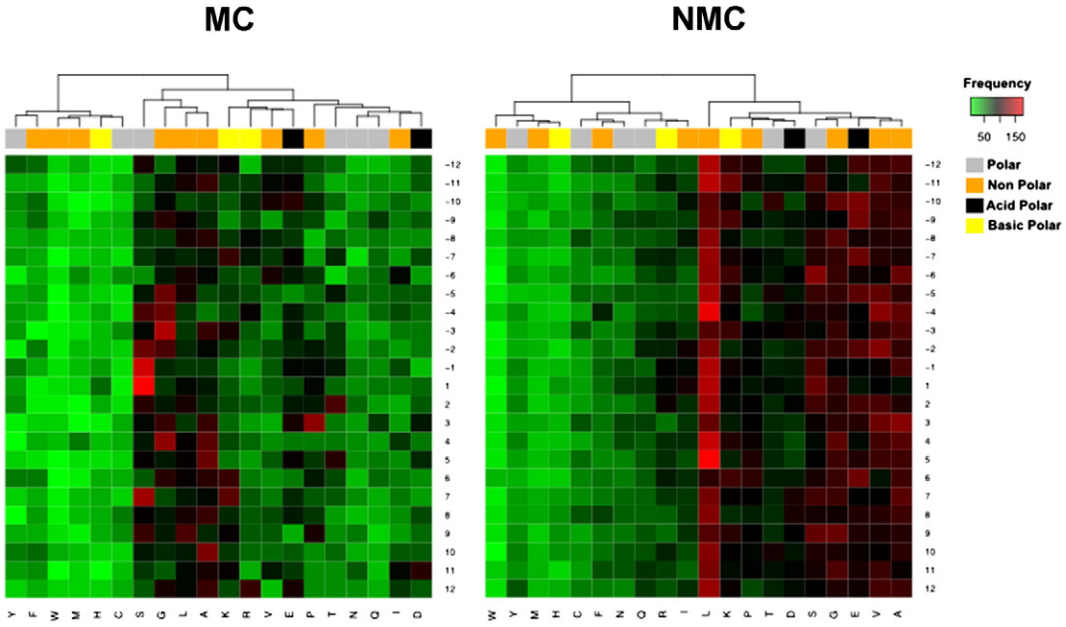

Fig. 4. Occurrence frequency of the amino acids in peptides flanking Cys sites. For each amino acid (row), each column corresponds to a location within a window size of 25 amino acids centred on a Cys residue (location 0, not shown). Results of the hierarchical clustering are shown by dendrograms.

Table 2
Positions where significant differences in the transition matrices have been detected.

TM	Position	Amino acids	*P*-value
MC	−2	G/S	0.082
	1	C/S	0.006
NMC	1	C/F	0.002
	1	C/R	0.007
	10	P/G	0.063

that MC have no propensity for a specific secondary structure indicating that this feature could not allow to classify MC and NMC.

3.2.4. Accessibility of the sulphur atoms

Compared to the NMC residues, the mean value of the solvent accessibility of the sulphur atom was significantly higher for the examined MC residues (0.3 ± 0.2 and 8.6 ± 2.9 Å^2, respectively; $P < 0.01$, *t*-test; Table 3), indicating that the accessibility of the sulphur atom could play a role in determining the susceptibility of a Cys residue to the succination, and could be considered for predictive purpose. Nevertheless, although some enrichment in highly exposed sulphur atoms was determined for the MC, some of them were predicted to be buried (i.e., with exposure value ≤1.0 Å^2), and not accessible to a 1.4 Å-probe (to mimic the water molecule).

3.2.5. Acid dissociation constant of the thiol functional groups

Thiolate anions are known to be much more reactive toward Michael reactions than their protonate counterparts. The acid dissociation constant (pK_a) of the sulfhydryl functional group correlates with the reactivity of the Cys residue, with lowered pK_a indicating an activated Cys. Thereby, the pK_a values of the selected Cys residues were computed with the empirical predictor PropKa. The average pK_a was 8.4 ± 0.5, and comparable values were found for MC and NMC (8.3 ± 0.5 and 8.5 ± 0.8, respectively; Table 3). In addition, although uncommon low pK_a values were determined for some Cys residues, they were in both subsets, suggesting that the pK_a value could not allow classification of Cys residues with regard to their susceptibility to succination.

