Supporting Information

Ionization and Conformational Equilibria of Citric Acid: Delocalized Proton Binding in Solution

Sergio Madurga[a*], Miroslava Nedyalkova[b], Francesc Mas[a] and Josep Lluís Garcés[c]

[a] Materials Science and Physical Chemistry Department & Research Institute of Theoretical and Computational Chemistry (IQTCUB) of Barcelona University (UB), C/ Martí i Franquès, 1. 08028 Barcelona (Catalonia, Spain)

[b] Inorganic Chemistry Department, Faculty of Chemistry and Pharmacy "St Kliment Ohridski", University of Sofia, 1 James Bourchier Blvd. 1164, Sofia (Bulgaria)

[c] Chemistry Department and AGROTECNIO, University of Lleida (UdL), Rovira Roure, 191, 25198, Lleida (Catalonia, Spain)

* corresponding author. e-mail: s.madurga@ub.edu; `www.ub.edu/biophyschem`

The reported best-fitted parameters correspond to the unique local minimum of the fitting procedure. An additional argument to the robustness of the found microscopic binding parameters.

Let us outline an argument to show that the best-fitted parameters shown in this work correspond to a unique minimum of the fitting process. Firstly note that by rewriting the equations

$$\begin{aligned} \delta &= \delta_0 + \delta_{13}\theta_1 + (\delta_{\mathrm{P}} - \delta_0 - \delta_{13})\,\theta_2 \\ \Delta &= \Delta_0 + \Delta_{13}\theta_1 + (\Delta_{\mathrm{P}} - \Delta_0 - \Delta_{13})\,\theta_2 \end{aligned} \tag{1}$$

replacing

$$\begin{cases} \theta_1\,(a_{\mathrm{H}}) = \pi_1\,(\mathrm{H\Theta\Theta})\,P_1\,(a_{\mathrm{H}}) + (1 - \pi_2\,(\mathrm{HH\Theta}))\,P_2\,(a_{\mathrm{H}}) + P_3\,(a_{\mathrm{H}}) \\ \theta_2\,(a_{\mathrm{H}}) = (1 - 2\pi_1\,(\mathrm{H\Theta\Theta}))\,P_1\,(a_{\mathrm{H}}) + 2\pi_2\,(\mathrm{HH\Theta})\,P_2\,(a_{\mathrm{H}}) + P_3\,(a_{\mathrm{H}}) \end{cases} \tag{2}$$

in eqn. (1), δ and Δ can be expressed in terms of the macroscopic probabilities $P_i\,(a_{\mathrm{H}})$, which only depend on the macroscopic dissociation pK_i-values *via*

$$P_n\,(a_{\mathrm{H}}) = \frac{\overline{K}_n a_{\mathrm{H}}^n}{\Xi}\ ;\ \Xi = \sum_{i=0,\ldots,N} \overline{K}_i a_H^i\ ;\ \mathrm{p}K_i = -\log K_i = \log\left(\overline{K}_{N-i+1}/\overline{K}_{N-i}\right) \tag{3}$$

Introducing Eqns. (2) in (1) we obtain δ and Δ as a linear combination of the macroscopic probabilities. As a result we obtain the equations

$$\begin{aligned} \delta &= \delta_0 + X\,P_1\,(a_{\mathrm{H}}) + Y\,P_2\,(a_{\mathrm{H}}) + (\delta_{\mathrm{P}} - \delta_0)\,P_3\,(a_{\mathrm{H}}) \\ \Delta &= \Delta_0 + x\,P_1\,(a_{\mathrm{H}}) + y\,P_2\,(a_{\mathrm{H}}) + (\Delta_{\mathrm{P}} - \Delta_0)\,P_3\,(a_{\mathrm{H}}) \end{aligned} \tag{4}$$

where X, Y, x and y are new parameters to be fitted and defined as

$$\begin{aligned} X &= (2\delta_0 - 2\delta_{\mathrm{P}} + 3\delta_{13})\,\pi_1\,(\mathrm{H\Theta\Theta}) + (\delta_{\mathrm{P}} - \delta_0 - \delta_{13}) \\ Y &= (2\delta_{\mathrm{P}} - 2\delta_0 - 3\delta_{13})\,\pi_2\,(\mathrm{HH\Theta}) + \delta_{13} \\ x &= (2\Delta_0 - 2\Delta_{\mathrm{P}} + 3\Delta_{13})\,\pi_1\,(\mathrm{H\Theta\Theta}) + (\Delta_{\mathrm{P}} - \Delta_0 - \Delta_{13}) \\ y &= (2\Delta_{\mathrm{P}} - 2\Delta_0 - 3\Delta_{13})\,\pi_2\,(\mathrm{HH\Theta}) + \Delta_{13} \end{aligned} \tag{5}$$

