Online Supporting Information

Determination of free acidic and alkaline residues of protein via moving reaction boundary titration in microdevice electrophoresis

Hou-yu Wang[1], Si Li[1], Yun-yun Tang[1], Jing-yu Dong[1, 2], Liu-yin Fan[1], and Cheng-xi Cao[1]*[1]

[1]*Laboratory of Bioseparation and Analytical Biochemistry, State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China*

[2]*School of Chemistry and Molecule Engineering, East China University of Science and Technology, Shanghai 200234*

Table S1 Parameters of seven model proteins.

Protein	MW(KDa)	Isoelectric point	N_{ACR} [a)]	N_{ALR} [a)]
Mb	17.5	7.07	23	31
Hb	67	7.23	64	84
BSA	66	7.8	149	124
Cyt C	13.8	4.47	17	23
Trypsin	23.3	10.5	30	26
Pepsin	35	8.1	67	27
CT	11.7	4.9	17	16

[a)] The value of N_{ACR} and N_{ALR} are calculated through the corresponding amino acid sequences of seven proteins according to the protein database of NCBI, the detailed calculation process is shown in the section of Data Analysis.

Chemicals

Two standard reagents of potassium hydrogen phthalate and sodium tetraborate were obtained from Shanghai Chemical reagent Co. (Shanghai, China). Acrylamide and Bis-acrylamide were bought from Fluka (Switzerland). N, N, N', N-tetramethyl- ethylenediamine (TEMED) was from Sigma (USA). The other chemicals used herein were ammonium persulfate (analytical reagent grade, AR), sodium hydroxide (AR), hydrochloric acid (AR) and potassium chloride (AR) and phenolphthalein (chemical reagent grade, CR) as well as bromophenol blue (CR) were purchased from Shanghai Chemical reagent Co. (Shanghai, China).

* The correspondent author: Prof. Cheng-Xi Cao, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; Fax: +86 21-3420 5820; E-mail: cxcao@sjtu.edu.cn.

Data Analysis

Different concentrations (0, 1, 2, 3 and 4 mg/mL) of model proteins were chosen for experiments of MRB titration. A blank control (no any protein) was used for obtaining the values of v_{OH}^- and v_H^+ in PAG, correcting the values of v_{OH}^- and v_H^+ used in Eq. (5)-(6) and Eq. (7)-(8) when detecting free acidic and alkaline groups of immobilized protein, respectively. A same experiment of MRB titration was repeated three times, and the distance of the boundary motion was referred with a ruler.

After a base titration (Scheme 1A), the boundary velocity could be computed with Eq. (1) and the photographs, and the acidic groups and residues of protein could be calculated with Eq. (6) and Eq. (7), respectively. Similarly, the boundary velocity in the acid titration (Scheme 1B) could be computed with Eq. (1) and the relevant photographs, and the alkaline groups and residues of protein could be calculated with Eq. (9) and Eq. (10), respectively.

The titrating sites of acidic residues in protein concerned Asp (e), Glu (d), Tyr (y) and Cys (c). And the titrating sites of alkaline residues included His (h), Lys (k), 2×Arg (r), Trp (w). The actual values of N_{ACR} and N_{ALR} were calculated from the National Center of Biotechnology Information (NCBI) protein database which was publicly available for community use.

1. The amino acids sequence of Mb from equine skeletal muscle is listed as follow:
   mglsdgewqq vlnvwgkvea diaghgqevl irlftghpet lekfdkfkhl kteaemkase
   dlkkhgtvvl talggilkkk ghheaelkpl aqshatkhki pikylefisd aiihvlhskh
   pgdfgadaqg amtkalelfr ndiaakykel gfqg

The amino acids sequence can be referred to the URL of NCBI:
(http://www.ncbi.nlm.nih.gov/protein/NP_001157488.1)
The N_{ACR} of Mb is 23 (8d +13e +2y+1c).
The N_{ALR} of Mb is 31 (11h + 16k + 2r + 2w).

2. The amino acids sequence of one chain of Hb from human is listed as follow:
   mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk
   vkahgkkvlg afsdglahld nlkgtfatls elhcdklhvd penfr

The amino acids sequence can be referred to the URL of NCBI
(http://www.ncbi.nlm.nih.gov/protein/ABG47031.1)
The N_{ACR} of Hb (4Chains) is 64: 4(7d +7e +1y+1c).
The N_{ALR} of Hb (4Chains) is 84: 4(5h + 8k +2 × 3r +2w).