3.3. Performance of the succination degree as descriptor of modifiable and non-modifiable sites

As significant differences were determined when the composition of the Cys-flanking peptides where examined, a sequence-based descriptor of Cys residues is here proposed (see Method). A better performance as descriptor of modifiable and non-modifiable sites has been obtained considering the SD ($x = 12$ amino acids) computed using the PFMs (Fig. 5A) compared to TMs (Fig. 5B), as indicated by the area under the ROC curve values that were 0.854 ± 0.015 for PFMs and 0.596 ± 0.052 for TMs. These results indicate that the PFMs have a descriptive power higher than the TMs, suggesting that the type of residues and their position are critical for a good description but each position is not directly dependent on the previous one.

Then, to evaluate the effect of the window size on the performance of the PFM-based SD, ten sets of fragments (five MC and five NMC) were generated by adopting increasing window sizes ($x = 3$, 6, 9, 12, and 15 amino acids, respectively) and analysed as previously described. As shown in Fig. 5C and Table S6, a window size-dependent increase in the value of all measurements employed was obtained, thus indicating that a better performance of the PFM-based SD as a descriptor on modifiable and non-modifiable sites was achieved by employing larger window sizes.

4. Discussion

Identification of succinated proteins with their 2SC sites is essential for deciphering the molecular mechanism and role of protein succination. In this study, we have compiled a list of proteins known to contain 2SC sites at established locations, and we have analysed the *succinate proteome* in order to identify features related to the specificity of protein succination.

The adduction of FA to Cys is currently retained to proceed by a spontaneous nucleophilic attack of the thiolate form of the sulfhydryl group of certain Cys residues [30,43]. Our findings indicate that discrete features at or near the functional site govern its specificity. For example, mitochondrial proteins have been commonly found in the published sets of succinated proteins, thus suggesting the hypothesis that the

Table 3
Structural features of a subset of representative modifiable and nonmodifiable Cys residues.

Site		Protein	Uniprot ID	Cys residue (position)	Model (PDB code)	Secondary structure	Solvent accessibility (Å^2)	pK_a
		Peroxiredoxin-1	P35700	173	2z9s	NA	0.0	9.6
		Heterogeneous nuclear ribonucleoprotein L	Q8R081	469	3tyt	NA	12.5	5.8
	Hydrophilic	Calbindin	P12658	187	2f33	α-Helix	17.3	6.7
		Destrin	Q9R0P5	23	1ak6	NA	4.5	9.5
		Aconitate hydratase, mitochondrial	Q99KI0	451	7acn	α-Helix	0.2	10.6
Modifiable								
		Glyceraldehyde-3-phosphate dehydrogenase	P16858	150	1u8f	α-Helix	8.0	5.4
		Aconitate hydratase, mitochondrial	Q99KI0	448	7acn	NA	0.0	8.6
	Hydrophobic	Protein DJ-1	Q99LX0	106	1q2u	NA	5.4	10.1
		Creatine kinase U-type, mitochondrial	P30275	317	1qk1	α-Helix	9.1	7.3
		NFU1 iron–sulphur cluster scaffold homologue, mitochondrial	Q9QZ23	213	1veh	α-Helix	28.5	9.0
		Carbonyl reductase [NADPH] 3	Q8K354	150	2hrb	α-Helix	0.0	9.9
		Creatine kinase U-type, mitochondrial	P30275	397	1qk1	α-Helix	1.1	9.5
	Hydrophilic	Nuclear cap-binding protein subunit 1	Q3UYV9	456	3fey	α-Helix	0.0	11.4
		Creatine kinase U-type, mitochondrial	P30275	180	1qk1	NA	1.5	7.6
		S-methyl-5′-thioadenosine phosphorylase	Q9CQ65	145	3ozc	α-Helix	0.6	5.5
Nonmodifiable								
		Enoyl-CoA delta isomerase 1, mitochondrial	P42125	148	1xx4	α-Helix	0.0	10.6
		Alcohol dehydrogenase 1	P00329	287	1u3w	NA	0.0	9.4
	Hydrophobic	Cytoplasmic dynein 1 heavy chain 1	Q9JHU4	3323	3j1t	α-Helix	0.0	4.1
		Aconitate hydratase, mitochondrial	Q99KI0	205	7acn	β-Strand	0.0	11.8
		Thioredoxin-dependent peroxide reductase, mitochondrial	P20108	109	1zye	α-Helix	0.2	5.9

NA = no secondary structure was assigned.