But we have very good values for the macroscopic dissociation constants, obtained previously by potentiometry. Recalling too that δ_{P}, δ_0, Δ_{P} and Δ_0 do not need to be fitted (they are directly obtained from the chemical shifts at very low and very high pH-values), we get that Eqns. (4) reduce to a *linear* fitting of X, Y, x and y. As a result, only *one minimum* can be obtained in fitting eqn. (4) to the experimental NMR curves. Thus, the values of X, Y, x and y are uniquely determined. The obtained values are (within the experimental and numerical error) X=0.137, Y=0.266, x=-15.14, y=-5.65.

With those values for X, Y, x and y at hand, the problem reduces to show that the solution of the system of Eqns. (5), containing four unknowns ($\pi_1\,(\mathrm{H\Theta\Theta})$, $\pi_2\,(\mathrm{HH\Theta})$, δ_{13} and Δ_{13}) is unique for values of $\pi_1\,(\mathrm{H\Theta\Theta})$ and $\pi_2\,(\mathrm{HH\Theta})$ lying between 0 and 1. This is an elementary exercise of algebra and numerical computation by eliminating δ_{13} and Δ_{13} from the system of equations and analysing the solutions for the resulting 2-equations system for $\pi_1\,(\mathrm{H\Theta\Theta})$ and $\pi_2\,(\mathrm{HH\Theta})$. It is found that the only solution (within the experimental and numerical error) corresponds to

$$\pi_1\,(\mathrm{H\Theta\Theta}) \simeq 0.5\,;\; \pi_2\,(\mathrm{HH\Theta}) \simeq 0.20 \tag{6}$$

which are very close to the obtained values by direct fitting of all the microscopic binding parameters to the experimental NMR shift data. Furthermore, recalling that

$$\pi_2\,(\mathrm{HH\Theta}) = \frac{1}{2 + \frac{k_2}{k_1}} \simeq 0.20 \quad \Rightarrow k_2/k_1 \simeq 3 \tag{7}$$

which is very close to the ratio obtained by direct fitting. Replacing (7) in the expression for $\pi_1\,(\mathrm{H\Theta\Theta})$

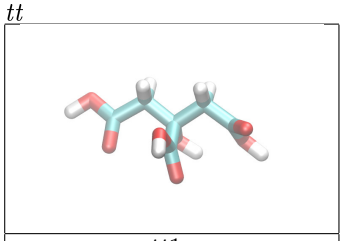
$$\pi_1\,(\mathrm{H\Theta\Theta}) = \frac{1}{2 + \frac{k_1 v}{k_2 u}} \simeq 0.5 \;\;\Rightarrow\;\; \frac{v}{3u} \simeq 0 \tag{8}$$

so that the repulsive interaction between terminal charged groups must be larger than the interaction terminal-central group.

FIGURES of Supplementary Information

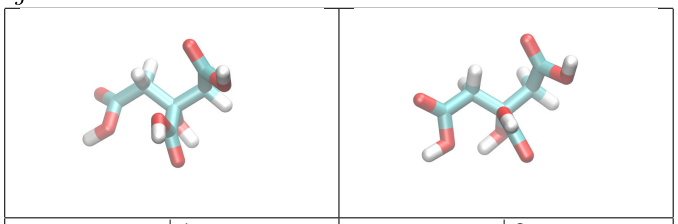
Free energy minimum conformations of citric acid obtained from *ab initio* optimizations in solvent for fully protonated (Fig. S1), mono-deprotonated (Fig. S2), di-deprotonated (Fig. S3) and fully deprotonated (Fig S4) forms. In Fig S2 and S3, it is also indicated the labelling to specify the state of each of the three carboxylic groups: "H" specifies a protonated carboxylic group and "Θ" specifies a deprotonated carboxylic group. The participation of a carboxylic group in a intramolecular hydrogen bond is indicated by a overbar over the corresponding "H" or "Θ" site.