3. The amino acids sequence of BSA from bovine is listed as follow:
   mkwvtfisll llfssaysrg vfrrdthkse iahrfkdlge ehfkglvlia fsqylqqcpf
   dehvklvnel tefaktcvad eshagceksl htlfgdelck vaslretygd madccekqep
   ernecflshk ddspdlpklk pdpntlcdef kadekkfwgk ylyeiarrhp yfyapellyy
   ankyngvfqe ccqaedkgac llpkietmre kvltssarqr lrcasiqkfg eralkawsva
   rlsqkfpkae fvevtklvtd ltkvhkecch gdllecaddr adlakyicdn qdtissklke
   ccdkplleks hciaevekda ipenlpplta dfaedkdvck nyqeakdafl gsflyeysrr
   hpeyavsvll rlakeyeatl eeccakddph acystvfdkl khlvdepqnl ikqncdqfek
   lgeygfqnal ivrytrkvpq vstptlvevs rslgkvgtrc ctkpesermp ctedylslil
   nrlcvlhekt pvsekvtkcc teslvnrrpc fsaltpdety vpkafdeklf tfhadictlp

dtekqikkqt alvellkhkp kateeqlktv menfvafvdk ccaaddkeac favegpklvv stqtala

The amino acids sequence can be referred to the URL of NCBI
(http://www.ncbi.nlm.nih.gov/protein/CAA76847.1)
The N_{ACR} of BSA is 149 (39d + 58e + 19y+33c).
The N_{ALR} of BSA is 124 (15h + 57k + 2 × 25r +2w).

4. The amino acids sequence of Cyt C from bovine is listed as follow:
mgdvekgkki fvqkcaqcht vekggkhktg pnlhglfgrk tgqapgfsyt danknkgitw geetlmeyle npkkyipgtk mifagikkkg eredliaylk katne

The amino acids sequence can be referred to the URL of NCBI
(http://www.ncbi.nlm.nih.gov/protein/NP_001039526.1)
The N_{ACR} of Cyt C is 17 (3d + 9e +3y+ 2c).
The N_{ALR} of Cyt C is 23(2h +18k + 2r + w).

5. The amino acids sequence of trypsin from porcine is listed as follow:
ivggytcaan sipyqvslns gshfcggsli nsqwvvsaah cyksriqvrl gehnidvleg neqfinaaki ithpnfngnt ldndimlikl sspatlxsrv atvslprsca aagteclisg wgntkssgss ypsllqclka pvlsdsscks sypgqitgnm icvgfleggk dscqgdsggp vvcngqlqgi vswgygcaqk nkpgvytkvc nyvnwiqqti aan

The amino acids sequence can be referred to the URL of NCBI
(http://www.ncbi.nlm.nih.gov/protein/1C9P_A)
The N_{ACR} of trypsin is 30 (6d +5e +8y+11c).
The N_{ALR} of trypsin is 26 (4h +10k +2×4r+ 4w).

6. The amino acids sequence of pepsin from porcine gastric mucosa is listed as follow:
lvkvplvrkk slrqnlikdg klkdflkthk hnpaskyfpe aaaligdepl enyldteyfg tigigtpaqd ftvifdtgss nlwvpsvycs slacsdhnqf npddsstfea tsqelsityg tgsmtgilgy dtvqvggisd tnqifglset epgsflyyap fdgilglayp sisasgatpv fdnlwdqglv sqdlfsvyls snddsgsvvl lggidssyyt gslnwvpvsv egywqitlds itmdgetiac sggcqaivdt gtslltgpts aianiqsdig asensdgemv iscssidslp divftidgvq yplspsayil qdddsctsgf egmdvptssg elwilgdvfi rqyytvfdra nnkvglapva

The amino acids sequence can be referred to the URL of NCBI
（http://www.ncbi.nlm.nih.gov/protein/3PSG_A）
The N_{ACR} of pepsin is 67 (32d +13e +16y+6c).
The N_{ALR} of pepsin is 27 (2h +13k + 2 × 4r +4w).

7. The amino acids sequence of from CT from bovine is listed as follow:
ivngeeavpg swpwqvslqd ktgfhfcggs linenwvvta ahcgvttsdv vvagefdqgs ssekiqklki akvfknskyn sltinnditl lklstaasfs qtvsavclps asddfaagtt cvttgwgltr y

The amino acids sequence can be referred to the URL of NCBI
(http://www.ncbi.nlm.nih.gov/protein/1GHB_F)
The N_{ACR} of CT is 17 (6d +5e +2y+4c).
The N_{ALR} of CT is 16 (2h + 8k + 2 × r + 4w).

Optimization of Conditions

To achieve reproducible experiments with the developed method, we at first optimized the electric field strength (*E*), NaOH and HCl concentrations, and ionic strength in the whole system. Fig. S1A showed the influence of electric field strenght on the boundary velocity in the MRB titratin system. As for applied voltage, at a relative low field strength (250 V/m), the boundary moved slowly. When high field strength (2500 V/m) was used, the boundary migrated very quickly, but a poor linearity was present between the boundary displacement distance (*D*) and running time (*t*). In addition, at high field strength, the capillary was felt very hot. Condisering fair boundary displacement but not much Joule heating, 1250 V/m was finally chosen as the optimized field strenght (Fig. S1A).