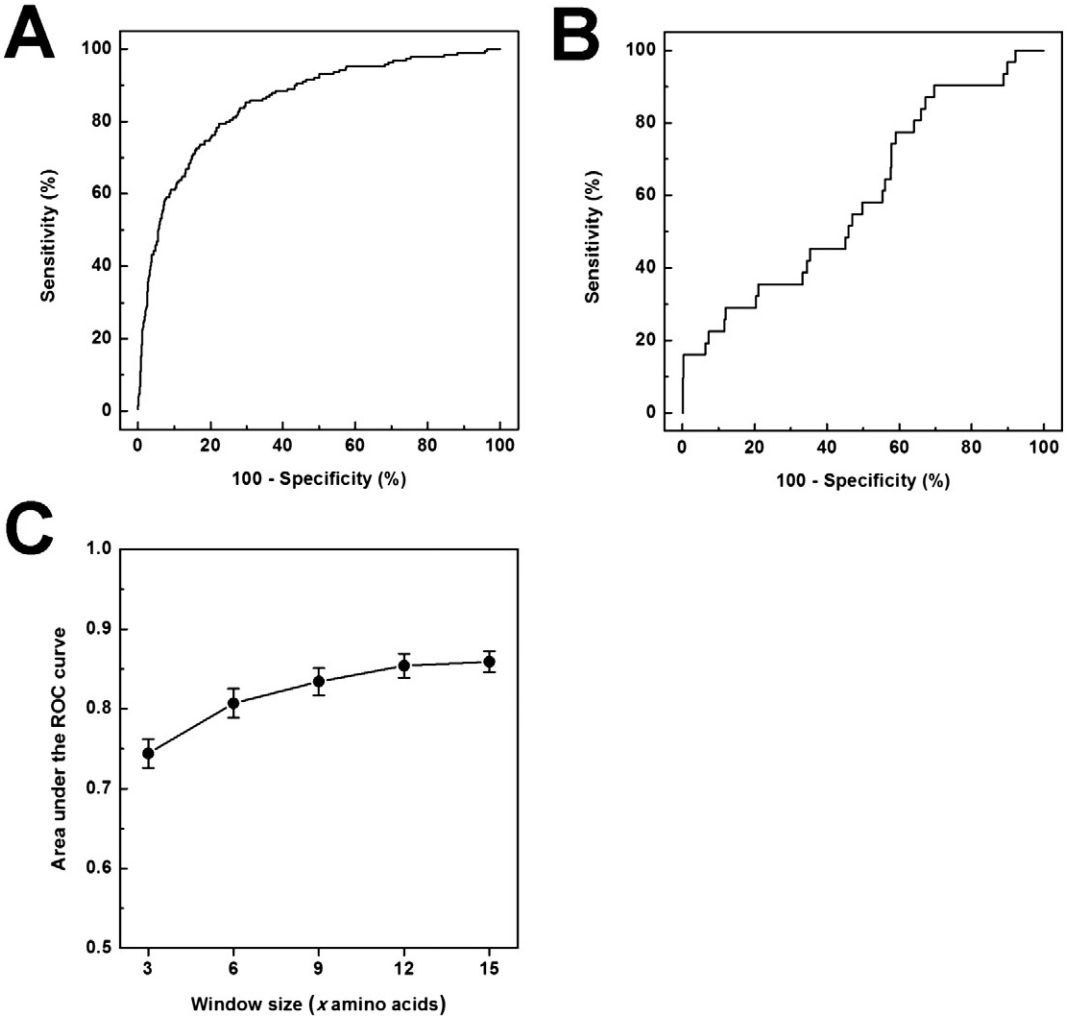

Fig. 5. ROC curves for different succination degrees. (A) ROC curve obtained employing the succination degree defined in Eq. (3); the score functions defined in Eqs. (1) and (2) were instantiated with the PFM entries. (B) ROC curve obtained when the succination degree was computed employing the score functions instantiated with the TM entries. (C) Effect of the window size on the area under the ROC curve. Succination degree was computed employing the score functions instantiated with the PFM entries and increasing the *x* values.