tt



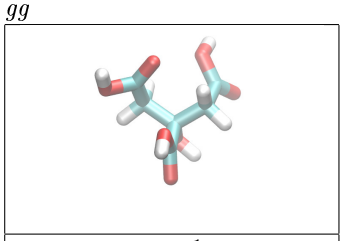
$tt1_0$
0.0

tg



tg^+1_0	tg^+2_0
1.5	1.8

gg



$g^-g^-1_0$
2.9

Figure S1: *tt, tg* and *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with the 6-311++G(d,p) basis set for charge state 0. The labeling indicates the conformation of each of the two principal dihedral angles (*t* for *trans*, *g* for *gauche*). The (+,-) superscript of *g* indicates the sign of the *gauche* angle. Relative free energies with respect to the most stable minimum conformation for each charge state in kcal/mol. The color of atoms correspond to blue for carbon, red for oxygen, white for hydrogen. Intermolecular hydrogen bonds are also indicated by red dashed lines.

tt

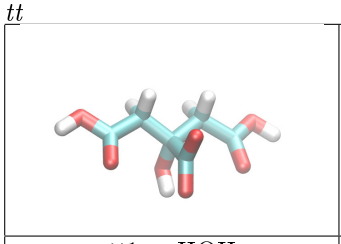
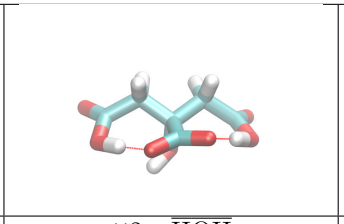
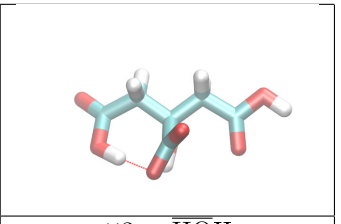
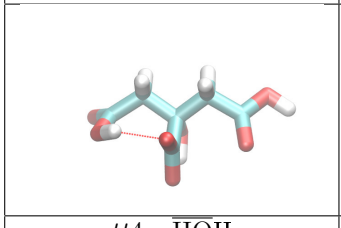
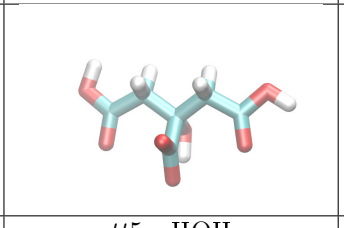
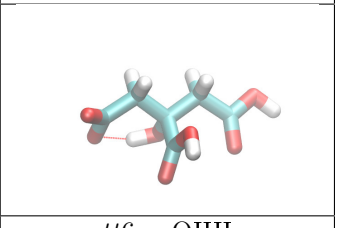
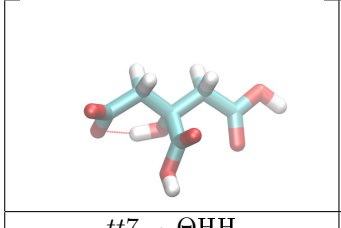
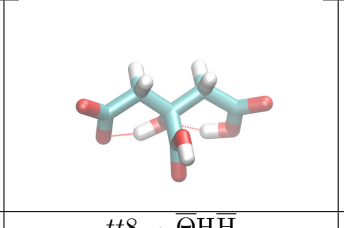
		
$tt1_{-1}$ HΘH	$tt2_{-1}$$\overline{\text{HΘH}}$	$tt3_{-1}$ $\overline{\text{HΘ}}$H
0	1.5	1.6
		
$tt4_{-1}$$\overline{\text{HΘ}}$H	$tt5_{-1}$HΘH	$tt6_{-1}$ ΘHH
2.1	2.2	2.5
		
$tt7_{-1}$ ΘHH	$tt8_{-1}$ $\overline{\text{Θ}}$H$\overline{\text{H}}$	
2.6	2.7	

Figure S2: *tt* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

tg

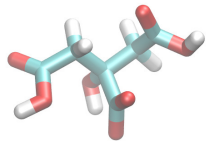
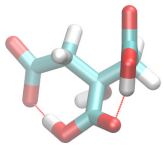
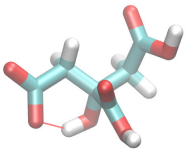
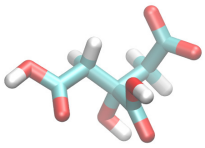
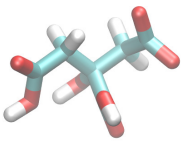
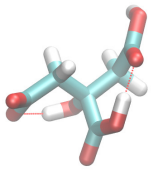
		