It was further observed in Fig. S1B that the boundary indicated by the acid-base indicator (phenolphthalein, or bromophenol blue) moved slowly and was not clear if a low concentration NaOH or HCl (less than 15 mM) was used; while the boundary migrated fast and became clear if high concentration base or acid (20 mM) was applied in the given system. Whereas, there was a poor linearity between *D* and *t*, besides much Joule heating if higher concentration NaOH or HCl was used. Finnally, 20 mM NaOH and HCl were chosen for base and acid titration the experiments, respectively (Fig. S1B).

Fig. S1C exhibited the impact of background electrolyte of potassium chloride on the movement of MRB. It was demonstrated in Panel C that high concentration background eletrolyte (KCl) had no obvious effect on the boundary velocity, but could result in much more Joule heating. The ionic strength of the anodic solution, cathodic solution and PAG was selected at 100 mM to keep similar conductance through the whole titration system.

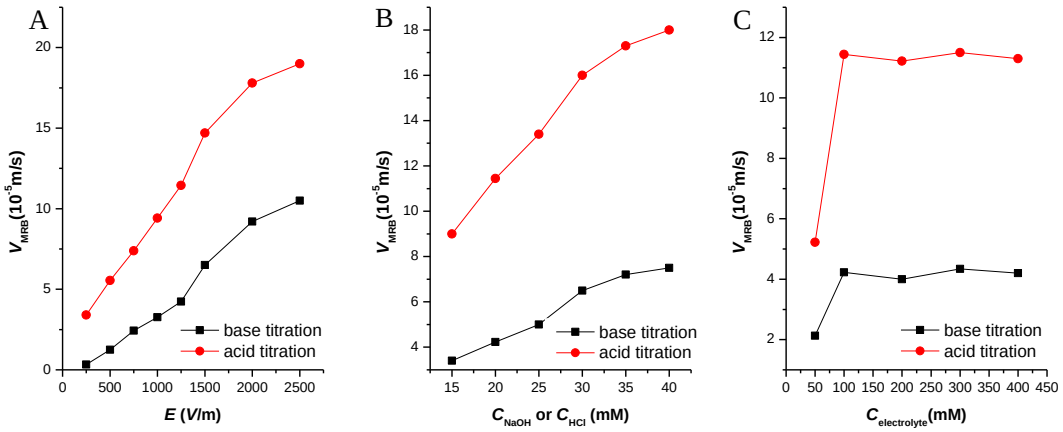

Fig. S1 Influence of electric field strength (*E*) (**A**), NaOH and HCl concentration (**B**) and ionic strength of background electrolyte (**C**) on boundary velocity (V_{MRB}). Conditions in Panel A: 20 mM NaOH or 20 mM HCl, 2.0 mg/mL BSA, 0.02μL 1.0% (w/v) phenolphthalein or 0.1% (w/v) bromophenol blue, 500-2500 *V*/m, PAG (15% T and 4.0% C), 100 mM background electrolyte, 500 μm i.d. and 4 cm length capillary, and air-conditioned room (22-25℃). Conditions in Panel B: 15–40 mM NaOH or HCl. the other conditions are the same as those in Panel A. Conditions in Panel C: 50-400 mM background electrolyte. The other conditions are the same as those in Panel A.

Table S2 Linearity curves between the boundary movement (*D*) and the running time (*t*) in the MRBT system

	MRB base titration		MRB acid titration	
Protein	Linearity equation [a]	*R* [b]	Linearity equation [c]	*R*
Mb	*D*=0.0048*t*-0.0603	0.9989	*D*=0.0103*t*+0.08	0.9926
Hb	*D*=0.0039*t*-0.0103	0.9985	*D*=0.0101*t*+0.17	0.996
BSA	*D*=0.0044*t*+0.0414	0.9922	*D*=0.0107*t*+0.095	0.994
Cyt C	*D*=0.0033*t*+0.1	0.9999	*D*=0.0104*t*+0.07	0.9955
Trypsin	*D*=0.0044*t*-0.0029	0.9901	*D*=0.0113*t*+0.05	0.9948
Pepsin	*D*=0.0036*t*+0.0633	0.9947	*D*=0.0113*t*+0.09	0.9987
CT	*D*=0.0034*t*-0.02	0.9965	*D*=0.0115t+0.05	0.9921

[a] The linearity equations correspond with those in Fig. 1A.

[b] R is the correlation coefficient.

[c] The linearity equations correspond with those in Fig. 2A.