mitochondrial environment could favour this post-translational Cys modification. The thiolate form of a sulfhydryl group is thought to be much more nucleophilic than its protonated counterpart, and is much more readily alkylated by electrophiles [7,44,45]. Given a Cys-sulfhydryl group with a common $pK_a = 9$, at the equilibrium the percentage of the thiolate form is about 2% at a pH 7.2 (i.e., cytosol). However, the percentage is about 10-fold higher at a pH 8.3 (i.e., mitochondrial matrix). A less activating environmental pH may explain the lower abundance of succinated proteins, relative to mitochondria, found in different subcellular compartments. Moreover, FA concentrations in different subcellular compartments (e.g., mitochondria vs cytosol) could also differentially affect protein succination. Accurate determination of FA in different subcellular compartments is technically difficult. However, indirect evidence allows to postulate that FA could persistently accumulate into mitochondria, as a consequence of either persistent nutrient excess or FH-deficiency (see the Introduction section). As protein succination has been related to the intracellular accumulation of FA, our data on the relative abundance of mitochondrial proteins in the succinate proteome support this paradigm.

The local microenvironment surrounding the sulfhydryl group could also explain the specific modifications of intrinsically reactive Cys residues [8]. For example, $Cys^{human\_152/mouse\_150}$ in the active site of glyceraldehyde-3-phosphate dehydrogenase has been repeatedly reported as a 2SC site [29–31]. The low pK_a value (5.4) of the sulfhydryl group of these residues could explain their atypical reactivity toward succination. However, our results indicate that the pK_a value alone could not distinguish MC from NMC. It is reasonable that additional features, which are known to influence the reactivity of Cys residues toward electrophiles (e.g., binding to metals, helix macrodipole) [8] could underlie the specificity of protein succination.

The 2SC sites so far identified form a heterogeneous set. Some features discussed in the literature to explain the reactivity of Cys residues toward electrophiles often could not allow a distinction between MC and NMC, and therefore could not be used for predictive purpose. In contrast, we found that solvent accessibility of the sulphur atom and the amino acid composition of the Cys-flanking peptide could be more useful. In particular, compared with the NMC residues, a higher solvent accessibility of the sulphur atom was found for the MC sites. This finding indicates that the steric hindrance could play a role in determining this modification type. A nucleophilic attack on the β-position of the carbonyl conjugated double bond in FA is the crucial step of a Michael reaction. It is reasonable that exposed, rather than buried, Cys residues could be more easily modified by electrophiles, including FA. Conformational changes due to the mobility (i.e., flexibility and local unfolding) of the structure may be considered to explain the low solvent accessibility of the sulphur atom of certain MC computed from static data, as well as unexpected pK_a values [46]. Nevertheless, as the estimation of the

sulphur accessibility requires 3D-structural information, which is not always available, this feature has limited usefulness. In contrast, the amino acid composition of the Cys-flanking peptides could be employed to perform a fast screening on a proteome-wide scale. Our analysis reveals the absence of a consensus sequence motif around the 2SC sites. This finding is consistent with results previously reported on other types of post-translational Cys modification including, for instance, nitrosylation [36,38,39], and oxidation to sulfenic acid [47]. However, significant differences at distinct positions were determined when both PFMs and TMs were compared. Interestingly, most of these differences involve amino acids in close proximity to the Cys site in the central position, thus suggesting a possible role for these amino acids in governing the succination reaction. Furthermore, the differences in the Cys-flanking peptide allow us to propose a new descriptor of Cys residues. Our findings on the performance of this descriptor not only confirm the hypothesis that protein succination is not a random modification, but also they indicate that features related to the sequence of the Cys-flanking peptides could provide useful data to describe and perhaps predict 2SC sites.