$tg^{+}1_{-1}$ HΘH	$tg^{+}2_{-1}$ $\overline{\Theta HH}$	$tg^{+}3_{-1}$ ΘHH
1.2	3.2	3.7
		
$tg^{+}4_{-1}$ HHΘ	$tg^{+}5_{-1}$ HHΘ	$tg^{+}6_{-1}$ $\Theta\overline{HH}$
4.9	6.8	7.0

Figure S3: *tg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

Fig. S3. Minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

gg

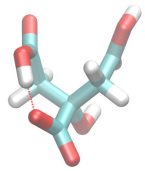
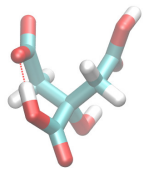
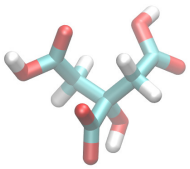
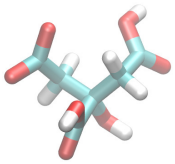
		
$g^{-}g^{-}1_{-1}$ $\overline{\mathrm{H\Theta}}\mathrm{H}$	$g^{-}g^{-}2_{-1}$ $\overline{\Theta\mathrm{H}}\mathrm{H}$	$g^{-}g^{-}3_{-1}$ $\mathrm{H\Theta H}$
0.0	0.9	2.6
		
$g^{-}g^{-}4_{-1}$ $\Theta\mathrm{HH}$		
8.4		

Figure S4: *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -1. The same preferences and units as indicated in Fig S1 were used.

tt

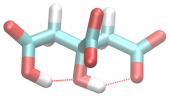
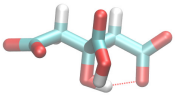
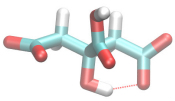
		
$tt1_{-2}$ $\mathrm{H\Theta\Theta}$	$tt2_{-2}$ $\Theta\mathrm{H\Theta}$	$tt3_{-2}$ $\Theta\mathrm{H}\overline{\Theta}$
3.4	3.9	4.8

Figure S5: *tt* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

tg

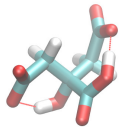
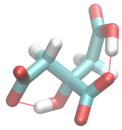
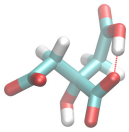
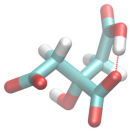
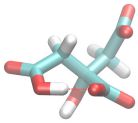
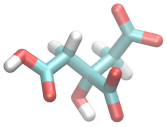
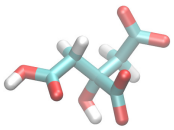
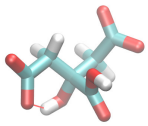
		
$tg^{+}1_{-2}$ $\Theta\overline{\mathrm{H}\Theta}$	$tg^{+}2_{-2}$ $\Theta\overline{\Theta\mathrm{H}}$	$tg^{+}3_{-2}$ $\Theta\overline{\Theta\mathrm{H}}$
0.8	1.2	1.5
		
$tg^{+}4_{-2}$ $\Theta\overline{\Theta\mathrm{H}}$	$tg^{+}5_{-2}$ $\overline{\mathrm{H}\Theta}\Theta$	$tg^{+}6_{-2}$ $\mathrm{H}\Theta\Theta$
3.5	5.0	5.0
		
$tg^{+}7_{-2}$ $\mathrm{H}\Theta\Theta$	$tg^{+}8_{-2}$ $\Theta\mathrm{H}\Theta$	
5.1	6.1	

Figure S6: *tg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

gg

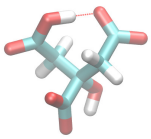
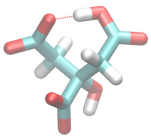
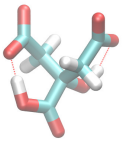
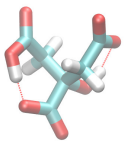
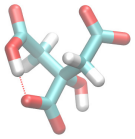
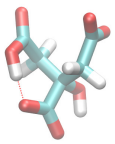
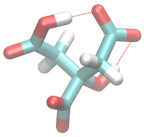
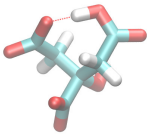
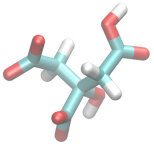
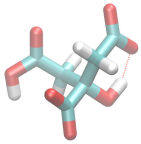
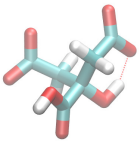
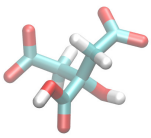
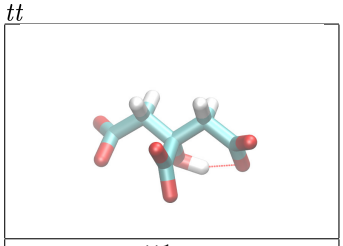
		