With regard to the long-term goal of developing methods to predict 2SC sites in proteins, our results indicate some important drawbacks. The number of succinated proteins with their 2SC sites so far identified, together with the heterogeneity of these sites are major limits to the development of reliable predictive methods. Future studies focused on the identification of additional succinated proteins could compensate these drawbacks. Furthermore, although the prediction/identification of 2SC is a crucial step in the effort of deciphering the pathophysiological role of protein succination, functional data should be also collected in future studies. Indeed, to date only in few cases the effects of protein succination have been established [12,13,29,48]. Functional data could not only give an insight into the role of protein succination as a new *thiol-switch* mechanism [49,50], but also reveal unknown functional sites to regulate protein function, and functional protein systems [4] targeted by FA. Finally, post-translational Cys modification resulting from the adduction of FA could be mimicked by FA related compounds. They include maleate, the FA's geometric isomer, and dimethyl fumarate, a cell permeable FA derivative, and a drug used to treat patients with psoriasis or relapsing-remitting multiple sclerosis. So far, no attempt has been made to determine whether the relative reactivity of FA vs maleate would make a difference, while the adduction of dimethyl fumarate to proteins has been scantily investigated [51]. Future studies could shed light on these topics, and our results could be of interest to elucidate the mode of action of these effective, albeit enigmatic compounds [52].

5. Conclusions

To the best of our knowledge this is the first computational analysis of the *succinated proteome*. Our results indicate that a relative small proportion of Cys in proteins is modified by FA. In addition, they highlight distinct features related to Cys reactivity which may be considered in future investigations aimed at the development of computational methods to predict 2SC sites in proteins and the deciphering of the role of protein succination.

Author contributions

Conceived and designed the experiments: GM and FC. Performed the experiments GM, ADS, EV, MB, and FC. Analysed the data: GM and MTG. Wrote the paper: GM and FC.

Conflict of interest

The authors declare that they have no conflict of interest.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

This work was supported by funding of the Università degli Studi di Torino, Ricerca Locale Ricerca Locale 2013 (to GM).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2015.11.003.

References

[1] Y.L. Deribe, T. Pawson, I. Dikic, Post-translational modifications in signal integration, Nat. Struct. Mol. Biol. 17 (2010) 666–672.
[2] P. Minguez, L. Parca, F. Diella, D.R. Mende, R. Kumar, M. Helmer-Citterich, A.C. Gavin, V. van Noort, P. Bork, Deciphering a global network of functionally associated post-translational modifications, Mol. Syst. Biol. 8 (2012) 599.
[3] E.C. Miller, J.A. Miller, R.W. Sapp, G.M. Weber, Studies on the protein-bound aminoazo dyes formed in vivo from 4-dimethylaminoazobenzene and its C-monomethyl derivatives, Cancer Res. 9 (1949) 336–343.
[4] S.G. Codreanu, J.C. Ullery, J. Zhu, K.A. Tallman, W.N. Beavers, N.A. Porter, L.J. Marnett, B. Zhang, D.C. Liebler, Alkylation damage by lipid electrophiles targets functional protein systems, Mol. Cell. Proteomics 13 (2014) 849–859.
[5] P. Cohen, The regulation of protein function by multisite phosphorylation — a 25 year update, Trends Biochem. Sci. 25 (2000) 596–601.
[6] B.T. Weinert, S.A. Wagner, H. Horn, P. Henriksen, W.R. Liu, J.V. Olsen, L.J. Jensen, C. Choudhary, Proteome-wide mapping of the drosophila acetylome demonstrates a high degree of conservation of lysine acetylation, Sci. Signal. 4 (2011), ra48.
[7] S.M. Marino, V.M. Gladyshev, Analysis and functional prediction of reactive cysteine residues, J. Biol. Chem. 287 (2012) 4419–4425.
[8] N.J. Pace, E. Weerapana, Diverse functional roles of reactive cysteines, ACS Chem. Biol. 8 (2013) 283–296.
[9] G. Aldini, M.R. Domingues, C.M. Spickett, P. Domingues, A. Altomare, F.J. Sánchez-Gómez, C.L. Oeste, D. Pérez-Sala, Protein lipoxidation: detection strategies and challenges, Redox Biol. 5 (2015) 253–266.
[10] N.L. Alderson, Y. Wang, M. Blatnik, N. Frizzell, M.D. Walla, T.J. Lyons, N. Alt, J.A. Carson, R. Nagai, S.R. Thorpe, J.W. Baynes, S-(2-succinyl)cysteine: a novel chemical modification of tissue proteins by a Krebs cycle intermediate, Arch. Biochem. Biophys. 450 (2006) 1–8.
[11] R. Nagai, J.W. Brock, M. Blatnik, J.E. Baatz, J. Bethard, M.D. Walla, S.R. Thorpe, J.W. Baynes, N. Frizzell, Succination of protein thiols during adipocyte maturation: a biomarker of mitochondrial stress, J. Biol. Chem. 282 (2007) 34219–34228.
[12] M. Blatnik, S.R. Thorpe, J.W. Baynes, Succination of proteins by fumarate: mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes, Ann. N. Y. Acad. Sci. 1126 (2008) 272–275.
[13] N. Frizzell, M. Rajesh, M.J. Jepson, R. Nagai, J.A. Carson, S.R. Thorpe, J.W. Baynes, Succination of thiol groups in adipose tissue proteins in diabetes: succination inhibits polymerization and secretion of adiponectin, J. Biol. Chem. 284 (2009) 25772–25781.
[14] S.A. Thomas, K.B. Storey, J.W. Baynes, N. Frizzell, Tissue distribution of S-(2-succino)cysteine (2SC), a biomarker of mitochondrial stress in obesity and diabetes, Obesity (Silver Spring) 20 (2012) 263–269.
[15] N. Frizzell, S.A. Thomas, J.A. Carson, J.W. Baynes, Mitochondrial stress causes increased succination of proteins in adipocytes in response to glucotoxicity, Biochem. J. 445 (2012) 247–254.
[16] J.S. Isaacs, Y.J. Jung, D.R. Mole, S. Lee, C. Torres-Cabala, Y.L. Chung, M. Merino, J. Trepel, B. Zbar, J. Toro, P.J. Ratcliffe, W.M. Linehan, L. Neckers, HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability, Cancer Cell 8 (2005) 143–153.
[17] P.J. Pollard, J.J. Brière, N.A. Alam, J. Barwell, E. Barclay, N.C. Wortham, T. Hunt, M. Mitchell, S. Olpin, S.J. Moat, I.P. Hargreaves, S.J. Heales, Y.L. Chung, J.R. Griffiths, A. Dalgleish, J.A. McGrath, M.J. Gleeson, S.V. Hodgson, R. Poulsom, P. Rustin, I.P. Tomlinson, Accumulation of Krebs cycle intermediates and over-expression of HIF1α in tumours which result from germline FH and SDH mutations, Hum. Mol. Genet. 14 (2005) 2231–2239.
[18] H. Ashrafian, L. O'Flaherty, J. Adam, V. Steeples, Y.L. Chung, P. East, S. Vanharanta, H. Lehtonen, E. Nye, E. Hatipoglu, M. Miranda, K. Howarth, D. Shukla, H. Troy, J. Griffiths, B. Spencer-Dene, M. Yusuf, E. Volpi, P.H. Maxwell, G. Stamp, R. Poulsom, C.W. Pugh, B. Costa, C. Bardella, M.F. Di Renzo, M.I. Kotlikoff, V. Launonen, L. Aaltonen, M. El-Bahrawy, I. Tomlinson, P.J. Pollard, Expression profiling in progressive stages of fumarate-hydratase deficiency: the contribution of metabolic changes to tumorigenesis, Cancer Res. 70 (2010) 9153–9165.
[19] L. O'Flaherty, J. Adam, L.C. Heather, A.V. Zhdanov, Y.L. Chung, M.X. Miranda, J. Croft, S. Olpin, K. Clarke, C.W. Pugh, J. Griffiths, D. Papkovsky, H. Ashrafian, P.J. Ratcliffe, P.J. Pollard, Dysregulation of hypoxia pathways in fumarate hydratase-deficient cells is