$g^-g^-1_{-2}$ $\overline{\mathrm{H}}\Theta\overline{\Theta}$	$g^-g^-2_{-2}$ $\overline{\Theta}\Theta\overline{\mathrm{H}}$	$g^-g^-3_{-2}$ $\overline{\Theta}\overline{\mathrm{H}}\Theta$
0.0	0.3	0.9
		
$g^-g^-4_{-2}$ $\overline{\mathrm{H}}\overline{\Theta}\Theta$	$g^-g^-5_{-2}$ $\overline{\mathrm{H}}\overline{\Theta}\Theta$	$g^-g^-6_{-2}$ $\overline{\mathrm{H}}\overline{\Theta}\Theta$
1.0	1.3	2.8
		
$g^-g^-7_{-2}$ $\overline{\mathrm{H}}\Theta\overline{\Theta}$	$g^-g^-8_{-2}$ $\overline{\Theta}\Theta\overline{\mathrm{H}}$	$g^-g^-9_{-2}$ $\Theta\Theta\mathrm{H}$
3.0	3.3	6.7
		
$g^-g^-10_{-2}\mathrm{H}\Theta\Theta$	$g^-g^-11_{-2}$ $\Theta\mathrm{H}\Theta$	$g^-g^-12_{-2}$ $\Theta\mathrm{H}\Theta$
7.6	7.7	11.1

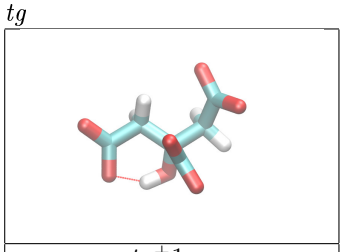
Figure S7: *gg* minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -2. The same preferences and units as indicated in Fig S1 were used.

tt

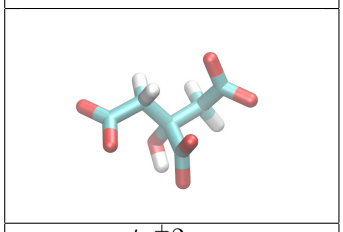


$tt1_{-3}$
0.0

tg

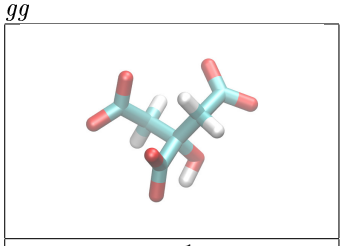


$tg^{+}1_{-3}$
0.4



$tg^{+}2_{-3}$
1.6

gg



$g^{-}g^{-}1_{-3}$
2.6

Figure S8: *tt*, *tg* and *gg* Minimum conformations of citric acid obtained from optimizations at MP2 level with the SMD water model with th 6-311++G(d,p) basis set for charge state -3. The same preferences and units as indicated in Fig S1 were used.

TABLES of Supplementary Information

Tables S1 and S2 show relative free energies and proportions of all conformations of citrate in water for charge states -1 and -2 obtained by MP2 calculations.

Table S1: Relative Free energies (in kcal/mol) in water of different conformations for the mono-deprotonated form of citrate. Conformational proportions are calculated at 25ºC. "H" specifies a protonated carboxylic group and "Θ" specifies a deprotonated carboxylic group. The participation of a carboxylic group in a intramolecular hydrogen bond between two carboxylic groups is indicated by an overbar over the corresponding "H" or "Θ" site.