independent of defective mitochondrial metabolism, Hum. Mol. Genet. 19 (2010) 3844–3851.
[20] C. Frezza, L. Zheng, O. Folger, K.N. Rajagopalan, E.D. MacKenzie, L. Jerby, M. Micaroni, B. Chaneton, J. Adam, A.A. Hedley, G. Kalna, I.P. Tomlinson, P.J. Pollard, D.G. Watson, R.J. Deberardinis, T. Shlomi, E. Ruppin, E. Gottlieb, Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase, Nature 477 (2011) 225–228.
[21] I.P. Tomlinson, N.A. Alam, A.J. Rowan, E. Barclay, E.E. Jaeger, D. Kelsell, I. Leigh, P. Gorman, H. Lamlum, S. Rahman, R.R. Roylance, S. Olpin, S. Bevan, K. Barker, N. Hearle, R.S. Houlston, M. Kiuru, R. Lehtonen, A. Karhu, S. Vilkki, P. Laiho, C. Eklund, O. Vierimaa, K. Aittomäki, M. Hietala, P. Sistonen, A. Paetau, R. Salovaara, R. Herva, V. Launonen, L.A. Aaltonen, Multiple Leiomyoma Consortium, Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer, Nat. Genet. 30 (2002) 406–410.
[22] J. Adam, E. Hatipoglu, L. O'Flaherty, N. Ternette, N. Sahgal, H. Lockstone, D. Baban, E. Nye, G.W. Stamp, K. Wolhuter, M. Stevens, R. Fischer, P. Carmeliet, P.H. Maxwell, C.W. Pugh, N. Frizzell, T. Soga, B.M. Kessler, M. El-Bahrawy, P.J. Ratcliffe, P.J. Pollard, Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling, Cancer Cell 20 (2011) 524–537.
[23] C. Bardella, M. El-Bahrawy, N. Frizzell, J. Adam, N. Ternette, E. Hatipoglu, K. Howarth, L. O'Flaherty, I. Roberts, G. Turner, J. Taylor, K. Giaslakiotis, V.M. Macaulay, A.L. Harris, A. Chandra, H.J. Lehtonen, V. Launonen, L.A. Aaltonen, C.W. Pugh, R. Mihai, D. Trudgian, B. Kessler, J.W. Baynes, P.J. Ratcliffe, I.P. Tomlinson, P.J. Pollard, Aberrant succination of proteins in fumarate hydratase-deficient mice and HLRCC patients is a robust biomarker of mutation status, J. Pathol. 225 (2011) 4–11.
[24] A. Ooi, J.C. Wong, D. Petillo, D. Roossien, V. Perrier-Trudova, D. Whitten, B.W. Min, M.H. Tan, Z. Zhang, X.J. Yang, M. Zhou, B. Gardie, V. Molinié, S. Richard, P.H. Tan, B.T. Teh, K.A. Furge, An antioxidant response phenotype shared between hereditary and sporadic type 2 papillary renal cell carcinoma, Cancer Cell 20 (2011) 511–523.
[25] C. Frezza, P.J. Pollard, E. Gottlieb, Inborn and acquired metabolic defects in cancer, J. Mol. Med. (Berl) 89 (2011) 213–220.
[26] M. Yang, T. Soga, P.J. Pollard, J. Adam, The emerging role of fumarate as an oncometabolite, Front. Oncol. 2 (2012) 85.
[27] L. Zheng, S. Cardaci, L. Jerby, E.D. MacKenzie, M. Sciacovelli, T.I. Johnson, E. Gaude, A. King, J.D. Leach, R. Edrada-Ebel, A. Hedley, N.A. Morrice, G. Kalna, K. Blyth, E. Ruppin, C. Frezza, E. Gottlieb, Fumarate induces redox-dependent senescence by modifying glutathione metabolism, Nat. Commun. 6 (2015) 6001.
[28] G.G. Piroli, A.M. Manuel, A.C. Clapper, M.D. Walla, J.E. Baatz, R.D. Palmiter, A. Quintana, N. Frizzell, Succination is increased on select proteins in the brainstem of the Ndufs4 knockout mouse, a model of Leigh syndrome, Mol. Cell. Proteomics (2015)http://dx.doi.org/10.1074/mcp.M115.051516 (*In press*).
[29] N. Ternette, M. Yang, M. Laroyia, M. Kitagawa, L. O'Flaherty, K. Wolhulter, K. Igarashi, K. Saito, K. Kato, R. Fischer, A. Berquand, B.M. Kessler, T. Lappin, N. Frizzell, T. Soga, J. Adam, P.J. Pollard, Inhibition of mitochondrial aconitase by succination in fumarate hydratase deficiency, Cell Rep. 3 (2013) 689–700.
[30] E.D. Merkley, T.O. Metz, R.D. Smith, J.W. Baynes, N. Frizzell, The succinated proteome, Mass Spectrom. Rev. 33 (2014) 98–109.
[31] M. Yang, N. Ternette, H. Su, R. Dabiri, B.M. Kessler, J. Adam, B.T. Teh, P.J. Pollard, The succinated proteome of FH-mutant tumours, Metabolites 4 (2014) 640–654.