conformation		ΔG_{rel}	Relative population
$tt1_{-1}$	HΘH	0	23%
$tt2_{-1}$	$\overline{\text{H}\Theta\text{H}}$	1.5	3.7%
$tt3_{-1}$	$\overline{\text{H}\Theta}$H	1.6	3.2%
$tt4_{-1}$	$\overline{\text{H}\Theta}$H	2.1	1.5%
$tt5_{-1}$	HΘH	2.2	1.1%
$tt6_{-1}$	ΘHH	2.5	0.8%
$tt7_{-1}$	ΘHH	2.6	0.6%
$tt8_{-1}$	ΘHH	2.7	0.5%
$tg^{+}1_{-1}$	HΘH	1.2	6.1%
$tg^{+}2_{-1}$	$\overline{\Theta\text{HH}}$	3.2	0.2%
$tg^{+}3_{-1}$	ΘHH	3.7	0.1%
$tg^{+}4_{-1}$	HHΘ	4.9	0.0%
$tg^{+}5_{-1}$	HHΘ	6.8	0.0%
$tg^{+}6_{-1}$	$\Theta\overline{\text{HH}}$	7.0	0.0%
$g^{-}g^{-}1_{-1}$	$\overline{\text{H}\Theta}$H	0.0	48%
$g^{-}g^{-}2_{-1}$	$\overline{\Theta\text{H}}$H	0.9	10%
$g^{-}g^{-}3_{-1}$	HΘH	2.6	0.6%
$g^{-}g^{-}4_{-1}$	ΘHH	8.4	0.0%

Table S2: Relative Free energies (in kcal/mol) in water of different conformations for the di-deprotonated form of citrate. Conformational proportions are calculated at 25ºC. Charged carboxylic groups and intramolecular hydrogen bonds are indicated as in Table S1.

conformation		ΔG_{rel}	Relative population
$tt1_{-2}$	$\mathrm{H}\Theta\Theta$	3.4	0.1%
$tt2_{-2}$	$\Theta\mathrm{H}\Theta$	3.9	0.1%
$tt3_{-2}$	$\Theta\mathrm{H}\overline{\Theta}$	4.8	0.0%
$tg^{+}1_{-2}$	$\Theta\overline{\mathrm{H}}\overline{\Theta}$	0.8	10%
$tg^{+}2_{-2}$	$\Theta\overline{\Theta}\overline{\mathrm{H}}$	1.2	5.3%
$tg^{+}3_{-2}$	$\Theta\overline{\Theta}\overline{\mathrm{H}}$	1.5	3.1%
$tg^{+}4_{-2}$	$\Theta\overline{\Theta}\overline{\mathrm{H}}$	3.5	0.1%
$tg^{+}5_{-2}$	$\overline{\mathrm{H}}\overline{\Theta}\Theta$	5.0	0.0%
$tg^{+}6_{-2}$	$\mathrm{H}\Theta\Theta$	5.0	0.0%
$tg^{+}7_{-2}$	$\mathrm{H}\Theta\Theta$	5.1	0.0%
$tg^{+}8_{-2}$	$\Theta\mathrm{H}\Theta$	6.1	0.0%
$g^{-}g^{-}1_{-2}$	$\overline{\mathrm{H}}\Theta\overline{\Theta}$	0.0	37%
$g^{-}g^{-}2_{-2}$	$\overline{\Theta}\Theta\overline{\mathrm{H}}$	0.3	24%
$g^{-}g^{-}3_{-2}$	$\overline{\Theta}\overline{\mathrm{H}}\Theta$	0.9	8.3%
$g^{-}g^{-}4_{-2}$	$\overline{\mathrm{H}}\overline{\Theta}\Theta$	1.0	6.8%
$g^{-}g^{-}5_{-2}$	$\overline{\mathrm{H}}\overline{\Theta}\Theta$	1.3	4.0%
$g^{-}g^{-}6_{-2}$	$\overline{\mathrm{H}}\overline{\Theta}\Theta$	2.8	0.3%
$g^{-}g^{-}7_{-2}$	$\overline{\mathrm{H}}\Theta\overline{\Theta}$	3.0	0.3%
$g^{-}g^{-}8_{-2}$	$\overline{\Theta}\Theta\overline{\mathrm{H}}$	3.3	0.1%
$g^{-}g^{-}9_{-2}$	$\Theta\Theta\mathrm{H}$	6.7	0.0%
$g^{-}g^{-}10_{-2}$	$\mathrm{H}\Theta\Theta$	7.6	0.0%
$g^{-}g^{-}11_{-2}$	$\Theta\mathrm{H}\Theta$	7.7	0.0%
$g^{-}g^{-}12_{-2}$	$\Theta\mathrm{H}\Theta$	11.1	0.0%