[32] D.J. Jones, Radical-free biology of oxidative stress, Am. J. Physiol. Cell Physiol. 295 (2008) C849–C868.
[33] D.E. Fomenko, W. Xing, B.M. Adair, D.J. Thomas, V.N. Gladyshev, High-throughput identification of catalytic redox-active cysteine residues, Science 315 (2007) 387–389.
[34] Y. Xue, Z. Liu, X. Gao, C. Jin, L. Wen, X. Yao, J. Ren, GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm, PLoS One 5 (2010), e11290.
[35] L.L. Hu, S.B. Wan, S. Niu, X.H. Shi, H.P. Li, Y.D. Cai, K.C. Chou, Prediction and analysis of protein palmitoylation sites, Biochimie 93 (2011) 489–496.
[36] B.Q. Li, L.L. Hu, S. Niu, Y.D. Cai, K.C. Chou, Predict and analyze S-nitrosylation modification sites with the mRMR and IFS approaches, J. Proteomics 75 (2012) 1654–1665.
[37] C. Sun, Z.Z. Shi, X. Zhou, L. Chen, X.M. Zhao, Prediction of S-glutathionylation sites based on protein sequences, PLoS One 8 (2013), e55512.
[38] T.M. Greco, R. Hodara, I. Parastatidis, H.F. Heijnen, M.K. Dennehy, D.C. Liebler, H. Ischiropoulos, Identification of S-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 7420–7425.
[39] S.M. Marino, V.N. Gladyshev, Structural analysis of cysteine S-nitrosylation: a modified acid-based motif and the emerging role of trans-nitrosylation, J. Mol. Biol. 395 (2010) 844–859.
[40] X. Zhao, Q. Ning, H. Chai, Z. Ma, Accurate in silico identification of protein succinylation sites using an iterative semi-supervised learning technique, J. Theor. Biol. 374 (2015) 60–65.
[41] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[42] D.E. Schones, P. Sumazin, M.Q. Zhang, Similarity of position frequency matrices for transcription factor binding sites, Bioinformatics 21 (2005) 307–313.
[43] H. Lin, X. Su, B. He, Protein lysine acylation and cysteine succination by intermediates of energy metabolism, ACS Chem. Biol. 7 (2012) 947–960.
[44] A. Higdon, A.R. Diers, J.Y. Oh, A. Landar, V.M. Darley-Usmar, Cell signalling by reactive lipid species: new concepts and molecular mechanisms, Biochem. J. 442 (2012) 453–464.
[45] R.M. Lopachin, T. Gavin, A. Decaprio, D.S. Barber, Application of the hard and soft, acids and bases (HSAB) theory to toxicant-target interactions, Chem. Res. Toxicol. 25 (2012) 239–251.
[46] S.M. Marino, Protein flexibility and cysteine reactivity: influence of mobility on the H-bond network and effects on pKa prediction, Protein J. 33 (2014) 323–336.
[47] F.R. Salsbury Jr., S.T. Knutson, L.B. Poole, J.S. Fetrow, Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid, Protein Sci. 17 (2008) 299–312.
[48] G.G. Piroli, A.M. Manuel, M.D. Walla, M.J. Jepson, J.W. Brock, M.P. Rajesh, R.M. Tanis, W.E. Cotham, N. Frizzell, Identification of protein succination as a novel modification of tubulin, Biochem. J. 462 (2014) 231–245.
[49] B. Groitl, U. Jakob, Thiol-based redox switches, Biochim. Biophys. Acta 1844 (2014) 1335–1343.
[50] L.I. Leichert, T.P. Dick, Incidence and physiological relevance of protein thiol switches, Biol. Chem. 396 (2015) 389–399.
[51] A.M. Manuel, N. Frizzell, Adipocyte protein modification by Krebs cycle intermediates and fumarate ester-derived succination, Amino Acids 45 (2013) 1243–1247.
[52] G. Miglio, A. Romanò, A. Sabatino, F. Cordero, M. Beccuti, M.T. Giraudo, E. Veglia, R. Fantozzi, The Mode of Action of Dimethyl Fumarate: Protein Succination and Anti-pyroptotic Effects, Proceedings of the British Pharmacological Society at http://www.pA2online.org/abstracts/Vol12Issue3abst104P.pdf